Fostering Conceptual Understanding in Mathematical Statistics

Authors: Jennifer L. Green and Erin E. Blankenship

Fostering Conceptual Understanding in Mathematical Statistics

Jennifer L. Green*

Department of Mathematical Sciences, Montana State University

and

Erin E. Blankenship

Department of Statistics, University of Nebraska–Lincoln

July 1, 2015

Abstract

In many undergraduate statistics programs, the two-semester calculus-based mathematical statistics sequence is the cornerstone of the curriculum. However, ten years after the release of the Guidelines for the Assessment and Instruction in Statistics Education (GAISE) College Report (GAISE, 2005) and the subsequent movement to stress conceptual understanding and foster active learning in statistics classrooms, the sequence still remains a traditional, lecture-intensive course. In this paper, we discuss various instructional approaches, activities and assessments that can be used to foster active learning and emphasize conceptual understanding while still covering the necessary theoretical content students need to be successful in subsequent statistics or actuarial science courses. In addition, we share student reflections on these course enhancements. The course revision we suggest doesn't require substantial changes in content, so other mathematical statistics instructors can implement these strategies without sacrificing concepts in probability and inference that are fundamental to the needs of their students. (*Note: Supplementary materials, including code used to generate class plots and activity handouts, are available online.*)

Keywords: Active learning; Activities; Assessments; Course Revision; Writing

1 Introduction

Creating a pipeline of statistically savvy students is critical for building and sustaining the nation's workforce and research enterprise (Manyika et al., 2011). With the growing presence of data in society and the need for analytically minded employees, undergraduate programs in statistics are charged with preparing graduates who are able to "think with data" (ASA, 2014). The newly released document, *2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science* (ASA, 2014), advances such efforts by providing a set of curriculum recommendations for undergraduate students majoring or minoring in statistics. These guidelines recommend that programs carefully integrate "statistical theory, statistical application, data manipulation, computation, mathematics and communication" (p. 9) in multiple, authentic ways that allow students to advance their statistical and critical thinking skills.

While the guidelines state that statistical theory is an integral part of an undergraduate statistics curriculum, there is not a consensus on how a modern statistical theory course should be structured. The two-semester calculus-based introductory mathematical statistics sequence often remains a traditional, lecture-intensive course, discussing probability and distribution theory during the first semester and frequentist inferential theory during the second semester. This often entails the theory of point estimation, including sufficiency, the derivation of likelihood ratio and other hypothesis tests, and confidence intervals. Yet, others suggest transforming the course by including non-parametric modeling, decision theory, computing and problem solving; excluding optimal testing; and de-emphasizing asymptotics (JSM 2003 Panel Session, 2003).

Inherent to these discussions lies the overarching principle that understanding the theory enables statistics majors to see statistics as more than a set of fixed methods; equipped with this theoretical toolset, statisticians can continue to change and revolutionize current statistical methods and practice (Cobb, 2011). However, it can be challenging to structure learning opportunities that help students learn and explain the "interplay between mathematical derivations and statistical applications" (ASA, 2014, p. 12), particularly in a traditional, lecture-intensive mathematical statistics course. In addition to learning statistical theory, students need opportunities to "develop flexible problem-solving skills;

. . . [synthesize] theory, methods, computation, and applications; . . . work in teams; . . . [and] refine communication skills" (ASA, 2014, p. 13).

Within the undergraduate mathematical statistics course, instructors have proposed a variety of approaches for creating these types of learning opportunities. Such efforts include the use of in-depth case studies and statistical software to integrate statistical theory and practice (Nolan and Speed, 1999, 2000); the use of statistical software, such as R, to enhance student understanding of concepts (Buttrey et al., 2001; Horton et al., 2004; Nolan and Temple Lang, 2003); and the use of student-centered problem-based learning approaches to develop students' statistical thinking and problem-solving skills and foster collaborative learning (Bates Prins, 2009; Horton, 2013). To date, these efforts primarily focus on the second semester of the mathematical statistics sequence and often require substantial changes to curriculum (e.g., Horton, 2013; Nolan and Speed, 1999). However, such changes may not be feasible if the courses are pre-requisites for professional exams or follow-up coursework in allied/client departments, such as actuarial science.

In this paper, we discuss various approaches, activities and assessments that can be used to foster active learning and emphasize conceptual understanding while still covering the necessary theoretical content students need to be successful in subsequent statistics or actuarial science courses. We begin by describing our various instructional approaches, including the context in which we teach. Then we highlight some of the in-class activities and assessments we've developed and used in both semesters of the calculus-based mathematical statistics sequence. We conclude by sharing student reflections about these course enhancements. The course revision we suggest requires very little change in content, so other mathematical statistics instructors can implement these strategies without sacrificing concepts in probability and inference that are fundamental to the needs of their students. In addition, several of the methods we use are not topic specific and can be easily adapted for other topics in the course. By emphasizing conceptual understanding as opposed to mere memorization and calculation, we encourage students to apply the theoretical concepts to a variety of new settings and contexts, thus strengthening their critical thinking and problem solving skills needed for future academic and real-life problems they might encounter.

2   Approach

In this section, we describe the two-semester calculus-based undergraduate mathematical statistics sequence and the context in which these courses are taught at each campus. We also describe our teaching philosophies and approaches for these courses.

2.1   Course Description

The calculus-based mathematical statistics sequence at Montana State University (MSU) is a set of two 3-credit hour courses, while at the University of Nebraska-Lincoln (UNL), it is a set of two 4-credit hour courses in which three of the weekly contact hours are spent in a "lecture" setting and one takes place in smaller recitation sections. At both institutions, the courses simultaneously serve undergraduate and graduate students from varying disciplines, with some classes composed primarily of students majoring in mathematical sciences and others composed primarily of students from other disciplines, including actuarial science, engineering, economics and computer science. The course content at both institutions is similar, encompassing traditional topics about distribution theory and statistical inference as well as more modern topics such as simulation and bootstrap methods. These align with the probability and statistical theory topics (http://www.amstat.org/education/pdfs/GuidelineTopics.docx) suggested by the *2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science* (ASA, 2014). The required text is *Mathematical Statistics with Applications* (Wackerly et al., 2008), but the treatment and order of most topics is similar to that of Casella and Berger (2002) in *Statistical Inference.*

In 15 years of combined experience teaching the mathematical statistics sequence, we have taught undergraduate and graduate versions of the course sequence. We have also taught classes ranging in size from 11 to 84 students per section. In addition, we have taught the courses both with and without students' use of statistical software.

2.2   Methodology

In our courses, we emphasize conceptual understanding as students apply the concepts to a variety of new settings and contexts, thus strengthening their critical thinking (Wong, 2007)

and problem solving skills (Snyder and Snyder, 2008). We expect them to be curious, ask questions, seek opportunities to learn, and be open and responsive to constructive feedback. These skills, which align with several of the statistical practice and problem solving topics suggested by the *2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science* (ASA, 2014), are important for thinking and reasoning statistically (GAISE, 2005) and will extend to future academic and real-life problems students might encounter.

We use a variety of instructional strategies that allow students to actively engage with the material and think critically about what they are learning (Bonwell and Eison, 1991; Freeman et al., 2014). Along with traditional discussion and individual practice, writing plays a key role. We use writing, in the form of a short answer exam response or a written report, to assess understanding. We also use writing to promote student learning (Langer and Applebee, 1987; Newell, 1984). For instance, at the end of class, students turn in a summary of the key idea they took away from the lecture that day and/or any questions they still have about what we discussed (Bean, 2001; Singleton and Newman, 2009). Students also write brief responses to prompts such as the following:

> *We've spent time finding the distribution of an order statistic. Give at least two examples of how we could use the distribution of an order statistic.*

Such open-ended writing encourages students to think about their learning and how it applies to other topics without the fear of being assessed (Elbow, 1997; Murray, 1978). It also motivates whole class and small group discussions, where students are given the opportunity to share what they've written and justify their responses (Bean, 2001).

In addition, we use collaborative and cooperative learning techniques in which groups of students help one another learn and apply statistical concepts (Bruffee, 1995; Johnson and Johnson, 1999; Roseth et al., 2008; Slavin, 2012). Such collaborations primarily occur through group activities and peer review. These instructional strategies allow students to explore alternative approaches for answering questions, identify key concepts, clarify misunderstandings and develop a deeper conceptual understanding of the material (Mulder et al., 2014; Roseth et al., 2008). They also foster reflective learning (Dochy et al., 1999) and enhance students' statistical reasoning and critical thinking skills (Roseth et al., 2008) by giving students practice evaluating their own work and the work of others, articulating their

thought processes and justifying the approaches they used while simultaneously learning from others.

Participation, homework, traditional exams and projects are major components of a student's overall course grade. We let students know that exams will evaluate their *understanding* of the material, as well as their ability to *synthesize* and *transfer* that knowledge to other scenarios and situations; questions assess conceptual understanding as opposed to mere memorization.

In the next two sections, we describe specific examples of activities and assessments we've created and used in the mathematical statistics courses. These examples were chosen to help illustrate our methodology and how we try to align activities and assessments with the type of learning we value most in the course.

3 Activities

During class, we use a variety of activities that typically involve a combination of small group work and class discussion. These activities allow students to explore and discuss challenging topics collaboratively, with the intention of developing statistical thinking and conceptual understanding as opposed to merely practicing mechanics (Roseth et al., 2008). To encourage open exploration and discussion, students receive little to no credit for these activities, with grading based solely on completion. In the following sections, we discuss several of the activities we use to help teach discrete random variables, point estimation and sufficiency.

3.1 Discrete Probability Distributions

This activity was adapted from the *Can You "Beat" Randomness* lesson intended to introduce beginning statistics students to the concept of randomness (Zieffler and Catalysts for Change, 2013) and takes about 30 minutes of class time. In the original lesson, students play an applet-based game in which they try to guess, after observing a sequence of red and green lights, whether subsequent lights will be red or green. After playing the game, students are told that the probabilities used to generate the light colors were 3/4 for green

and 1/4 for red. They are then asked to compare two different guessing strategies:

- **Strategy A:** Guess green 3/4 of the time and red 1/4 of the time, trying to choose "randomly" with these probabilities each time.

- **Strategy B:** Guess green every time.

Using simulation, students determine whether Strategy A is better than Strategy B (Zieffler and Catalysts for Change, 2013).

We use an adapted version of this activity after we have introduced the binomial distribution, but before we have discussed the geometric and negative binomial distributions. Similar to the *Can You "Beat" Randomness* lesson, students play the game and then discuss the various strategies they used. After we present the two guessing strategies, we extend the activity by having groups of 2-3 students calculate, rather than simulate, the expected number of correct guesses out of 50 for each strategy. By actually calculating these expected values by hand, students get to review probability concepts when determining the theoretical probabilities of success. This then leads to a class discussion about which strategy they think is "better."

We continue with another extension of the activity by asking students to answer the following questions:

- Suppose you are interested in the random variable that gives the number of trials needed until the first correct guess is made. Using what you know about the binomial probability mass function (pmf), what do you think the pmf for this random variable will be? Explain your reasoning.

- Suppose you are interested in the random variable that gives the number of trials needed until the 20th correct guess is made. Using what you know about the binomial pmf, what do you think the pmf for this random variable will be? Explain your reasoning.

Through discussion, students are able to "derive" the pmfs of the geometric and negative binomial distributions. This helps impress upon the students that pmfs are not formulae to be memorized, but logically arise from the definition of the random variable. By using

statistical software to simulate Bernoulli random variables until the first or 20th success, students can also explore the geometric and negative binomial distributions empirically. The R script we have used in class for this simulation may be found in the Supplementary Materials.

3.2   Maximum Likelihood Estimation

We employ two activities when covering maximum likelihood estimation: one at the beginning of the unit to introduce the concept and another at the end of the unit to both summarize the process and provide a more complicated example.

3.2.1   Introducing Maximum Likelihood Estimation: Capture-Recapture

The Capture-Recapture activity serves as an introduction to maximum likelihood estimation and assumes students have already been introduced to the hypergeometric distribution and/or counting. It requires approximately 40 minutes of class time, and students may complete it individually or in groups with up to five students each. This activity is adapted from one for introductory statistics students, and the details may be found in Scheaffer et al. (2004).

For this activity, students receive a small pond (Ziploc bag) of cheddar goldfish and use capture-recapture sampling techniques to estimate the total number of fish in each pond. They assume the same number of fish are in each pond. A handout (see Supplementary Materials) guides them through the process of sampling fish with a net (3 oz paper cup) and "tagging" captured cheddar goldfish by replacing them with the same number of pretzel goldfish. Students then mix the cheddar and pretzel goldfish together and take a second sample with the same net. The number of tagged (pretzel) fish, as well as the total number of fish in the second sample are recorded, and each group uses the results to estimate the population size, N.

After students have obtained their estimates, we discuss their estimation strategies and use the hypergeometric distribution to motivate maximum likelihood estimation (see Supplementary Materials for a video describing this process). Then the instructor inputs each group's information into an R or SAS program file (see Supplementary Materials) to
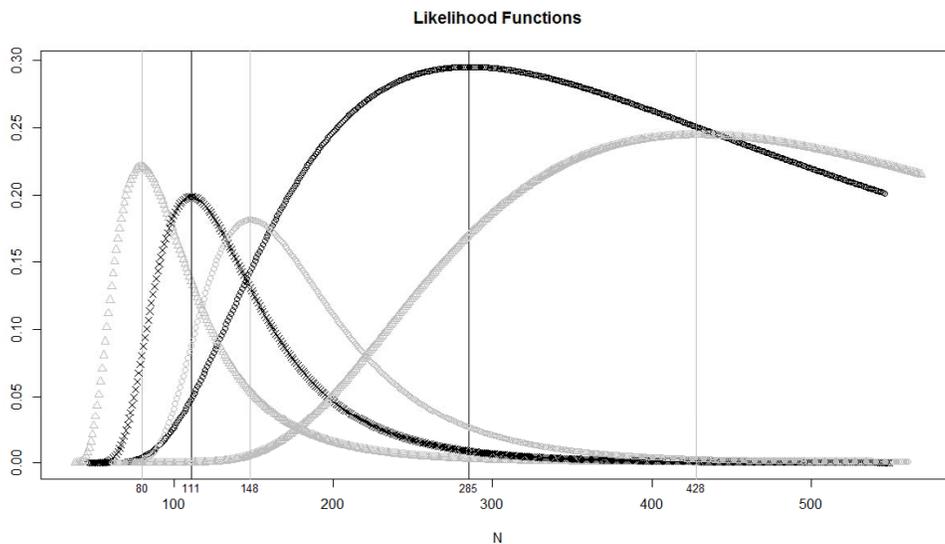
Likelihood Functions

Figure 1: Five likelihood functions, each resulting from the sample data obtained by one of five groups of students in a class that completed the Capture-Recapture activity.

graph the resulting likelihood functions. The class discusses the graph (e.g., see Figure 1), and the instructor uses this shared information to transition to a more formal discussion on how to find maximum likelihood estimators (MLEs).

3.2.2 Summarizing Maximum Likelihood Estimation: Spies versus Agents

The Spies versus Agents activity was adapted from one published by Fewster (2014) that focuses on finding an MLE for a mixture proportion. During the activity, students are given secret instructions, which vary depending on whether they have been randomly selected to be a "spy" or an "agent." The instructions have them roll a die to determine whether or not a series of "missions" are successful; spies and agents have different given probabilities of success. The results of the individual missions are pooled in a common envelope across a team of 10-12 students. The parameter of interest is the proportion of spies in the team. Specific details on the activity may be found in Fewster (2014), and it takes approximately one 75-minute class period to complete.

Because agents and spies have different probabilities of success, the teams are faced with estimating a mixture proportion, which is a more difficult estimation task than they have
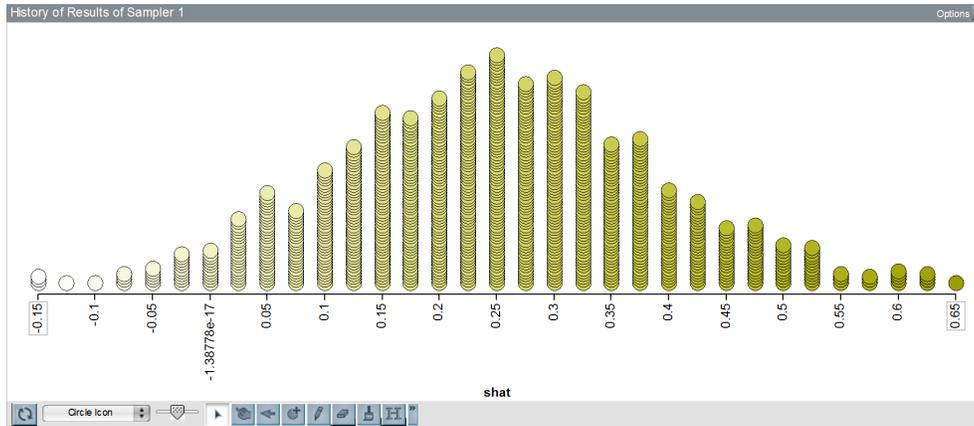
9

Figure 2: 1000 simulated ML estimates of the spies and agents mixture proportion.

previously encountered. First, they must determine the relationship between the parameter of interest and the given probabilities of success for both agents and spies. In addition, they must realize that the way in which they define their random variable(s), either as a single binomial with multiple trials pooled over their entire team or as independent Bernoullis representing each individual die roll, will determine the appropriate likelihood function.

This activity yields several teachable moments. As an activity wrap-up, students are asked to reflect on the purpose of the activity. Most students reference the multiple potential ways to represent the random variable, seeing this as an important concept. Some students, however, note the discrepancy between their ML estimate and the true mixture proportion of 0.25, and comment that perhaps a different estimator should be used. This leads to a simulation study, so that we can, as a class, explore the behavior of the MLE. The simulated sampling distribution is displayed in Figure 2. This particular empirical sampling distribution elicits a productive discussion about bias of estimators, as well as potential invalid estimates outside the parameter space. In addition, students can see the variability associated with the estimator and realize (perhaps for the first time) that estimates aren't typically "right" or necessarily even "close."

3.3 Conceptualizing Point Estimation

The Conceptualizing Estimation activity is used as practice for finding MLEs and method of moments estimators. Students work in pairs, and the activity takes approximately 45

minutes of class time. A pair of students is presented with a distribution and is asked to find the MLE for each unknown parameter. At the same time, a partner pair is presented with the same distribution but is asked to find a method of moments estimator. For each problem, students draw a line down the middle of their papers to make two separate columns. In the first column, the students complete the work needed to find the appropriate estimator. In the second column, the students use words to explain what they are doing and why they are doing it. This helps emphasize concepts over process. After they have found the estimator, they explain their thought process to the partner pair and answer any questions they have about the solution. This ensures that the students must engage with the problem in three different ways, by (1) practicing mechanics, (2) translating those mechanics into words that describe the process and (3) effectively communicating that process to others.

After completing this activity, some students begin explaining their work on subsequent homework and exam problems. We have observed this is most likely to occur with lower-performing students. Sometimes this "self-coaching" helps them successfully solve a problem. In instances where students are not successful, this type of articulation of their thought process helps us better diagnose the conceptual stumbling blocks and/or recognize when their difficulty is with the underlying mathematics (e.g., taking derivatives), not the statistical theory.

While the approach works well when finding estimators, it can also be used with any other process in which students would benefit from explaining their reasoning and understanding for each step of a potentially formulaic process.

3.4 Sufficiency

The Cell Phone Battery Life activity is used as an introduction to the concept of sufficiency. Prior to this, students have learned how to find MLEs and method of moments estimators. In addition, they have seen that best unbiased estimators are defined to have minimum variance among all other unbiased estimators. However, students have not yet discussed what it means for a statistic to be sufficient. Instead, sufficiency is introduced through the following group activity.

Table 1: The six scenarios provided to students for the Cell Phone Battery Life activity.

Statistic Provided	Distribution 1 $f_X(x|\theta_1) = \frac{1}{75}\exp\left(\frac{-(x-\theta_1)}{75}\right), \quad x > \theta_1$	Distribution 2 $f_X(x|\theta_2) = \frac{1}{\theta_2}\exp\left(\frac{-x}{\theta_2}\right), \quad x > 0$
A. Entire data set $(n = 25)$	$\{21, 23, \ldots, 181, 203\}$	$\{1, 3, \ldots, 161, 183\}$
B. Sample minimum	$x_{(1)} = 21$	$x_{(1)} = 1$
C. Sample mean	$\bar{x} = 92.48$	$\bar{x} = 72.48$

To prepare for this activity, we create six different scenarios for modeling the lifetime, in hours, of a cell phone battery. Each scenario provides one distribution and one statistic, as displayed in Table 1. The numeric values for each statistic are obtained by first simulating data for the second distribution ($\theta_2 = 75$) and then shifting the values $\theta_1 = 20$ hours to obtain simulated values for the first distribution. This creates six possible scenarios, each consisting of one of the following distributions: (1) displaced exponential with mean $= 75 + \theta_1$ or (2) exponential with mean $= \theta_2$, along with exactly one of the following statistics: (A) the entire data set, (B) the sample minimum or (C) the sample mean.

During class, each group is given one of the six scenarios as well as the sample size, $n = 25$, and instructed to estimate the unknown parameter (θ_1 or θ_2) using the information provided. (Ideally, there should be at least 12 groups of two to four students so that each scenario is given to two or more different groups. If this is not possible, groups that finish early can be given another scenario and asked to complete the activity again.) Even though we tell students there is no right answer, most automatically want to find a method

Table 2: Sample class results for the Cell Phone Battery Life activity.

Scenario Provided	Parameter	Group Estimates
1A - Data	$\theta_1 = 20$	21, 21, 21
1B - Min	$\theta_1 = 20$	21, 21, 21
1C - Mean	$\theta_1 = 20$	15.71, 92.48, ?
2A - Data	$\theta_2 = 75$	72.48, 72.48, 72.48
2B - Min	$\theta_2 = 75$	25, 18, 1
2C - Mean	$\theta_2 = 75$	72.48, 72.48, 72.48

of moments estimator or MLE. Consequently, groups that receive a sufficient statistic for their unknown distributional parameter have no problem calculating their estimates and typically provide similar values (e.g., Scenarios 1A, 1B, 2A and 2C in Table 2). Other groups, however, often become frustrated by the limited amount of information they receive and use a variety of different strategies to come up with an estimate. As shown in Table 2, this can lead to varying estimates among groups who receive the same scenario. In one class, groups that received the exponential distribution with mean $= \theta_2$ along with the observed sample minimum, $x_{(1)} = 1$, (i.e., Scenario 2B) produced three different estimates of the unknown parameter. To get the first estimate of 25, one group calculated the expected value of the sample minimum, $E(X_{(1)}) = \theta_2/n$, and set it equal to the observed sample minimum. For the same scenario, another group made an "educated" guess of 18, while another just wrote the value of the statistic received, $x_{(1)} = 1$. This same type of variability also occurs for Scenario 1C, providing valuable opportunities for discussing the concept of sufficiency.

Within 5-10 minutes most groups have generated their estimates and written them on the board. The class then takes an additional 5-10 minutes to compare the group estimates and discuss the different estimation strategies and struggles they encountered. At this point, they are instructed to note key observations. Using the results (e.g., see Table 2), students identify that some information consistently leads to the same estimate of the unknown parameter, while other information does not. Additionally, they recognize

that the usefulness of a statistic is dependent upon the parameter they are estimating. For example, the sample mean is more useful for estimating a population mean than it is for estimating a population minimum. These key concepts are important for understanding sufficiency and its applications to estimation.

4 Assessments

When we teach the mathematical statistics sequence, we use multiple types of assessments including low- and high-stakes writing assignments, as well as in-class exams. We discuss each of these types of assessments and provide examples for how we use them to stress conceptual understanding and foster active learning.

4.1 Low-Stakes Writing Assignments

Low-stakes writing, also called exploratory writing, is *writing to learn*, as opposed to *writing to demonstrate learning*. According to Bean (2001), "the writing process itself provides one of the best ways to help students learn the active, dialogic thinking skills" (p. 19). Depending on an assignment's learning objectives, instructors may grade for accuracy and/or completion. In Section 2.2 we describe a few short, in-class low-stakes writing assignments we have used to summarize a concept or focus a discussion. Here, we describe an assignment to introduce hypothesis testing and errors. For this assignment, we instruct students to select and read one of the following articles:

- Erotokritou-Mulligan, I., P. Sonksen, and R. Holt (2011). Beyond reasonable doubt: catching the drug cheats at the London Olympics. *Significance*, *8*(1), 5-9.

- Gastwirth, J. and W. Johnson (2011). Dare you buy a Henry Moore on eBay? *Significance*, *8*(1), 10-14.

- Wainer, H. and R. Feinberg (2015). For want of a nail: Why unnecessarily long tests may be impeding the progress of Western civilisation. *Significance*, *12*(1), 16-21.

Because *Significance* is geared toward a lay audience, these articles serve as gentle introductions to the logic of hypothesis testing and errors. In addition, by providing students a

series of critical reading questions (see Supplementary Materials) to answer while they read their chosen article, we guide students through the readings and encourage them to reflect on important concepts. This part of the assessment is typically assigned as homework, with the assurance to students that they will be graded on the thoughtfulness and completeness, as opposed to the accuracy, of their answers.

During class, students discuss their readings and answers within groups, focusing specifically on the questions addressing errors and any other questions that were challenging to answer. Groups are instructed to prepare an overall summary of the articles read and discuss hypotheses and errors within the articles' respective contexts. After approximately 10-20 minutes, one group for each article is selected to share their summary with the rest of the class. By sharing these brief summaries, students are able to learn about the other two articles they did not read, enabling them to participate in conversations about how the different concepts (e.g., hypotheses, errors, etc.) apply to each of the three situations. Because the articles address topics most students find interesting, this can lead to lively, yet productive, discussions about the statistical concepts and their applications. Through these discussions, instructors can quickly gauge student understanding and answer questions without having to read each student's set of responses. In addition, this activity serves as a nice introduction to hypothesis testing, allowing students to learn about hypothesis testing and errors on their own and providing instructors the opportunity to assess current student understanding of the relevant topics.

4.2   High-Stakes Writing Assignments

Unlike low-stakes assignments, high-stakes writing assignments are graded for accuracy and represent a substantial portion of a student's overall grade. While these formal writing assignments may be designed to support learning (Bean, 2001), they are intended primarily for evaluating understanding. In the following sections, we describe two formal writing assignments used to assess students' understanding of probability distributions and concepts related to estimation and hypothesis testing.

4.2.1 Writing Distribution Questions

We use the Writing Distribution Questions assignment to assess student understanding of probability distributions. This assessment has three different parts, each of which takes approximately one week to complete. During the unit on named distributions, each group of students is assigned two distributions (one discrete, one continuous) and instructed to create at least one original question and solution key for each. For example, one group could be assigned to work with a Poisson distribution and an Exponential distribution.

One week later, two groups trade and review rough drafts of their questions and solutions via a course management system such as D2L or Blackboard, or in person during recitation. During the third week, each group then uses their peers' feedback and recommendations to edit their work before uploading their final products to an electronic dropbox. Throughout this process, students are aware their final questions and solutions will be run through plagarism software, such as *TurnItIn*, that will identify what portion of the work has been copied, along with the original source(s).

The discrete distribution question and solution set constitutes 40% of the overall grade for this project and is evaluated using the following criteria:

- Problem requires the use of the discrete distribution assigned and demonstrates the group understands the statistical concepts involved

- Problem is framed within a content-appropriate real-life (or imaginary) context or scenario and is logical, clear and well-written

- Correct solution is provided with thorough explanations and work

The continuous distribution question and solution set is scored similarly. The remaining 20% of the overall grade is based on the peer review component. This encourages students to provide thoughtful, thorough and clearly stated feedback that addresses the relevant criteria outlined in the grading rubric and to invest an adequate amount of effort in reviewing the set of questions and solutions.

The instructor may post the final questions for other groups to view and study; depending on the questions' originality and intellectual merit, some may be selected or adapted for

the exam. Some examples of student-generated scenarios, questions and solutions include the following:

Example 1:

Question: Emails arrive in a student's inbox at a rate of 3 per hour. One-third of all emails received are non-spam emails. Find the total expected number of emails received in two hours and the probability that the student receives 2 or more non-spam emails in two hours.

Solution: If t represents time in hours, then the email arrival rate is $\lambda = 3t$. Let p be the probability of a spam email, so $p = 2/3$. Using this information, the arrival rate for spam emails is $\lambda_S = \lambda p = 3tp = 2t$ and the arrival rate for non-spam emails is $\lambda_N = \lambda(1 - p) = 3t(1 - p) = t$. So, if we define the random variable Y = number of non-spam emails in $t = 2$ hours, then $\lambda_N = 2$, and $Y \sim$ Poisson(2). Then:

$$
\begin{aligned}
P(Y \geq 2) &= 1 - P(Y < 2) = 1 - (e^{-2} + 2e^{-2}) = 0.59399, \text{ and} \\
E(Y) &= \lambda_N = 2.
\end{aligned}
$$

Example 2:

Question: A cellular phone company accepts trade-ins when customers upgrade to a new device. After the phones are traded in, they are sent to a distribution center and distributed to various stores. 1000 iPhones were traded in and sent to the distribution center. Demand for pre-owned iPhones is soaring, and so the company doesn't check to see whether the iPhones are functioning. It turns out that 50 of the 1000 iPhones are malfunctioning. A store receives a shipment from the distribution center of 25 pre-owned iPhones. What is the probability that the store receives more than 1 malfunctioning iPhone?

Solution: $N = 1000$, $r = 50$, $n = 25$, X = number of malfunctioning iPhones received by the store

$$
\begin{aligned}
P(X > 1) &= 1 - P(X = 1) - P(X = 0) \\
&= 1 - \frac{\binom{50}{1}\binom{1000-50}{25-1}}{\binom{1000}{25}} - \frac{\binom{50}{0}\binom{1000-50}{25-0}}{\binom{1000}{25}} \\
&= 1 - 0.3685 - 0.2730 = 0.3585.
\end{aligned}
$$

By completing this assignment and incorporating the peer review process, students reevaluate their understanding of probability distributions and consider the various distributional assumptions and real-life (or imaginary) situations that can be modeled. Because the assessment occurs concurrently with instruction, the instructor can gauge student understanding and comprehension of the material, adapting instruction as needed. In addition, students (as well as the instructor) receive a repository of extra problems to use for practice and/or further assessments.

4.2.2 Meaningful Paragraph

At one or two times during the second semester of the course, students complete a short writing assignment that gives them the opportunity to write about concepts from the course. The structure of the Meaningful Paragraph assigment is attributed to entomologist Elaine Backus (Jordan, 2008). This assessment may be completed individually or in pairs, and the final product should be a continuous piece of writing that uses all of the words in a list and has sentences that "make sense and hang together". In the past, lists of terms have focused on either estimation or hypothesis testing. The terms provided for each have included:

- Estimation: Estimator, Parameter, Estimate (noun), Random Variable, Random Sample, Bias, Variance, Sufficient Statistic, Best Unbiased Estimator

- Hypothesis Testing: Size, Power, Power Function, Null Hypothesis, Alternative Hypothesis, Best Test, P-value

The ideas in the paragraph should illustrate that the student understands the terms in a way that allows him/her to write "meaningfully" about them. The sentences need to demonstrate the relationships between the terms, and the paragraph must be framed within

a content-appropriate real-life (or imaginary) context or scenario. The meaningful paragraphs are graded using a rubric (see Supplementary Materials) that assesses the writing for both knowledge of the terms and fluency of use, as demonstrated by the organization of the paragraph and the originality of the context or scenario. Bonus points are awarded to paragraphs that show evidence of extreme innovation.

While students are required to have a written piece, they may also submit additional materials (e.g., audio, video, etc.). For example, during the 2014 spring semester, one pair of students created a satirical news website (http://gemarsa1.wix.com/dailyestnebraskan), similar to *The Onion*. On the website they included articles, such as "Statistics students finally discover Google, Yahoo! answers for homework help," which demonstrated that, by using the estimation terms in a humorous way, they understood the concepts in enough depth to employ them in puns and jokes. Other examples of exemplary paragraphs include comic books, student-performed raps (see Supplementary Materials) and children's storybooks. As evidenced by the rap, the students are demonstrating both knowledge of the terms and a fluency of use that testifies to a deeper understanding of the concepts than can be established by traditional problems.

While we have not used this assignment in the first semester, it could be easily adapted to those earlier concepts. For example, asking students to write about the terms

- Random, Probability, Independence, Random Varible, Probability Density Function (pdf), Cumulative Distribution Function (CDF), Moment, Conditional Distribution

could be used to assess their understanding of fundamental probability theory concepts.

4.3   Exams

While we have incorporated both low- and high-stakes writing assignments, pencil-and-paper exams still play an important role in our evaluation of student understanding. Exams are a mix of "conceptual" and "traditional" questions. The traditional questions tend to assess students' procedural fluency, while the conceptual questions aim to assess students' ability to apply their understanding of the concepts in new ways. For example, an exam question used recently in the first semester of the course sequence was:

The Heisenberg Auto Dealership has several salesman, two of whom are Walt and Jesse. Let X denote the proportion of cars sold by Walt and let Y denote the proportion of cars sold by Jesse. The joint pdf of X and Y is given by

$$f_{XY}(x, y) = 6x.$$

Circle any supports below that you believe would be reasonable for this situation.

$$x > 0, y > 0 \qquad 0 < x < y \qquad 0 < y < x \qquad x > 0 \qquad \text{none of these}$$

Explain your reasoning in words. Any answers relying on calculations will automatically earn a score of 0.

The scenario for this question was adapted from one submitted by a student group (see Section 4.2.1). The goal of the question is to assess whether students can think critically about the assumptions imposed by random variables representing proportions. The grading focuses on the students' explanations, rather than their choice of support. If the students select anything other than "none of these," but adequately explain the correct requirements for the support, they receive full credit. For example, it was not unusual for students to select the support "$x > 0, y > 0$" and add the additional requirement $0 < x + y < 1$ in their explanations.

The most common mistake students have made on the question is not recognizing that the supports need an upper bound, as well as a lower bound. Several students recognize that there is dependence between X and Y, but miss that the dependence cannot be completely represented by either $0 < x < y$ or $0 < y < x$. Additionally, some students choose the distractor support $x > 0$, citing the absence of y from the joint density function.

During the second semester of the course sequence, we continue to use both traditional and conceptual questions on exams. The following is an example of both types of questions:

A vice president in charge of sales for a large company claims that the average number of sales contacts per week has decreased from what it was 10 years ago. To check this claim, 5 salespeople are selected at random, and the number of contacts made by each is recorded for a single randomly selected week. To
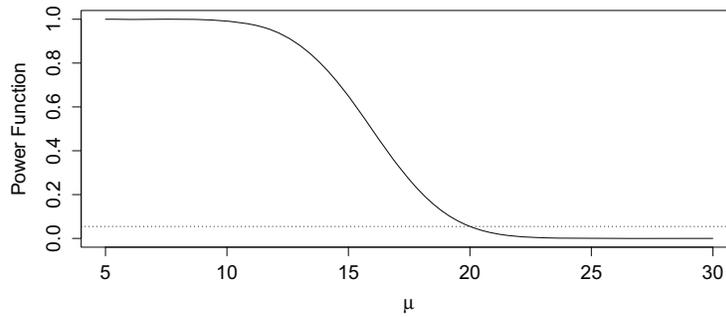
Figure 3: Power function for example test question.

test her claim, the vice president is presented with the following hypothesis test function:

$$\xi_M(X_1, \ldots, X_n) = \begin{cases} 1 & X_{(n)} < c \\ 0 & \text{elsewhere} \end{cases}$$

The plot (see Figure 3) of the power function for ξ_M was produced, and a horizontal line was drawn at the significance level $\alpha = 0.0546$.

a) Explain why a Poisson distribution with mean μ would be a reasonable model for each of the random variables $X_1, \ldots, X_n$. Your explanation should address the assumptions of the Poisson distribution. Your explanation should also include **in words** what n, $X_1, \ldots, X_n$, μ and $x_1, \ldots, x_n$ are in this situation, and, if applicable, give the numerical values of each.

b) Based on the power curve, what are the hypotheses the vice president wants to test? Be sure to include any appropriate numerical values and use words to describe any symbols used.

c) The vice president was presented with another hypothesis test function:

$$\xi_S(X_1, \ldots, X_n) = \begin{cases} 1 & \sum_{i=1}^{n} X_i < k \\ 0 & \text{elsewhere} \end{cases}$$

- **On the plot provided**, sketch what you think the power function of this hypothesis test would look like in relation to the power function of

21

ξ_M. Assume approximately the same significance level is used. (Note: Power functions must be drawn clearly on the plot provided above in order to receive credit.)

- **In words**, explain thoroughly **why** the power function for ξ_S should look similar to what you sketched. A complete explanation should also include rationale for the relative positions chosen for the power functions over the **entire** plot.

Part a) is an example of a traditional question, where students are expected to explain why a Poisson probability distribution would reasonably model each of the random variables, $X_1, \ldots, X_n$. In addition, students must identify and define what n, $X_1, \ldots, X_n$, μ and $x_1, \ldots, x_n$ are for this scenario. While this is a review of material covered during the first semester of the course sequence, most students are able to answer this question correctly.

The remaining parts to this problem are examples of conceptual questions. Part b) requires that students apply their knowledge of power functions and significance levels in a new way. Instead of working with specified hypotheses, students must use the definitions of significance levels, power functions and test functions in order to provide the hypotheses of interest. In the past, students have struggled with identifying the null value of 20 and specifying the correct inequality in the alternative hypothesis.

Part c) builds on these concepts, as well as the concept of sufficiency. This question has students compare two hypothesis test functions, one in which the test statistic is a complete, sufficient statistic for the parameter of interest and one in which the test statistic is not. Because this question is used after the class has discussed estimation and hypothesis testing, but not optimal testing, students must translate the concepts underlying best estimation to a testing situation in order to identify which test should be "more powerful." After they've identified and described why the test they've identified is more powerful, they then have to sketch the power function for that test. While they are not asked to find the exact power function for this new hypothesis test function, students do need to know what the power function of a more powerful test looks like relative to one that is not as powerful at the same significance level.

While a majority of students are able to identify the test that should be more powerful than the other, they struggle with identifying how that translates to a power function. Many students draw relatively inaccurate power functions, where the two functions intersect at values other than the null value of 20 or do not intersect at all. Others flip the relative positions of the two functions, having the power function of the test they've identified as more powerful lie below the power function of the other test over the alternative parameter space, but above it over the null parameter space. The ensuing descriptions explaining their rationale for the relative positions chosen vary in quality and offer valuable insight about their logic. Students who are able to sketch the power function correctly do not always do so for the appropriate reasons. Therefore, this type of question helps reveal students' thought processes when faced with a problem that assesses their conceptual understanding, as opposed to their rote understanding, of various statistical topics.

5   Reflection

We think the revisions we have made to the mathematical statistics sequence have been valuable, both in terms of student understanding and in our ability to gauge that understanding. We are also interested in whether or not students see value in the course modifications. To explore students' perceptions of the mathematical statistics sequence, we recently began asking them to write a "letter of learning" at the end of each semester. In this letter, students reflect on and write about their experiences as learners in the course. Specifically, we have them write a narrative that describes how the work they have done has or has not supported the learning that is most important to them. For example, we ask students to share how the course expectations and activities have impacted their learning. In addition, we have them reflect on their engagement in the course and how the intellectual community of the class influenced their learning.

For this paper, we reviewed students' letters of learning and course evaluations from the most recent offering of the year-long sequence of courses. We found that students recognize that the approaches, activities and assessments we use are unconventional for a traditional mathematical statistics class. Some students resist such changes, claiming to prefer methods that essentially emphasize memorization and calculation. For example,

one student states, "[C]reative writing assignments . . . do not belong in a 400-level course. Assigned group work should not take place in a math course."

However, a majority of students were positive when referencing our teaching approaches for the course, noting that they appreciate actively engaging with their peers in order to obtain a deeper understanding of the concepts presented. One student wrote, "[T]he format of the class allows for everyone to contribute to class discussions and emphasizes collaborative understanding, which I really like. So I feel that I was able to help my peers learn and vice versa through the discussion in the classroom." Another student recognized the value of a collaborative learning environment, specifically referencing peer review: "Peer review of work is undeniably valuable as mistakes are made that cannot be picked up by one's own eye and of course many concepts are beneficial to hear from a third party for clarification."

Students also commented on how the activities aided in their learning. One student observed, "The in class learning activities . . . were fun and I thought they were more productive than just lecturing would have been." Another student noted, "The group activities were helpful, I especially found the sufficiency activity useful in explaining the concept of sufficiency in a new way that clicked for me."

The students also shared how the various assessments, such as the Writing Distribution Questions project, supported their learning. For example one student reflected, "Determining a problem given the constraints in a teaching environment really causes one to flex their . . . thinking-muscle. I found myself thinking of 'probability' questions all the time." Another student saw benefit in encountering challenging exam questions that require students to form connections and apply concepts in new ways:

> [E]xam questions always got me thinking hard as well. In particular the question on the second take home exam. . . really forced me to think hard to apply the material we learned in class but I also had to go back and review most of the topics we had covered last semester. . . Seeing how all the ideas tied together and related to one another also was really fascinating and beneficial because [it] put the entire course sequence into perspective.

Together, these approaches, activities and assessments are meant to help students rec-

ognize and experience the theoretical relationship between mathematics and statistics, strengthening their critical thinking and problem solving skills for future work in statistics. One student aptly summarized the broader goals of the course:

> *In my opinion this was the main concept of the class . . . , using statistics in ways that are different and informative. Obviously we learned a lot of definitions and formulas but I thought the main idea of the class [was] how to manipulate and use these concepts on problems that were not simply 'find the answer' type questions. In addition this will definitely benefit me in the future as it will enable me to apply statistics to any number of problems rather than simply plug and chug.*

With the course revisions we made, students are required to synthesize various mathematical concepts and apply them to new situations when solving difficult problems. Developing such problem solving skills is a different, but important, focus for a mathematical statistics course. Overall, students appreciated the opportunity to learn in a collaborative environment where they actively engage with and apply concepts in different contexts to deepen their understanding of statistics.

6 Conclusions

In this paper, we discussed various instructional approaches, activities and assessments that can be used to foster active learning and emphasize conceptual understanding in the undergraduate mathematical statistics sequence. As evidenced by the student comments, the course enhancements encourage students to see the connections between the concepts and not view them simply as disparate mathematical facts. In addition, the majority of students are now better able to see mathematical statistics as a *statistics* course, rather than a mathematics course. Yet, because the theoretical content of probability and inference is still covered, the students remain well prepared for follow-up statistics and actuarial science courses; many of our students take the Society of Actuaries Exam P (Probability) either concurrently with or subsequently to the mathematical statistics sequence. Furthermore, several of the approaches, activities and assessments we described are not topic specific and

can be easily adapted for other material in the course, not just the concepts provided in our examples. For instance, we have found peer review to be a valuable addition to many courses, ranging from introductory statistics to graduate-level experimental design.

Innovation in the introductory statistics class, instigated by the GAISE College Report (2005), has improved the teaching and learning of introductory statistics at our institutions and throughout the country. We have shown that similar reform can be implemented in the most advanced undergraduate courses as well, even when the constraints of the course require more traditional curriculum. However, additional reforms are needed as the field of statistics continues to evolve. With growing enrollments in undergraduate and graduate programs of statistics and data science, the mathematical statistics curriculum will need to move beyond traditional topics and skills. To remain current, the course will need to help students develop the skills necessary to be successful in a data-centric world. While the course revisions we proposed are meant to help students develop critical thinking and problem solving skills, we anticipate additional revisions. Some examples include students active use of software to solve problems, incorporation of current topics, such as Bayes and simulation methods, and explicit connection-building across other theoretical and applied coursework. With these changes, we envision the primary challenge will be learning how to address the ensuing tensions arising from trying to make such changes while still meeting the needs of client disciplines.

SUPPLEMENTARY MATERIAL

R program for discrete probability distributions: The R file the instructors use to simulate empirical distributions for the discrete probability distributions activity. (.R file)

Capture-Recapture class handout: The student handout used as part of the capture-recapture in-class activity. (.pdf file)

Capture-Recapture video: A video demonstrating how the capture-recapture student handout is used as part of an interactive lecture. (.mp4 file)

R program for Capture-Recapture: The R file the instructors use to produce the graph of the class likelihood functions. (.R file)

SAS program for Capture-Recapture: The SAS file the instructors use to produce the graph of the class likelihood functions. (.sas file)

Article questions: The questions provided to students as part of the article reading and reflection low-stakes writing assignment. The file contains the questions for all articles referenced in Section 4.1.(.pdf file)

Meaningful paragraph rubric: The grading rubric used for the Meaningful Paragraph high-stakes writing assessment. (.pdf file)

STAT Rap: An example of an exemplary Meaningful Paragraph. (.m4a file)

References

ASA (2014). *2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science.* Alexandria, VA: American Statistical Association. `http://www.amstat.org/education/curriculumguidelines.cfm`.

Bates Prins, S. C. (2009). Student-centered instruction in a theoretical statistics course. *Journal of Statistics Education 17*(3). `http://www.amstat.org/publications/jse/v17n3/batesprins.html`.

Bean, J. C. (2001). *Engaging Ideas: The Professor's Guide to Integrating Writing, Critical Thinking, and Active Learning in the Classroom.* San Francisco, CA: Jossey-Bass.

Bonwell, C. C. and J. A. Eison (1991). *Active Learning: Creating Excitement in the Classroom.* Washington, DC: School of Education and Human Development, George Washington University.

Bruffee, K. A. (1995). Sharing our toys: Cooperative learning versus collaborative learning. *Change: The Magazine of Higher Learning 27*(1), 12–18.

Buttrey, S., D. Nolan, and D. Temple Lang (2001). Computing in the mathematical statistics course. In *Proceedings of the 2001 Joint Statistics Meetings*, Alexandria, VA. American Statistical Association. `http://www.amstat.org/sections/srms/proceedings/y2001/proceed/00113.pdf`.

Casella, G. and R. L. Berger (2002). *Statistical Inference, 2nd Edition*. Pacific Grove, CA: Duxbury.

Cobb, G. W. (2011). Teaching statistics: Some important tensions. *Chilean Journal of Statistics 2*(1), 31–62.

Dochy, F., M. Segers, and D. Sluijsmans (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education 24*(3), 331–350.

Elbow, P. (1997). High stakes and low stakes in assigning and responding to writing. *New Directions for Teaching and Learning 1997*(69), 5–13.

Fewster, R. M. (2014). Teaching statistics to real people: Adventures in social stochastics. In K. Makar, B. de Sousa, and R. Gould (Eds.), *Sustainability in Statistics Education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS 9, July 2014), Flagstaff, Arizona, USA.*, Voorburg, The Netherlands. International Statistical Institute. `http://icots.info/9/proceedings/pdfs/ICOTS9_PL3_FEWSTER.pdf`.

Freeman, S., S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences 111*(23), 8410–8415.

GAISE (2005). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report College Report*. American Statistical Association. `http://www.amstat.org/education/gaise/`, last accessed June 15, 2015.

Horton, N. J. (2013). I hear, I forget. I do, I understand: A modified Moore-method mathematical statistics course. *The American Statistician 67*(4), 219–228.

Horton, N. J., E. R. Brown, and L. Qian (2004). Use of R as a toolbox for mathematical statistics exploration. *The American Statistician 58*(4), 343–357.

Johnson, D. W. and R. T. Johnson (1999). Making cooperative learning work. *Theory Into Practice 38*(2), 67–73.

Jordan, J. (2008). Writing assignments in an introductory statistics course. In *CAUSE Teaching and Learning Webinar Series; May 13, 2008.* `https://www.causeweb.org/webinar/teaching/2008-05/`.

JSM 2003 Panel Session (2003). Is the math stat course obsolete? `http://www.amstat.org/sections/educ/MathStatObsolete.pdf`, last accessed June 15, 2015.

Langer, J. A. and A. N. Applebee (1987). *How Writing Shapes Thinking: A Study of Teaching and Learning.* Urbana, IL: National Council of Teachers of English.

Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers (2011). Big data: The next frontier for innovation, competition, and productivity. `http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation`, last accessed June 15, 2015.

Mulder, R., C. Baik, R. Naylor, and J. Pearce (2014). How does student peer review influence perceptions, engagement and academic outcomes? a case study. *Assessment & Evaluation in Higher Education 39*(6), 657–677.

Murray, D. M. (1978). Write before writing. *College Composition and Communication 29*(4), 375–381.

Newell, G. E. (1984). Learning from writing in two content areas: A case study/protocol analysis. *Research in the Teaching of English*, 265–287.

Nolan, D. and T. Speed (2000). *Stat Labs: Mathematical Statistics Through Applications.* New York, NY: Springer-Verlag.

Nolan, D. and T. P. Speed (1999). Teaching statistics theory through applications. *The American Statistician 53*(4), 370–375.

Nolan, D. and D. Temple Lang (2003). Case studies and computing: Broadening the scope of statistical education. In *Proceedings of the 2003 International Statistical Institute*

Meeting. International Statistical Institute. `https://oz.berkeley.edu/users/nolan/Papers/isi03.pdf`.

Roseth, C. J., J. B. Garfield, and D. Ben-Zvi (2008). Collaboration in learning and teaching statistics. *Journal of Statistics Education 16*(1). `http://www.amstat.org/publications/jse/v16n1/roseth.pdf`.

Scheaffer, R. L., A. Watkins, J. Witmer, and M. Gnanadesikan (2004). *Activity-Based Statistics, Student Guide, 2nd Edition*. Emmeryville, CA: Key College Publishing.

Singleton, A. and K. Newman (2009). Empowering students to think deeply, discuss engagingly, and write definitely in the university classroom. *International Journal of Teaching and Learning in Higher Education 20*(2), 247–250.

Slavin, R. E. (2012). Classroom applications of cooperative learning. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major, and H. L. Swanson (Eds.), *APA Educational Psychology Handbook, Vol. 3: Application to Learning and Teaching*, pp. 359–378. Washington, DC: American Psychological Association.

Snyder, L. G. and M. J. Snyder (2008). Teaching critical thinking and problem solving skills. *The Delta Pi Epsilon Journal 50*(2), 90–99.

Wackerly, D. D., W. Mendenhall III, and R. L. Scheaffer (2008). *Mathematical Statistics with Applications, 7th Edition*. Belmont, CA: Brooks/Cole.

Wong, D. (2007). Beyond control and rationality: Dewey, aesthetics, motivation, and educative experiences. *The Teachers College Record 109*(1), 192–220.

Zieffler, A. and Catalysts for Change (2013). *Statistical Thinking: A Simulation Approach to Uncertainty, 2nd Edition*. Minneapolis, MN: Catalyst Press.