Redesigning a Large Introductory Course to Incorporate the GAISE Guidelines

Roger Woodard
Herle McGowan
North Carolina State University

Journal of Statistics Education Volume 20, Number 3 (2012),
www.amstat.org/publications/jse/v20n3/woodard.pdf

Key Words: Training graduate instructors; Capstone experience; Authentic assessment.

Abstract

In 2005, the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) college report described several recommendations for teaching introductory statistics. This paper discusses how a large multi-section introductory course was redesigned in order to implement these recommendations. The experience described discusses the key sections of the GAISE report and sheds light on the challenges that must be overcome in putting them in place. The result is a course which addresses both the "how to" and big picture of statistics.

1. Introduction

In 2005 the American Statistical Association (ASA) funded a project to develop guidelines on teaching introductory statistics. Through strategic initiative funding, they brought together a group of statistics educators who authored the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) college report (ASA 2005). These guidelines put forth six key recommendations on how to teach an introductory course:

1. Emphasize statistical literacy and develop statistical thinking;
2. Use real data;
3. Stress conceptual understanding rather than mere knowledge of procedures;
4. Foster active learning in the classroom;
5. Use technology for developing conceptual understanding and analyzing data;
6. Use assessments to improve and evaluate student learning.

This paper discusses a redesign of a large multi-section introductory course undertaken to implement these recommendations. The experience of redesign and implementation illuminates the challenges that must be overcome in putting these recommendations in place.

1.1 The Course

The course is a non-calculus based introductory statistics course taught at a large research-focused university in the south-eastern United States. It serves approximately 700 students per semester spread over 10 to 12 sections. These sections are primarily taught by graduate students, although one or two sections may be taught by department faculty members. The students taking the course are a mix of majors from across the university. Although no particular major dominates, many are from the biological sciences or the social sciences. The department offers other courses for engineering and business students. Issues pertaining to the redesign described in this paper would apply to specialized courses such as these. The course typically meets 3 hours per week in a lecture room with 65 students. A night section meets once per week and typically has around 100 students.

1.2 Need for Redesign

Prior to the redesign, instructors were given very limited training and oversight. Each class section covered different topics at different paces. For instance, some sections might cover a full range of topics from descriptive statistics through advanced inference such as ANOVA, whereas others might only touch on the basics of inference in the last few weeks. Grading methods and assessments were also unique to each section. Some sections had only exams while others had extensive data collection projects. Use of statistical software packages varied with some sections using JMP, Excel, or R, although most used no software at all. A course coordinator served primarily as a resource in the event of problems and to review exams.

Some of the key elements of this design included:

- Implementation of a common syllabus and policies.
- Development of common learning outcomes, support materials and final exam.
- Standardization of statistical software.
- A training program for new instructors.

Under the redesigned course all sections used the same calendar and order of topics. This simplification allowed the coordinator to hold discussions and plan support for first time instructors. Specifically, the use of detailed learning outcomes gave clarity in regard to the material that should be taught for each topic. A common final exam provided a more objective measure to facilitate comparison of student performance across different instructors. Standardization of software allowed for improved instruction and the advantages of software use in statistics classes.

The instructor training program consisted of two main phases. The first phase was a two day workshop. During this workshop instructors received a general overview of the course, an introduction to GAISE, training in presentation skills and exposure to key pedagogic principles (training materials from this workshop are available at http://www.amstat.org/publications/jse/v20n3/woodard/training.pdf). The second phase consisted of weekly meetings led by the course coordinator. These meetings began with a brief discussion of administrative details but focused primarily on further training of the instructors. The coordinator and other supporting faculty provided pedagogical advice on detailed examples and activities for the upcoming content. These training activities are discussed throughout this paper.

2. Incorporating GAISE

Throughout the redesign, the GAISE college recommendations were used as the guiding principles for making changes. Effort was undertaken to ensure compliance with these guidelines. Throughout the process we faced challenges that are common to all courses that are trying to implement GAISE, as well as some that are unique to large courses.

2.1 Emphasize Statistical Literacy and Develop Statistical Thinking

As defined by GAISE, statistical thinking is "the type of thinking that statisticians use when approaching or solving statistical problems" (ASA, pg. 14). The report goes on to describe the "Carpentry Analogy" (ASA, pg. 15). This analogy points out an experience to which many statistics students can relate. Material is presented to students in defined segments that relate to specific chapters in a text or specific topics listed on a course plan. These segments have names familiar and meaningful to statistics teachers: measures of central tendency, normal distributions, hypothesis tests for means, etc. Students learn these segments in a rather disconnected way, seeing them as the current hurdle they must overcome in their statistics course. The analogy links the learning of these topics to a carpenter learning to use different types of tools (e.g. saws, hammers, planes). The analogy continues to point out that in practice students are rarely given the opportunity to carry out the process of data collection and analysis from start to finish. This would be equivalent to taking a carpentry class where you never really built a final product such as a table.

To implement this recommendation, we needed to find ways in which students can be involved in the entire process of statistical enquiry, from posing a question, through data collection and analysis, to answering the question. In trying to choose such examples, we faced a variety of challenges:

1. The activity had to be carried out in a standard classroom with materials that could be brought into the class.
2. The activity had to be carried out with a class of 65 students (possibly working in small groups of 3-4 each) in one room.
3. The classrooms lacked computing technology so analysis had to be done by hand or with limited tools.
4. The students needed to be at a point in which they had seen sufficient material to think about the entire process, including designing an experiment to answer a practical question, carrying out that experiment and collecting the data, analyzing the results of the experiment by creating graphics and conducting inferential procedures, and making practical conclusions about the experiment.

The last of these constraints led us to envision a capstone activity that would combine a wide range of topics and help students to link the material together. Many authors have proposed data collection projects; the GAISE appendix suggests some of these activities.

The capstone activity used in this course is an adaptation of the experiment proposed by Utts in the GAISE appendix. In this activity students are asked to determine if wearing gloves significantly reduces the speed at which workers can carry out tasks that require fine motor skills. The students create a matched pairs experiment in which they perform a sorting task with and without gloves. They make decisions on the use of randomization, measurement, and other practical elements. They collect the data and analyze it as a class. They summarize the results and conclusions. Finally, they discuss real-world implications of the conclusions. (For those interested, there is a webinar which presents more detail on the capstone activity: www.causeweb.org/webinar/activity/2010-07/; the student handouts for the activity are available at http://www.amstat.org/publications/jse/v20n3/woodard/Capstone-1.doc and http://www.amstat.org/publications/jse/v20n3/woodard/Capstone-2.doc.)

An additional challenge in implementing a capstone activity is that the course is taught primarily by graduate students who have themselves rarely been asked to carry out the process of data collection and analysis. These graduate students often have not fully developed the art of statistical thinking themselves, so it is difficult for them to convey these ideas to others. To address this, the instructors are trained prior to enacting the activity in the classroom by completing the capstone activity in its entirety, exactly as their students would in class. They are presented with the motivating problem and asked to design an experiment to address the research question of interest, and are led through a discussion on the pros and cons of various ideas until a unified design is settled upon. They carry out this design and analyze the data so that they have a better understanding of what their students will experience. This also provides an opportunity for the trainer to point out areas where the students might struggle (e.g. knowing how to enter the paired data into a spreadsheet for analysis), as well as give examples of questions the instructors can ask their students that emphasize the connections between the various pieces of the project (e.g. seeing how data collection influences choice in the method of analysis).

2.2 Use Real Data

The guidelines point out:

> *It is important to use real data in teaching statistics to be authentic to consider issues related to how and why the data were produced or collected, and to relate the analysis to the problem context. Using real data sets of interest to students is also a good way to engage them in thinking about the data and relevant statistical concepts. (ASA 2005, pg. 16)*

The GAISE appendix justifies why it is important that real data (rather than simply realistic data) be used. In this course we planned several data collection examples that students could do, which allows them to understand the data collection process as well as to see and use the actual data. Following the example of Shaffer, Gnanadesikan, Watkins, and Witmer (1996) and others, we implemented an in-class survey of students as a way to collect data. This anonymous survey is conducted during the first week of class. Data from this survey are aggregated over all sections to produce a data set of over 500 observations. On the day the data are collected and over the next few days the instructors used data from a previous semester as examples for graphics and numeric summaries. The students had just answered these questions themselves and have an intimate understanding of what they were thinking when they filled them out. This in turn helps them predict the distribution of the variables. For example, one question on the survey asks: "How much did you spend on textbooks this term?" The instructor could bring up questions like: "Why would some individuals answer zero for this question?" (e.g. they may borrow textbooks from friends or the library) or "Why would the typical value be around $500?" (e.g. a typical course load is 5 classes and students spend around $100 per textbook). Other variables from the survey could be used to contrast distributions. For example, we could examine year of birth using a survey from two years ago to discuss basic features of the distribution (e.g. why it is skewed to the left and why there is a very large cluster in a two-year span), and then also to discuss how the data collected from this semester's class would differ (e.g. the shape would be basically the same but shifted over by two years). The key is that the instructor should delve into the *why* of the data—not just the mechanics of creating graphics. Using data about the students also allows students to form hypotheses about the data before they ever see it. For example, in the hypothesis testing section of the course we ask students to tell us about how many hours of sleep they had the night before. We also ask the class to decide if the amount of sleep for college students would be greater than or less than the recommended eight hours.

To further encourage use of real data in the classroom, instructors were exposed to a mix of in-class data collections and archival real data sets to use when illustrating specific points. Although we want our graduate instructors to prepare lectures and develop their own examples, we provide them with some specific examples of real data. By exposing them to effective examples that are well planned, we expect them to see the advantage of using real data as well as how to best do so. Many of these examples are introduced in the initial orientation and are further discussed in weekly staff meetings.

Asking instructors to use real data is natural in a statistics class, and one would expect this to occur without any difficulty. Indeed, all of our instructors readily included real data in the classroom. One challenge we faced, however, was that not all of them achieved the use of the data in the way described in GAISE. During in-class observations of graduate instructors, we found that they used real data but did not use it in a way that stimulated students to think about the data. For example, an instructor used a set of data on the weight and city gas mileage of cars. The instructor brought up a PowerPoint slide with the two variables listed and also gave the students a copy of the data on a handout. The instructor's dialogue with the students went something like this:

> *Instructor: "Here is a set of the weights and mileage of 25 models of cars. The weights are given in pounds and the mileage is in miles per gallon."*
>
> *Students: (sitting silently)*
>
> *Instructor: "We can create a scatter plot of these variables and then summarize it with the correlation coefficient."*

In this exchange the instructor did not ask the students questions about the data and the students did not engage in the discussion. The instructor did not ask them to form any hypotheses or in any way interact with the data. Use of data in this manner is not uncommon even among seasoned instructors.

We revised our training to specifically to address this challenge. During a training session we ask our instructors to propose a series of questions they could ask the students to get them involved in this data. When prompted they easily came up with questions like “Who here owns a car?” “Who wishes their car got better gas mileage?” “Do you think there is a positive or a negative correlation between the mileage and weight?” “Do all cars that weigh the same amount get the same gas mileage?” “How do you think these data were collected?” The point of our training activity was to remind instructors that they should engage students in the data. The point of using real data is that it has a story to be discovered and thought about by the students.

2.3 Stress Conceptual Understanding Rather Than Mere Knowledge of Procedures

This recommendation poses perhaps the greatest challenge for any statistics instructor. The steps we took in order to meet it involved: 1) identifying a few core conceptual ideas around which to focus the course, 2) selecting individual topics to include that support these “big ideas”—and removing those topics which do not, and 3) ordering the topics in such a way that the big ideas are continually revisited and reaffirmed. We expand on each of these steps in the remainder of this section.

2.3.1 Identifying Core Conceptual Ideas

The GAISE report points out that “Many introductory courses contain too much material, and students end up with a collection of ideas that are understood only at a surface level, are not well integrated, and are quickly forgotten” (ASA 2005, pg. 17). To address this concern, we created a new schedule that would be followed by all sections in the course. In this revised schedule, we planned the amount of time that would be allotted to each topic, rearranged the order in which topics would be covered, and even removed topics that did not support development and conceptual understanding of a few core topics. These “big ideas” are consistent with the GAISE recommendations on what students should know. They are:

- *How data are collected or generated is important if we want to make inference from it. In both experiments and surveys, randomization is necessary to ensure appropriate inference.*
- *Statistical inference is possible because statistics have a predictable distribution called a sampling distribution.*
- *The sampling distribution allows us to quantify the variability in sample statistics, including how they differ from the parameter (the margin of error) and what type of variability would not be expected to happen by random chance (statistical significance).*

By focusing on a few core ideas, we developed an underlying theme for the course that we could emphasize with the students, allowing us to focus on a connecting thread between topics.

2.3.2 Selecting Topics to Support the Core Themes

For our course, the central tenet is that of the sampling distribution. In the redesigned schedule, tools that were needed to develop the idea of sampling distributions came first. For example, students need to understand the idea of a distribution to comprehend sampling distributions, which in turn requires an understanding of histograms, summaries like means and standard deviations, and models for the sampling distribution such as the normal distribution.

After presenting these tools the sampling distributions for means and proportions are introduced. Because this is a course about statistical inference, we discuss how inference is applied in practice. The ideas of statistical significance and confidence intervals are taught not as disconnected topics but as simple extensions of the sampling distributions idea. But, as can be seen in our "big ideas," the issues of how the samples are collected are equally if not more important. Throughout the course we also emphasize the idea that inference is dependent on using randomness in the production of the data.

In a course on inference the ideas of regression are often a departure from the theme of the course. Rather than remove regression from our syllabus we chose to make a clear connection with the basics of inference. As a result the topic of inference for the slope of a regression line became a part of the material. This topic is less common in introductory statistics courses. By including it we stress the idea that inference is not just for means and proportions; rather, we get to talk about inference in a "big picture" way. We can stress to the students that the process of inference has the same underlying tools regardless of the type of parameter we are learning about.

A challenge in any coordinated course is to communicate what topics are to be taught and to what depth. Although a syllabus may list broad topics or even text chapters, instructors may interpret those topics in very different ways. In addition, students are often overwhelmed by the volume of material and have difficulty focusing their work. To address this challenge, a list of approximately 100 task-level learning outcomes was developed. This list details for the graduate instructors and the students the exact concepts and tasks which the students are expected to know. (The list is available in the Appendix.) For example, one learning outcome asks students to: "Explain that confidence intervals are random quantities which vary from sample to sample and that they may miss the true population parameter. Explain that the confidence level is that proportion of possible samples for which the confidence interval will capture the true parameter." The learning outcomes also cover aspects of statistical literacy (e.g. "Given a study, identify the population, sample, parameter, sampling frame and statistic.") as well as procedures (e.g. "Given a confidence level C, determine the critical value [z^*] from the standard normal table needed to construct the confidence interval.") in addition to the more conceptual items. Furthermore, the list of learning outcomes is used as the basis to construct exams for the course. Since each learning outcome is task level they can be directly assessed. The learning outcomes serve as a checklist for the instructor to create exam questions. Students may use the learning outcomes as a roadmap to study for the exams, thus avoiding the student complaint "I didn't know that would be on the test."

2.3.3 Order of Topics

In addition to careful selection of the topics to be included in the course, we also gave much consideration to the order in which those topics should be presented. The ordering of topics is often discussed among faculty members who teach introductory statistics. This issue is so divisive that most textbook companies have now included the ability to customize chapter order for many of their introductory texts. The question of topic order was examined by Chance and Rossman (2001), who identified several points of agreement in the ordering of topics that we implement in the design of our course.

For our course we considered the importance of the "big ideas" and tried to structure the course appropriately. Our sequence of topics takes into account the central themes of the course as well as the perceived difficulty of the topics. Figure 1 gives our impression of the difficulty of the topics in relation to time.

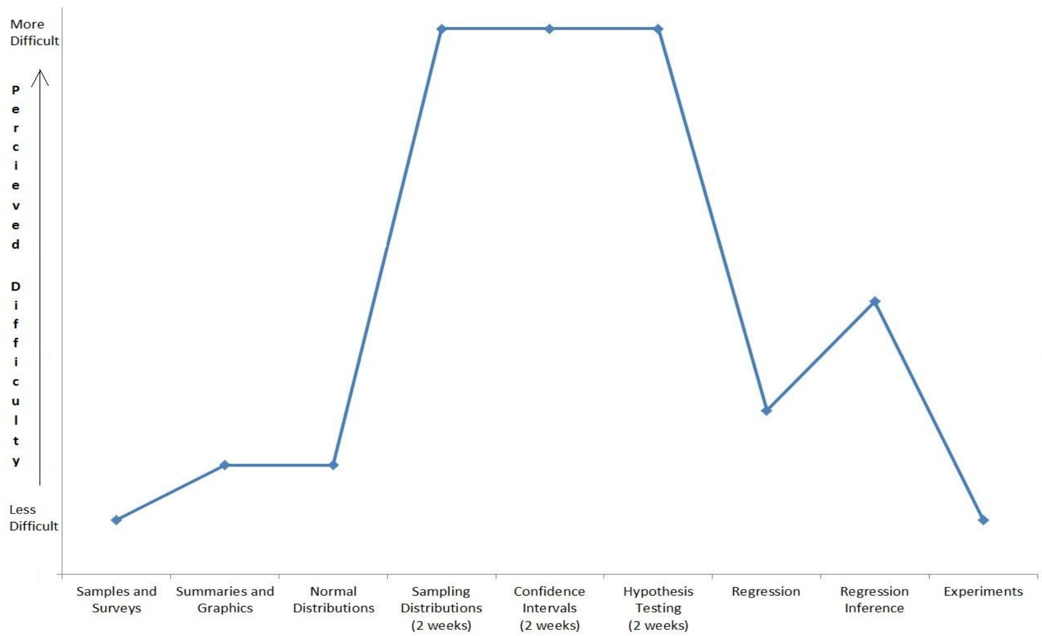

Figure 1: Perceived Difficulty of Re-ordered Course Topics

One of the conclusions put forth by Chance and Rossman is that "data production issues warrant serious attention" (2001, pg. 143). As such, we begin the semester with data collection and discussion of surveys and sampling. This gives students grounding in the principle idea of inference by discussing the ideas of populations, samples, parameters and statistics. Emphasis is given to the ideas of using randomness over convenience in choice of a sample.

The course proceeds very quickly through graphics and summary numbers. Most students have a good grounding in graphics and calculating numeric summaries from their K-12 work. In our experience students do not need to learn to calculate the mean or make a pie chart in a college class and find the material boring. Instead we try to focus on using graphics and numeric summaries to understand the larger concept of distribution and seeing the "story" behind the data. An example of this is incorporated in two of the learning outcomes:

- Given a graphical summary, propose an explanation of the distribution of the data.
- Given a description of a variable, predict what shape the histogram of that variable would take.

The course moves quickly through the idea of the normal distribution. The normal distribution is presented as a model for a histogram rather than in the complex framework of a probability distribution. Prior to the redesign, instructors might spend up to three weeks of the course on the normal distribution and present it as a focus of the course. Although we acknowledge it as a tool, it is only considered a minor part of the course and is allotted just one week. Chance and Rossman point out that "Fundamental ideas should be introduced early and revisited often" (2001, pg143).We feel that sampling distributions are conceptually the most difficult topic for students, and therefore it is important to cover this material early in the course. In our redesigned schedule we have reached this most important topic by the fifth week of the course. We feel it is important to reach this topic early and emphasize it throughout the remaining 10 weeks of the semester as the underlying principle that allows inference.

To achieve this early discussion of sampling distributions, some topics—such as two variable graphics, regression and correlation, and experimental design—were held until later in the course. The basic axioms of probability that discuss intersections, unions and conditional probability were removed from the course. Although many people see these axioms as an important part of the mathematical development of students, ideas of statistical inference can be developed without them and they often form an unnecessary stumbling point for students. (See, for example, Garfield and Ahlgren 1988)

Sampling distributions and the main tools of inference—confidence intervals and hypothesis testing—take up the middle six weeks of the course. These ideas are conceptually difficult for students so the difficulty level plateaus in the middle of the course. This allows students time to learn the material and then reflect on it over the remaining weeks. In traditional courses hypothesis testing is often the last topic in the courses and students are forced into a final exam within a few days of this topic. Instead we decrease the difficulty level by bringing in simpler topics, such as regression and basic experimental design. We approach correlation and regression with an emphasis on technology, thus minimizing formulas and the cognitive difficulty for students. We form a connection to the idea of inference by also teaching the hypothesis test for the slope of the regression line. The final major topic we discuss is experimental design. This ends the course with another topic about data collection, so that the course begins and ends with a topic about the source of the data. Our intention in doing this is to give students leaving the course the opinion that statistics is less about the calculation of formulas and more about the collection and analysis of data.

The last week of the course is dedicated to the capstone activity, which serves several purposes. One purpose is fostering statistical thought as discussed previously. The other is to serve as a review of topics covered in this course and to see them in combination.

The order of topics is intended to help with conceptual understanding, but these ideas must be accentuated with the graduate instructors. During our weekly meetings with the instructors we discuss the conceptual links between the topics. Faculty help the graduate instructors brainstorm how the links between these topics can be discussed and taught.

2.4 Foster Active Learning in the Classroom

The use of active learning in the classroom is not new and many of our graduate instructors have experienced this during their undergraduate studies. However, most of them have not thought about how to enact it in a class they are teaching. Few of the graduate courses they are taking incorporate active learning. Additionally, many of the international graduate students come from cultures that emphasize students quietly sitting and absorbing material. As a result, graduate instructors tend to rely on simple lecture as their classroom mode. To address this challenge, we incorporated several strategies intended to encourage the use of active learning.

The first, and perhaps simplest, strategy was to formalize a portion of the course grade dedicated to participation in in-class activities. These points were a relatively small proportion of the overall course grade (approximately 10%). By assigning points for in-class activities we obtain student buy-in to the completion of the activities used.

In addition to getting the students to be interested in the activities, we must also enable the section instructors to incorporate active learning in a meaningful way. This was accomplished by giving instructors training in a variety of strategies to encourage active learning as well as in implementing specific activities. For example, a problem that may arise during classroom activities is having a student or group simply waiting for the instructor to give the solutions. To avoid this, we encourage the instructors to call on a group at random to present their solution on the board. Students are told that they may be selected and feel the pressure to put effort into the activity.

Substantial time during the initial orientation is also used to illustrate appropriate working of problems in class. Students in this course are given “study guides” that contain lists of practice problems from the course text. When students ask about these problems, new instructors often feel obliged to simply go to the board and quickly write out the answer. We instruct new instructors on basic methods of working problems in a student-centered manner. In this method, the instructor acts as a moderator guiding the student’s thought process and recording their answers. Specific suggestions include:

- Ask students to read the question from the text.
- Model the decision process that students should go through and ask the students to take a vote on key decisions. For example, asking if the problem is about means or proportions.
- If numbers are given in the problem ask the class to decide what the numbers represent through a series of multiple choice questions. For example, is this a sample mean or population mean?

- If calculations are needed ask all students to take out their calculators and perform the calculations. Break the calculations into steps and have each student calculate the answer at each step.
- While the students are doing the calculations walk up the aisles and ask individuals for the numbers they have found. If some students want a "free ride" and do not bring their calculator we encourage instructors to take a few calculators along to class and hand them out to students.

Finally, during the initial orientation, the graduate students go through the capstone activity as if they were students (as described previously). Since there is a limited amount of time during the orientation session, we use time during the weekly staff meetings as a time to demonstrate activities that can be used in the upcoming weeks. We have found this just-in-time approach to be more effective with graduate instructors who are often thinking only about the next few lessons.

Most of the activities demonstrated during these sessions are classic illustrations used in introductory classes. For example, to illustrate the concept of confidence level groups of students are given bags of M&M candies, which they use to find the proportion of blue candies and then calculate a confidence interval for the true population proportion (Pearl and Woodard, 2006). The results from each group are then collected on the board so that students can examine the percent that contain the true population proportion, illustrating the concept of confidence level. Another activity asks students to determine if their dominant hand is stronger than their non-dominant hand. This is done using a simple hand dynamometer that measures grip strength (www.vernier.com). The activity serves as an introduction to the ideas of paired differences. It requires the students to formulate the appropriate hypotheses to test, determine the method for data collection, collect the data from volunteers, and form a conclusion based on the hypothesis test.

Throughout the demonstrations we continue to reiterate points that are described in GAISE such as:

- Ground activities in the context of real problems. Data should be collected to answer a question, not "collect data to collect data" (without a question).
- Collect data from students (anonymously).
- Encourage predictions from students about statistical results.
- Include assessment as an important component of an activity.

Along with the idea that assessment should be an important component of an activity, we also stress a point from Gelman and Nolan (2002) that the instructor should summarize the results of the activity. We encourage the instructors to have the "take-away" message included in their notes. This helps novice instructors make sure that the students see the point of the activity and don't become lost in the minutia of the activity itself.

Although demonstrating the activities is often enough to get many instructors to use them, we have found that two other support elements are needed to get wide adoption. First, any required equipment or materials should be provided. For example, some graduate instructors are unable to pay for the M&Ms for the confidence interval activity or the dynamometers for the grip strength activity. Accordingly our department purchases this equipment for them. The second major requirement is that the instructors must have enough time in the course schedule to incorporate the activities and still cover all the required topics. We designed the course syllabus to have reasonable coverage but not be overly aggressive. During the initial orientation workshop instructors are presented with strategies for covering material in an efficient way. For example, we encourage all instructors to produce guided note outlines of the course material. These outlines provide a general structure of a day's lecture, key equations and definitions, along with the text and graphics of any examples being used. These outlines can be provided to the students as either handouts or as postings to the section website. During class students no longer need to spend large amounts of time copying the text of examples from the board; instead they simply fill in the key parts. Thus, rather than using a class period for students to imitate copy machines, time is kept for them to actively engage in the material. In fact, Austin, Lee, Thibeault, Carr, and Bailey (2002) showed that the use of these outlines improved student retention and interaction in the classroom.

2.5 Use Technology to Help Understanding

The use of statistical software allows statisticians to explore data and find answers to questions that would not be possible otherwise. The GAISE report suggests that:

> *…technology should be used to analyze data, allowing students to focus on interpretation of results and testing of conditions, rather than on computational mechanics. Technology tools should also be used to help students visualize concepts and develop an understanding of abstract ideas by simulations (ASA 2005, pg. 19).*

Prior to this course redesign, the use of technology was both limited and varied from section to section. One challenge we faced when trying to integrate technology use across the sections was the sheer size of the course enrollment—there was insufficient space in the University's computer labs to hold organized sessions for all enrolled students. To meet this challenge, we have incorporated the web-based statistical software package StatCrunch® (West, Wu, and Heydt 2004). This software can be used individually by students on any computer and can be demonstrated in the classroom by the instructors.

The use of statistical software in a room in which each student does not have access to a computer can be challenging. Instructors have a main computer that can be used for technology demonstrations, but care needs to be taken so that students do not "tune out" during this time. During the graduate instructor training, we model the process of engaging students with the technology by asking them questions to guide what the instructor does next, as well as to make conjectures that can be tested and evaluated. For example, when demonstrating graphical and numeric summaries, instructors use data from the previously described background survey completed by the students. They ask students questions to engage with the analysis, such as "What type of graph would be appropriate to summarize this variable?" The instructor then creates and displays the specified graph. The instructor also asks students to speculate about the data, such as the characteristics of the distribution for a particular variable, or to compare distributions for various groups (e.g. "Do you think men or women will have higher values for this variable?"). These student conjectures are then tested (by the instructor creating the appropriate graphs or summaries using the software) and evaluated by the class. In these technology demonstrations, the instructor physically controls the software, but the students direct the analysis. This also helps ensure that students are thinking statistically about the data, not just crunching numbers.

There have been several good computer applets developed to demonstrate abstract statistical concepts that students often struggle with. As discussed previously, the idea of sampling distributions is central to the course, but it is also one of the most difficult topics of the semester. Fortunately, there are applets to simulate the process of repeated sampling and build the sampling distribution, so that students can “see” what is happening. To help students understand what the applet is doing, they first engage in a hands-on simulation that helps tie the applet to a real world problem. For example, to illustrate the Central Limit Theorem, students are asked to consider the distribution of the mint dates for the population of U.S. coins. Students are able to reason that the shape of this population is left-skewed, since most coins will have been minted in recent years while older coins tend to fall out of circulation. Students are asked to bring a sample of coins to class and calculate the average date for their sample (they work in groups of 3, with each student contributing 10 coins for a total sample size of 30 coins). Each sample average is then plotted on the board and the instructor uses the resulting dotplot to lead a class discussion of the distribution of the sample mean. Because each group placed an observation on the board, the students should understand what each dot represents. Furthermore, because they saw their classmates carrying out the same procedure, they should understand where the repeated samples are coming from. When the students realize that 20-30 samples do not give a very “full” distribution, the instructor moves to a computer applet that simulates the process they just experienced. This allows them to better understand what the applet is doing and, therefore, the origin of the sampling distribution. In turn, they can focus their attention on the important concepts behind sampling distributions and the Central Limit Theorem (e.g. what happens to shape, center, and variability as the sample size changes). As with all of the important activities in the course, the graduate instructors run through this entire activity during their weekly training, so that they know the important concepts and links to emphasize with students. As with other technology demonstrations used in the class, the instructors ask students to make conjectures that are tested using the applet and evaluated by the students.

2.6 Use Assessment to Improve Learning

The GAISE report indicates that it is important to use assessment not only to assign a grade but also to guide student learning. This can be difficult to accomplish in a timely manner in a large course. In our redesigned course we implement several assessment strategies that are designed to improve student learning.

One way we met this challenge is through a change in the structure of the weekly homework assignments. Historically students submitted hard copies of homework assignments to an instructor. In an ideal world, the students would receive timely feedback on their work. With a large class the grading of individual homework assignments becomes prohibitively time consuming. Even with dedicated graders assigned to the course, the amount of feedback on written assignments was often minimal. Students would receive scores with cryptic marks that may or may not have been informative. This process would typically take at least one week. To decrease the time before students see feedback and increase the quality of the feedback, we make use of an online homework system called WebAssign®.

Although the online system was introduced in part for convenience, it can also be used to improve student learning. The system can automatically score students' responses providing instantaneous feedback on an incorrect submission. Thus students know immediately when they have answered a question correctly and are more likely to remember the logic they used in answering the question.

We can capitalize further on this feedback by spurring the student's learning process after an incorrect answer. After a wrong submission students are given an opportunity to resubmit their answers. The system notifies them that they are wrong and students try to correct those incorrect answers. To help focus the student's learning we provide formative feedback for each question. This feedback contains connections to the textbook, where students can find a review of the relevant material. For instance, the student may incorrectly answer a question on a calculation of the normal distribution. The student may receive feedback that says "Review pages 305 to 308 of the textbook to see examples related to this concept. Then for practice try problem 8.63 from the text and check your work with the solutions in the student solutions manual." By giving students this type of feedback we encourage them to read the textbook and learn from the material presented there. By giving them worked solutions, we also encourage them to try additional problems that are not specifically assigned. Thus, the process of completing the homework assignments becomes a learning experience for the students.

Another benefit to using electronic homework is that the numbers in a problem can be randomized for each student in the class, ensuring that students cannot simply copy from a friend. This allows students to work together to determine how a problem is solved, which we encourage, but the student must still do their own version of the problem. It is even possible to randomly assign data sets to students. In the assignment on summarizing data, students must explore one of several datasets to determine which of the multiple-choice responses is correct for their data. In this example, each of the response choices includes a detailed conclusion about the data that can only be reached by exploring several variables in conjunction.

In addition to the homework, students are provided with optional problems in the study guides to expand their understanding. These problems come from the textbook, with solutions in the back of the book so that students can check their answers. The study guides not only provide opportunity for extra practice; they also connect the homework with the textbook and connect the content with the course learning outcomes, allowing each facet of the course to be well integrated.

Although homework is the most continuous source of formative feedback the students receive, we also strive to help them learn from the course exams. Students are able to use the study guides, learning outcomes, and practice problems to prepare for exams. After the exams have been graded and returned to students, we take class time to discuss common mistakes or misunderstandings. We also discuss which learning outcome was covered by each question, further reinforcing these outcomes as the "roadmap" for the course.

3. Discussion

In this paper, we reviewed many of the changes made to a large introductory statistics course in order to incorporate the GAISE guidelines. We also discussed the challenges that needed to be overcome in this redesign. Fortunately, reasonable solutions to these challenges could always be found, as summarized in Table 1. The result was a course where students were actively engaged in both the "how to" and big picture of statistics.

As with any major project, there are costs associated with implementation. In particular, this redesign required extensive amounts of time to be committed on the part of the faculty. A major component of this time included development of a training program for the graduate instructors. The cost of delivering this training to the graduate instructors is ongoing, as it will be needed for each group of new instructors. A coordinating faculty member also needs to meet with graduate instructors throughout the semester as part of the ongoing training. At our institution, we formalized this training by enrolling each graduate student in a 1-credit course called "Supervised Doctoral Teaching."

A second major component of the faculty's time commitment included development of the detailed learning outcomes. For this redesign, we started with a list of outcomes used by faculty members for other courses. For the reader that is considering a redesign of their own course, our learning outcomes (available in the Appendix) can serve as a starting point. We believe the process of putting together learning outcomes for a course is important and should include a group of interested faculty who can discuss the merits of each topic/outcome and how they fit together to form overall themes.

Faculty time was also involved in adapting activities such as the capstone activity for the course. We have found that available activities are unlikely to fit exactly into every course. Thus, the coordinator will need to adjust activities for their particular course. In addition to the cost of time, activities may incur a cost for materials. For example, the capstone activity required purchase of large amounts of dry beans, containers that could be used for storage and sorting, as well as many pairs of work gloves. While these costs were one-time only, other activities have ongoing costs. For example, the M&M activity required the purchase of large quantities of candy each semester.

Table 1: Summary of Implementation

GAISE Recommendation	Implementation Challenge	Implementation Solution
Emphasize statistical literacy and develop statistical thinking	Finding a way for students to experience statistical thinking by experiencing the full process of statistics, especially given the size of the course	A capstone project that could be conducted during class time
Use real data	Graduate instructors often did not understand how to use the reality of the data	Modeled process in training sessions
Stress conceptual understanding rather than mere knowledge of procedures	Course lacked common core themes to guide instruction	Formulating the “big ideas” for the course, and structuring the topics and sequencing to support these “big ideas”
Foster active learning in the classroom	Graduate instructors were often not prepared to implement active learning	Modeled process in training sessions and provided key activities
Use technology for developing conceptual understanding and analyzing data	Engage students with simulation to illustrate concepts, especially given class size	Coupling tactile simulations with computer simulations; Making use of web-based technologies
Use assessments to improve and evaluate student learning	Providing timely formative feedback	Capitalizing on online technology to provide feedback

Perhaps the most substantial cost associated with the redesign was in the initial implementation of the online homework. Although we incorporated a commercial system that was easy to use, a faculty member or graduate assistant needed to create the questions with the appropriate feedback in this system. We employed a graduate student over the course of two summers to create and then refine both the questions and the feedback. Fortunately, once the online homework is created, the cost associated with using it each semester is minimal.

In addition to the faculty costs, the redesign incorporates some costs for the student. The software package StatCrunch is relatively inexpensive (about $12 for 6 months at the time of this writing), as is access to Webassign (about $20 per course). We feel these monetary costs are offset by the substantial pedagogical gains produced by the adaptation of GAISE to our course. The student’s cost for Webassign is also offset by the advantage of instantaneous feedback. Additionally, the use of online homework resulted in a cost savings at the course-level. Previously each section of 65 students required a 20-hour per week graduate student to grade papers, in addition to the graduate instructor. The use of the online homework system has reduced this need to just 0.65 additional graduate students (at 20 hours per week) per section.

4. Future Directions for the Course

We are always seeking ways to improve the course. For example, we learned from students that, while we tried to build the homework as a learning experience, they tended to feel the assignments were too easy. We believe this may be because, rather than using the formative feedback provided to help them learn, they were employing more of a "guess and check" strategy (Pascarella, 2004). As such, we no longer allow multiple submissions on the homework. Students are still provided with formative feedback on each problem after they submit the assignment for grading, but they must think more about the concepts while they are working on it.

Our next goal in improving the course is to increase the efficiency of the learning experience for all students. Dennis Pearl, in his address at the 2011 United States Conference on Teaching Statistics, said:

> *…to effectively reach all learners we must provide a way to personalize each teacher's pedagogical options and each student's educational experience….this dream is within reach: support for statistics education has never been higher; resources have never been more abundant, and the technology required for personalization is now ubiquitous.*
> (http://www.causeweb.org/uscots/uscots11/speakers/plenary3.php)

We hope to build on the strengths of the current redesign to create a personalized experience that balances both individualized and group learning. This new redesign is in its early stages. It will be the subject of future papers.

5. Supplemental Materials

The following supplemental materials are available for download and modification as the reader wishes:

- Capstone: www.amstat.org/publications/jse/v20n3/woodard/Capstone-1.doc and www.amstat.org/publications/jse/v20n3/woodard/Capstone-2.doc
- Homework: www.amstat.org/publications/jse/v20n3/woodard/homework.doc
- TA Training: www.amstat.org/publications/jse/v20n3/woodard/training.pdf

Appendix:
Course Learning Outcomes

Data Collection and Surveys:

A1. Given a study, identify population, sample, parameter, sampling frame and statistic.
A2. Given a survey sample, determine if the sample is a voluntary response sample or a convenience sample.
A3. Given a study, recognize typical forms of biases such as potential undercoverage, nonresponse, and response bias.
A4. Given a study, determine whether a SRS, stratified random sample, cluster sample, or systematic sample was selected.
A5. Given a study's objective, decide when to use a stratified random sample, cluster sample, systematic sample, or SRS.
A6. Given a description of a study determine if the study is a census or a sample survey.

Summarizing with Graphics:

B1. Given a set of raw data, identify the individuals and the variables.
B2. Given a variable, determine whether it is categorical or quantitative.
B3. List which graphical methods (pie charts, histograms, etc.) are appropriate for categorical and for quantitative variables.
B4. Given a histogram, stem plot, or dot plot, determine the number of individuals in a particular range.
B5. Given a histogram, dot plot, or stemplot, describe the distribution's shape (skewed left, skewed right, symmetric, or multimodal), center, and spread.
B6. Given a histogram, dot plot, or stemplot identify values that would be considered outliers.
B7. Given a graphical summary, propose an explanation of the distribution of the data.
B8. Given a description of a variable, predict what shape the histogram of that variable would take.

Summarizing with Numbers:

C1. Explain how the mean and median are related for different shapes of a distribution (skewed left, skewed right or symmetric).
C2. List the following characteristics of the standard deviation
  a. The standard deviation must be greater than or equal to zero.
  b. When standard deviation is equal to zero, there is no spread – every number on the list is the same.
C3. Given a set of summary statistics (mean, median and standard deviation), find the summary statistics of a data set that would result from a linear transformation of the original data. (A linear transformation means adding or subtracting the same value from each observation and/or multiplying or dividing each observation by the same value).
C4. Given a histogram, be able to determine the approximate location of the median and quartiles.

C5. Match given histograms, dot plots, or boxplots to given sets of appropriate summary statistics. (For example, mean, median, standard deviation and quartiles).
C6. Explain the impact of outliers on summary statistics such as mean, median and standard deviation.
C7. Given a boxplot, determine the five number summary for that data.
C8. Given a boxplot, determine if a distribution is skewed right or skewed left.
C9. Given side-by-side boxplots, contrast key features of the groups represented by the boxplots.

The Normal Distribution:

D1. Explain that the normal distribution is a model for a bell-shaped histogram.
D2. List the key characteristics of the normal distribution.
D3. Given a mean μ, standard deviation σ, and observed value *x*, calculate the standardized value (z-score). Describe the characteristics of a standard score.
D4. Given a z-score, use a normal table to find the corresponding probability.
D5. Given a mean μ and standard deviation σ, find a specified percentile of the normal distribution. (e.g. Given a probability find the corresponding value of x.)

Sampling Distributions:

E1. Describe the sampling distribution of a statistic and define the standard deviation of a statistic.
E2. Given a study, describe the sampling distribution of x-bar as specifically as possible. This involves stating whether this distribution is at least approximately normal.
E3. Given a population standard deviation (σ), calculate the standard deviation of the sample mean $\bar{x}$, using the formula $\sigma/\sqrt{n}$.
E4. Given a population mean (μ), standard deviation (σ), sample size (n) and sample mean, calculate the standardized value (z-score) for a sample mean.
E5. Given a population proportion (p), calculate the standard deviation of the sample proportion, $\hat{p}$, using the formula $\sqrt{p(1-p)/n}$.
E6. Given a study, describe the sampling distribution of the sample proportion ($\hat{p}$) as specifically as possible. This involves stating whether this distribution is at least approximately normal.
E7. Given a population proportion (p), sample size (n) and sample proportion $\hat{p}$, calculate the standardized value (z-score) for a sample proportion.

Confidence Intervals:

F1. Define the standard error of a statistic.
F2. Calculate the standard error of the sample proportion $\hat{p}$ using the formula $\sqrt{\hat{p}(1-\hat{p})/n}$ or the sample mean $\bar{x}$ using the formula $s/\sqrt{n}$

F3. Given a study, determine whether the study meets the conditions under which inferences on a population proportion may be performed. (For example, requiring a simple random sample).
F4. Given a confidence level C, determine the critical value (z^*) from the standard normal table needed to construct the confidence interval.
F5. Explain that confidence intervals are random quantities which vary from sample to sample and that they may miss the true population parameter. Explain that the confidence level is that proportion of possible samples for which the confidence interval will capture the true parameter.
F6. Construct a confidence interval for a population proportion using the formula $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
F7. Given a study, interpret the result of a confidence interval in the context of the problem.
F8. Given a study, determine whether the study meets the conditions under which inferences on a population mean may be performed. (For example, requiring a simple random sample). Also explain how inferences based on the t-distribution are robust.
F9. Explain why we use the t-distribution instead of the normal distribution when making inference on the population mean.
F10. Given a sampling situation, determine the appropriate degrees of freedom associated with the t-distribution.
F11. Explain the differences and similarities between the normal and t-distributions. (For example, the t-distribution is more variable but approaches normality as *n* increases.)
F12. Given a confidence level C, determine the critical value (t^*) from the t-table needed to construct the confidence interval.
F13. Construct and interpret a one-sample confidence interval for the mean based on the t-distribution using the formula $\overline{x} \pm t \frac{s}{\sqrt{n}}$
F14. Given a study and confidence interval, describe how the following will affect the width of the confidence interval.
  a. Increasing the sample size
  b. Increasing the confidence level C

Tests of Hypothesis:

G1. Given a study objective, determine whether significance testing is appropriate.
G2. Given a study objective, choose appropriate null and alternative hypotheses, including determining whether the alternative should be one-sided or two-sided.
G3. Given a study and p-value, explain in context that p-value is a probability of getting a sample statistic as extreme or more extreme than what was seen in the sample given that the null hypothesis is true.
G4. Given a test statistic, calculate a p-value based on the standard normal distribution or t-distribution as appropriate.
G5. Given a study, interpret the results of a test of significance in context.
G6. Given a study objective, significance level (α) and summary statistics, conduct a formal test of significance on a population mean (or a population proportion) by conducting the

appropriate steps. (This includes choosing and stating hypotheses, calculating a test statistic, calculating and interpreting the p-value and interpreting the conclusion of the test in context.)

G7. Explain the relationship between a confidence interval and a two-sided hypothesis test.
G8. Given results from a hypothesis test, comment on the impact of sample size and the practical importance.
G9. Given a description of a study define statistical significance in context.

Inference for Means of Paired Differences:

H1. Identify a matched pairs design and when inference is appropriate in this situation.
H2. Given a study, describe the sampling distribution of the sample mean of paired differences as specifically as possible.
H3. Conduct a statistical inference (confidence interval or significance test) based on matched pairs data.

Correlation, Scatterplots, Introduction to Regression:

J1. Given a study, distinguish between explanatory and response variables.
J2. Given a set of raw data, make a scatterplot or regression output using appropriate software.
J3. Given a scatterplot, identify patterns such as positive and negative associations, non-linear patterns and outliers.
J4. Describe the characteristics of the correlation coefficient.
J5. Given two variables and their correlation coefficient, describe how the correlation changes if the units of either variable are changed.
J6. Given two variables (x and y), describe the correlation you would expect to find between x and y.
J7. Match given scatterplots with possible values of the correlation coefficient.
J8. Explain the relationship between the slope of the regression line and the correlation coefficient.
J9. Given the least squares line and a value of x, calculate the predicted value of y.
J10. Identify situations in which it is not appropriate to summarize the relationship between variables using a least squares line.
J11. Given standard regression output, identify and utilize key parts of the output (estimated slope, intercept, r^2 etc.)
J12. Given a study, explain in context that the regression method is used to estimate the average value of y when you know x and that individual values will vary around the predicted value of y.
J13. Given a study, interpret the value of the square of the correlation coefficient(r^2). That is, explain that it measures the proportion of the variance of one variable that can be explained by straight-line dependence on the other variable.
J14. Given a least squares line and an observation (x,y), calculate the residual for that observation.
J15. Identify situations where the correlation coefficient would not do a good job of summarizing the relationship between two variables.

J16. Given a scatterplot, contrast the influence of different outliers on the least squares regression or correlation coefficient.
J17. Given a study, explain why it might not be a good idea to use a least squares line to predict beyond the range of data that were used to create the line.
J18. Given a study, explain why correlation or association does not imply causation. That is, explain that the association may be due to common response, confounding or unusual events.

Inference for Regression:

K1. Given a study explain what is meant by the true regression line and describe the sampling distribution of the sample slope.
K2. State the assumptions of inference about the regression model.
K3. Explain the practical reason for testing that the slope is zero.
K4. Given a study objective, significance level (α) and summary statistics, conduct a formal test of significance on a slope based on the t-distribution by conducting the appropriate steps. (This includes choosing and stating hypotheses, calculating a test statistic, calculating and interpreting the p-value and interpreting the conclusion of the test in context.)
K5. Given standard regression output, interpret the results of the test of hypothesis about the slope.

Experiments:

L1. Given a study, determine whether it is an observational study or an experiment.
L2. Given a study, identify subjects and treatments.
L3. Given a study objective, explain the advantage of using a control group.
L4. Define the placebo effect and explain the purpose of a placebo.
L5. Given a study, identify whether a placebo and/or control group were used.
L6. Given a study, determine whether a completely randomized design was used.
L7. Given a study objective, describe how to implement a completely randomized design.
L8. Given a study, explain why randomization should be used.
L9. Given a study, determine whether the experiment was blinded or double-blinded.
L10. Given a study objective, explain the advantage of using a double-blind experiment.
L11. Given a study objective, describe if and how a matched pairs experiment could be used.
L12. Given a study objective, explain the advantage of using a matched pairs design.
L13. Given a study, determine whether a blocking design was used and describe the blocks.
L14. Given a study objective, explain the advantage of using a blocking design.
L15. Given a study objective, describe an appropriate comparative experiment using the principles of randomization, replication, control, blocking, double-blinding, placebo, and control group.

References

American Statistical Association. (2005), *Guidelines for Assessment and Instruction in Statistics Education*, available at http://www.amstat.org/education/gaise/index.cfm.

Austin, J.L., Lee, M.G., Thibeault, M. D., Carr, J. E., and Bailey, J.S. (2002), "Effects of Guided Notes on University Students' Responding and Recall of Information," *Journal of Behavioral Education*, *11(4)*, 243–254.

Chance, B. L. and Rossman, A. J. (2001), "Sequencing Topics in Introductory Statistics: A Debate on What to Teach When," The American Statistician, *55(2)*, 140-144.

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

Gelman, A. and Nolan, D. (2002), *Teaching Statistics: A Bag of Tricks*, New York NY: Oxford University Press.

Pascarella, A. M. (2004), "The Influence of Web-Based Homework on Quantitative Problem–Solving in a University Physics Class," in *Proceedings of the NARST 2004 Annual Meeting* (Vancouver, BC, Canada).

Pearl, D. K. and Woodard, R. D. (2006), *Labs and Activities Supplement for Moore and Notz's Statistics: Concepts and Controversies (6th ed.)*, New York NY: Freeman, 45-48.

Shaffer, R. L., Gnanadesikan, M., Watkins, A. and Witmer, J. (1996) *Activity-Based Statistics: Instructor Resources*, Emeryville, CA: Key College Publishing, 6-7.

West, W., Wu, Y. and Heydt, D. (2004), "An Introduction to StatCrunch 3.0," *Journal of Statistical Software, 9(5).*

Roger Woodard
Department of Statistics
North Carolina State University
5272 SAS Hall
Raleigh, NC 27695
Phone: 919-515-1938
Roger_Woodard@ncsu.edu

Herle McGowan
Department of Statistics
North Carolina State University
5262 SAS Hall
Raleigh, NC 27695

Phone: 919-515-0634
mcgowan@stat.ncsu.edu

Volume 20 (2012) | Archive | Index | Data Archive | Resources | Editorial Board | Guidelines for Authors | Guidelines for Data Contributors | Guidelines for Readers/Data Users | Home Page | Contact JSE | ASA Publications