Setting the Tone: A Discursive Case Study of Problem-Based Inquiry Learning to Start a Graduate Statistics Course for In-Service Teachers

Lawrence M. Lesser
The University of Texas at El Paso

Kerrie Kephart
University of Washington

Key Words: Counterintuitive; Cognitive conflict; Discourse; First day of class; Simpson's paradox; Representation; Teachers.

Abstract

The first day of a course has great potential to set the tone for the entire course, planting the seeds for habits of mind and questioning and setting in motion expectations for classroom discourse. Rather than let the first meeting contain little besides going over the syllabus, the instructor (Lesser) decided to use two sustained open-ended scenarios to put in place from the start the problem-based inquiry learning approach he wanted to use throughout most of the course. After reviewing the literatures involved, this paper shares a description of the lesson's design and instructional cycle and a discourse analysis of that lesson's implementation. Strategies identified by the case study analysis include varying participation structures, well-crafted problems, and the instructor's role as facilitator and co-learner.

1. Introduction

1.1 Active Learning: Promise and Challenge

Consistent with calls for active learning from influential reports such as Bransford, Brown, and Cocking (1999), the *Guidelines for Assessment and Instruction in Statistics Education (GAISE)*

College Report (ASA, 2010) recommend situating learning activities in authentic, real-world problems and providing opportunities for students to think through such problems, make predictions about them, and discuss their reasoning with peers. The shift in emphasis in such methods away from broad "coverage" of material through instructor-centered lecturing and toward more linguistically-based learning processes, that is, toward encouraging students to reason about abstract concepts out loud or in writing and to articulate and justify their thinking in words, can pose challenges to students, instructors, and educational researchers. Many students are socialized into mathematics and statistics primarily through "number-crunching" and procedures that lead to "right answers," and may initially feel exposed and uncomfortable in a classroom sharing their reasoning with their peers, questioning their own and others' assumptions, and reflecting on their own learning processes.

For their part, instructors may fear losing control of the teaching-learning process if they adopt approaches in which much of the learning happens in and through students' conversations and reflective writing. They wonder how to motivate and guide students to share their thought processes and how to evaluate and redirect the sometimes inarticulate and/or less-than-fully-formed (mis)conceptions students attempt to express in classroom discussions (Chazan & Ball, 1999). Additionally, they worry that time spent in peer-to-peer discussion groups and collaborative projects may lead to further muddying of the conceptual waters rather than clarification of concepts. Moreover, they wonder how to assess students' developing conceptual understanding: Are traditional, paper and pencil testing methods adequate for capturing what students know and have learned from participating in active, inquiry-based learning environments?

Researchers attempting to investigate the effects of innovations in teaching/learning processes have traditionally implemented studies based on quasi-experimental research designs comparing measurable outcomes on test scores or pass rates between treatment and control groups. Other variables studied often include student engagement or motivation, as measured through surveys of students' own perceptions. For example, in a quasi-experimental study of active learning in an introductory statistics course, Keeler and Steinhorst (1995) found that students engaged in cooperative learning had higher course grades and persisted to completion of the course in greater numbers than did students who took the course through a traditional lecture format.

Recently, however, there have been studies in mathematics and statistics education investigating the structure, nature, and challenges of classroom interaction. McClain and Cobb (2001) also focused on classroom conversations in order to analyze conceptual development over time in a class of middle school students studying statistics (distribution in data sets). They located evidence of conceptual development in students' articulations of their reasoning in justifying their solutions. McClain and Cobb emphasize the need for the development of classroom norms around what constitutes valid evidence in mathematical and statistical reasoning. Brodie (2007) found that conversations in inquiry-oriented mathematics classrooms often begin with class discussion on a provocative question raised by a student that belies some degree of misconception and come to a productive end with the teacher "taking account of the diversity of ideas that have been expressed, putting them into a relationship with each other, and bringing some resolution" (p. 22). Studies such as these may help demystify strategies and types of conversations in active learning classrooms. The present study aims to contribute to this same

vein of studies of the social and linguistic negotiation of meaning and understanding that comprises learning in inquiry-based classrooms.

1.2 Purpose

This paper reports on a case study of implementation of a problem-based, inquiry learning module in a statistics class for K-12 educators at a research intensive university in the southwestern U.S. The purpose of the study was to provide a detailed description of the pedagogical innovation, as well as to characterize the nature of the discussions students engaged in around the concepts under study. We develop a close analysis of a single class session (day one) to demonstrate the power of problem-based, inquiry learning and make a case for qualitative, discourse-based methods of analysis of teaching and learning. While our methods do not allow us to make strong claims about what students learned, we can show that they engaged in the kinds of conversations that "set the tone" for deeper conceptual engagement throughout the remainder of the course. Moreover, the analytical approach we take aims to help instructors understand what occurred and imagine how they might adopt or adapt such intervention in their classes. The existence of many resources (e.g., http://www.cmu.edu/teaching/designteach/teach/firstday.html) and scholarship specifically on the first day of class attests to its potential for impact and this literature is reviewed in the next section.

2. A Review of the 'First Day of Class' Literature

A delimitation of this study is to focus on the first meeting because of its potential to "set the tone" for the entire course and because of the difficulty in identifying the best use of the first meeting of a course before there has been any opportunity for homework or reading assignments. We now examine what the statistics education and broader literature has to say about the first day of class from both faculty and student perspectives and expectations.

2.1 Faculty Perspectives

In reviewing the overall literature of "faculty-oriented guidance for a successful course beginning," Perlman and McCann (1999, p. 277) report a consensus on some recommended goals ("setting a positive atmosphere, communicating course objectives, taking care of administrative details, grabbing the students' attention, and introducing yourself") and lack of consensus on others (covering course content or using the entire time for the first class meeting). Cowan and Piepgrass (1997, p. 105) note that many faculty feel that "'tyranny of content' demands that lecture begin in that first hour. Others feel that it sets the tone for the course, initiating a culture of rigor…but if the tone is interpreted as hostile, foreign, or intimidating it may have major implications for student success." Hulsizer and Woolf (2009) discuss various suggestions for a first statistics class that include an ice breaker in which students discuss how they would complete the sentence "A class in statistics is like…" or giving students a short pretest to assess if they have the necessary prerequisite mathematical knowledge for the course.

Brooks (1985) notes that "If you expect student participation, it must be encouraged the first day and every other day of class." With a reasonably large or mid-sized class, the first day could

include the efficient conducting of (and debriefing for) a simple experiment as a way of introducing students to the process of gathering data to investigate researchable questions. For an introductory psychology course, Wilson, Stadler, Schwartz & Goff (2009) used the simple vehicle of whether or not the instructor shook hands with the student when greeting the students as they came into the classroom for the first class. For an introductory sociology course, Dorn (1987) suggests the example of noting the arrangement of seating the students have chosen for the first class, to see if males are indeed more likely to sit in the back of the room and females in the front. The numbers of males and females sitting in the front half and in the back half of the room could be readily tabulated into a 2×2 table to begin a conversation (that will be formally concluded later in the course with a chi-squared calculation or difference of proportions hypothesis test).

2.2 Student Perspectives

There is also literature on students' preferences for the first day of class (e.g., Perlman & McCann, 1999; Henslee, Burgess, & Buskist, 2006). Wilson and Wilson (2007) give examples of student preferences (e.g., learning about the course, grading standards, work required) and dislikes (e.g., assigning homework, beginning course material, using the full class time). Although students appear to want the practical information a syllabus contains, several authors have suggested that faculty delay or even omit discussion of some of the details, letting students assume the responsibility of reading the syllabus for themselves. This choice not only results in more thoughtful and focused student questions about the syllabus, but it frees up time during that first meeting day to do something more interesting. Brown (2009) describes having the first day's class time focus on experiencing content and treating the syllabus as just "another reading assignment" that students can do on their own (with the option of quiz questions to provide encouragement to do so). Bennett (2004, p. 106) recommends delaying discussion of (or even passing out) the syllabus until after a demonstration, celebration exercise, or striking example. An example of the latter he uses is the birthday problem: "I go around the room and have people state their birthdays until we find a match. Suddenly, students are interested in math and are trying to figure out why the birthday paradox works."   Other striking examples appear in Sowey (2001).

2.3 Synthesis

Hermann and Foster (2008) note that the faculty first-day goal of immediate active engagement often conflicts with the student goal of gathering practical information, perhaps as part of "shopping" for classes. They reconciled this by setting up a reciprocal interview between students and instructor so that students were immediately actively involved in course-related discussion while getting to form relationships with their peers but also able to get all of their practical questions answered via the representative they selected for each of their small groups. Cowan and Piepgrass (1997, p. 106) report suggestions for the first day informed by both student and researcher feedback, including not giving "testable" material, but instead offering: subject-specific reading skills, a complete (but not overwhelming) syllabus, an informal introduction to the discipline, vehicles for students to recognize how much they already know, and stories of the instructor's own struggle and success.

Dorn (1987, p. 62) describes students on the first day of class as "essentially onlookers, and they may have a certain mental detachment causing them to see things more clearly and objectively." Brouillette and Turner (1992, p. 279) add, "The tone for the entire semester can be set on the first day. Students will never be more receptive…". The key they identify is having students experience content of the course (without already having to know many terms) rather than being told about it. Introductory science class students in focus group interviews conducted by Cowan and Piepgrass (1997) brought up the topic of first-day issues as critical and indicated a need to have their imaginations sparked "before they can be motivated to succeed in an unfamiliar field" (p. 105). Towards this end of piquing curiosity, Dorn (1987) suggests preparing a handout of various "common-sense" views about the subject and having students mark each statement as true or false. This can be a teaser of coming attractions as the students are told they will learn which statements are true as the course unfolds.

But is there any evidence that the quality of the experience on the first day can make a lasting impact for the entire course? Dorn (1987) cites several studies showing high correlation between student evaluations given after the first or second day of class and evaluations of the class and the instructor at the end of the term. More recently, Wilson and Wilson (2007) conducted an experiment in which students were randomly assigned to a positive or a negative first day experience and found that students who had the positive first-day experience reported higher motivation for the majority of the course, and ended the term with significantly higher grades ($p < 0.05$).

3. Background

3.1 Description of Course

The "Statistics in Research" graduate course is a course required in a Master of Arts in Teaching Mathematics degree program at a mid-sized research university near the US-México border. The main focus was to develop the students to be empowered consumers and occasional small-scale producers of (quantitative) mathematics education research, using the Vogt (2007) text to overview topics including: descriptive and inferential statistics, surveys, experiments, psychometrics, simple and multiple regression, ANOVA, chi-squared, and logistic regression.

While the course "starts from scratch", academic maturity is expected so that basic topics such as descriptive statistics can be skimmed or covered quite rapidly compared to more advanced topics. The course emphasizes critical thinking, conceptual understanding, and being able to generate and interpret technology output and research article reporting. Assessments include a final exam and five small-group projects that connected the major topics of the course to datasets, the literature, technology, and mathematical theory.

To accommodate working teachers' schedules, the course meets in the evening (usually two 1.5-hour meetings per week). This time – the spring 2009 semester – it was one 3-hour meeting per week and so it seemed all the more critical to develop a rich first-day (i.e., first-week) experience that could start building a foundation for the course without requiring prior reading. At the end of Section 7, we discuss how this scales down for shorter class periods (as well as how it scales up for larger class sizes).

3.2 Students

Enrollment for this course has fluctuated greatly (between 5 and 30 students) based on factors such as what grants were available that year to support or subsidize graduate students. During the particular semester of this study, there were seven students who attended the first class meeting, and one of them did not remain enrolled in the course, so the final enrollment was six. As will be discussed in Section 7, this limitation turns out not to be highly significant in being able to implement the innovation (as evidenced by the discussion in Section 6), and the innovation can readily "scale up" to be used in a larger class as well.

The seven students at the opening lesson were mostly female (86%), mostly Latina, and also included one student who is a Mexican national and one who is Asian. Both of these students are non-native English speakers. Most of the students are or will be middle/high school teachers. The modal gender and ethnicity of this class is consistent with those of the entire university. All students had taken (including one taking it concurrently) an undergraduate course in mathematical probability and most had also taken at least one undergraduate course in mathematical statistics. All students had interest in pedagogy, as four of the students were teaching then in local high schools, and the other three planned to teach in the future.

It turned out that the three students who the first author had taught in a previous class sat together as a group on the side of the room farthest from the door, while the other four students (including the one who later dropped the class) sat together as a group closer to the door. Because this class was not an elective for the students, it is possible some were anxious, ambivalent or even hostile about being there and perhaps arriving with preconceptions that statistics would be boring, useless, or overly difficult. The scholarship on statistics anxiety (e.g., Williams, 2010; DeVaney, 2010) includes the experiences of students in graduate courses.

3.3 Professor

The statistics professor, a white male native English speaker, had about two decades of overall teaching experience, with greatest concentration in the introductory statistics course. He had taught the course described in Section 3.1 three times previously (in the fall semesters of 2005, 2006, and 2007), the first two times using the book by Agresti and Finlay (1997), before switching to Vogt (2007). The professor has deep ongoing interest in pedagogy, as both a teacher and (statistics education) researcher, and received teaching awards from the Mathematical Association of America's Southwestern Section and the University of Texas System shortly after this study concluded. The professor (designer of the module described in Section 5 and first author of this paper) was a faculty participant in a three-year NSF CCLI (Phase II) research grant (http://2020engineer.iss.utep.edu/world/default.aspx) that funded STEM faculty from several institutions in the southwestern U.S. to gather in a workshop setting to share ideas for and develop inquiry-based modules based around counterintuitive concepts. More deeply nuanced understandings of the teaching and learning processes came from the collaboration of this professor (the "statistics content expert") with the educational linguist (this paper's second author) whose role as grant co-PI was to research the process of classroom implementation of these modules.

4. Data Collection and Methodology

The research design adopted is that of a case study developed around a close description of a lesson's design and a discourse analysis of that lesson's implementation. The broad purpose of this design is to use the analysis to exemplify the principles of the inquiry-based approach to the teaching and learning embodied in this lesson. A detailed description of the lesson's design follows in Section 5.1.

On the first meeting of the class (January 26, 2009), the first half-hour was devoted to procedural items (a/k/a the syllabus), and then the intervention activity was launched for the ensuing 75 minutes. (We note that this means the intervention would have "fit" within the time constraints of a class which meets for 80 minutes simply by having students read the syllabus on their own as homework.) This portion of the class was videotaped and the videotape was transcribed. Through iterative readings (Glaser & Strauss, 1967) of the transcript, five passages were selected and excerpted as illustrative of several commonly mentioned features of inquiry-based approaches. These five excerpts are shown and discussed in Section 6.

Two videocameras were set up on tripods – one for each of the two student groups. Students were able to readily arrange their (movable) desks so that they could see all members of their group. While the videocameras were halfway across the room from the groups, the audio was also recorded by a digital voice recorder placed in the center of each group. It was decided that it made sense for videotaping to occur only during group discussions, not during times when students worked quietly as individuals.

Further rigor came from several types of triangulation – multiple data sources, multiple respondents, and multiple researchers. Transcripts were checked for accuracy by a research assistant as well as by the second author. The two authors went through the transcripts independently before discussing them at several peer debriefing meetings held throughout the semester. During these meetings, the first author was able to share his statistics content knowledge to inform interpretation of a student's response and the second author was able to ask the first author about intentionality in the structure of certain aspects of in-class discussion. An example of the latter was this question the linguist emailed to the statistics professor on March 17, 2010 about Excerpt 5: "I think this is an interesting strategy, getting students to speculate on another students' answer or reasoning. Why did you do this, do you remember? Is this something you typically do?"

A delimitation is that this study was not designed to make strong claims (especially quantitative claims, given the small sample) about what students may have learned as a result of participating in the lesson, much less to argue that this instructional design is more effective than more didactic approaches in improving student learning. Such work has already been undertaken by other researchers, as mentioned in Section 1. So rather than spreading the data collection process thin over all 15 weeks of the course, the decision was made to get a concentrated, in-depth look at the beginning of the class. Thus, the design's strength is to isolate and describe key features and principles of inquiry-based teaching/learning in a graduate level statistics course. The study design also effectively allows us to "slow down" the quick pace of face-to-face discussions in order to "see" and speculate on the underlying thinking.

5. The Intervention Lesson

5.1 Design of Four-Step Intervention Cycle

As summarized in Figure 1 below, the instructional cycle of the intervention involved several phases.

Step 1) *Initial Individual Reflection*: 10 minutes for individuals to consider the Exploration scenario on their own in silence, as they write written responses to a set of questions which ask not only for mathematical answers but also for written reflections explaining the choices or assumptions they are making. The written reflections (i.e., the "Step 1" portions of Appendices A or B) were collected before Step 2 began.

Step 2) *Small Group Discussion*: 5 minutes for the Exploration to be discussed in small groups of 3-4 students each.

Step 3) *Whole Class Discussion*: 5 minutes for the Exploration to be discussed as a whole class, in which groups were given the opportunity to "report out" what observations and insights emerged.

Step 4) *Further Individual Reflection*: 15 minutes for individuals on their own in silence to complete a written reflection (i.e., the "Step 4" portions of Appendices A or B) which included writing about the most important thing they learned, what new questions arose, if the result and process were consistent with their expectations, and if they could think of additional examples of the phenomenon they encountered.
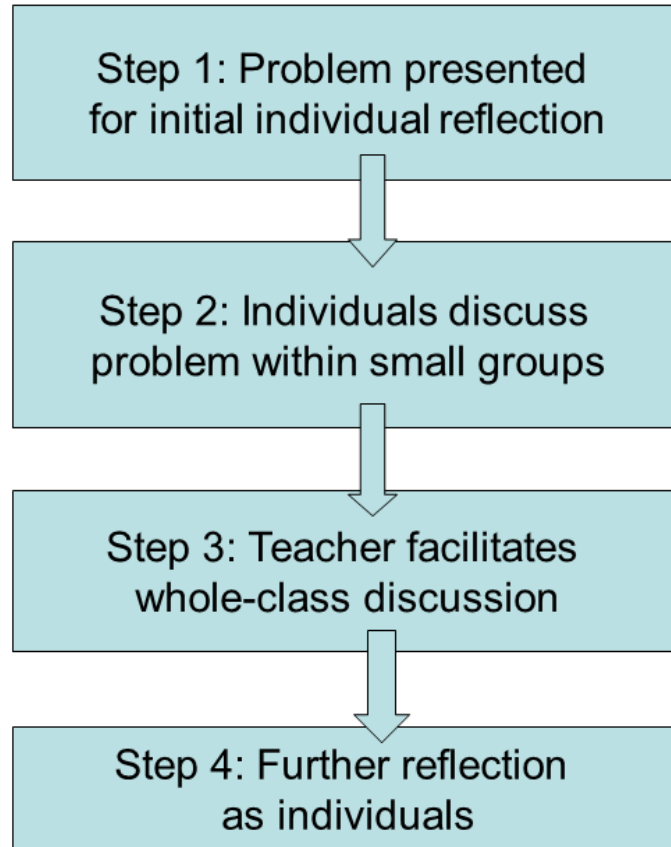
Figure 1: Four-Step Cycle of Instruction for Intervention

This cycle of instruction is consistent with constructivist pedagogical principles in which learners are encouraged to engage with new phenomena introduced through structured, thought-provoking activities and to make sense of their experiences by verbalizing their thinking, and doing so in ways that make it available to their peers and/or instructors (Gee, 2005; Lehrer & Schauble, 2006). Allowing students opportunities first to compose their thoughts individually and in writing, before having to articulate their ideas in conversation, is also supported by theories of second language learning, which suggest that providing second language learners time to focus and develop their ideas before having to express them orally in their second language, in front of classmates and/or the instructor, significantly reduces their anxiety and encourages deeper engagement with concepts at a pre-verbal or multi-lingual level (Fischer & Perez, 2008; Gibbons, 1998). The group discussion in "Step 2" is also valuable in helping such students create contexts for meaning (Rosebery, Warren & Conant, 1992). The first three steps of the cycle can be viewed as a variation of familiar techniques such as "think-pair-share" (e.g., Lyman, 1981) or "1-2-4-whole group" (e.g., Minich, 2010).

5.2 Content in Lesson

The scenarios for both of these cycles are designed to be open-ended in a way that helps introduce students to the idea that statistics is different from math – that there can be more than one reasonable way to interpret a data set. Indeed, the challenges of getting students to consider multiple solutions to a problem have been documented (Silver, Ghousseini, Gosen,

Charalambous, & Strawhun, 2005). The practice of statistics is filled with open-ended challenges – such as what to do when a dataset has an outlier – and so these scenarios are indeed setting the tone.

The two explorations were chosen deliberately for several reasons: (1) the context was related to students' backgrounds (i.e., their status as current or future teachers), (2) the prerequisite mathematics was limited to topics that could safely be assumed the students already understood (indeed, the building blocks of fractions and averages are part of the state's secondary mathematics standards that they themselves were teaching), (3) the questions were straightforward to pose, (4) the context of the scenarios would be revisited later in the course (in Chapters 16 and 3 of the textbook), and (5) the questions were sufficiently open-ended so as to support genuine reflection and multiple approaches/representations, possibly spurred further by an initial conclusion that was conflicting or counterintuitive. Using an example from geometry, Harper and Edwards (2011) give a very insightful demonstration how a question can be made progressively more inquiry-based, along with a rubric for assessment.

The idea of using counterintuitive examples to stimulate students' curiosity and engagement in statistics was inspired by several directions: (1) a major thread within the CCLI engineering grant discussed in Section 1.1 that has resulted in several papers at national engineering education conferences, (2) the call for active learning by ASA (2010), and (3) the emerging evidence from some of the literature in statistics education (discussed in Appendix C) for motivational potential to stimulate cognitive conflict.

5.2.1 "Student Test Performance" Exploration

The exploration used for the first cycle was driven by a dataset from Movshovitz-Hadar and Webb (1998, p. 113) consisting of number of tests passed and failed by two (presumably, hypothetical) students in each of two semesters, as well as for the full school year (i.e., both semesters combined). In this dataset, one student has a higher passing rate in the fall and in the spring semesters, but a lower passing rate for both semesters combined. This surprising lesser-known reversal phenomenon is known as Simpson's paradox and has also occurred in large-scale educational data (e.g., Terwilliger & Schield, 2004).   The term "Simpson's paradox" was not told to the students during the lesson because use of such explicit terms can inhibit student exploration (e.g., Harper and Edwards, 2011). Indeed, readers can verify that the list of questions for this cycle (see Appendix A) contains no academic statistics terms at all.

Pedagogical benefits or connections in addition to those listed in Lesser (2001, p. 130) for Simpson's paradox include possible connection to the "More A—More B" intuitive rule discussed by Stavy and Tirosh (2000) and its importance to real-world quantitative literacy. Former Mathematical Association of America president Lynn Arthur Steen (2001, p. 11) includes "Recognizing how apparent bias in hiring or promotion may be an artifact of how data are aggregated" on a list of systematic thinking needed for citizenship.

5.2.2 "Average Class Size" Exploration

The exploration used for the second cycle was driven by a dataset from Lesser (2009) consisting of exploring what would (or could) be the "average class size" for a particular set of classroom sizes at a small fictitious school. In the dataset, (at least) six distinct answers are possible, based on whether one interprets "average" to be the mean, median or mode, and (more subtly) whether one chooses to compute a per-classroom average or a per-student average. The importance and applicability of the mathematics underlying this particular exploration is established in a broad range of literature (e.g., Hemenway, 1982; Kadane, 2008; Lann & Falk, 2005; Schwenk, 2006; Wagner, 2009). See Appendix B for the list of questions used in the cycle for this exploration. We note that average class size is by no means the only instance of an ambiguous average. For example, Falk and Lann (in press) state: "Consider the real-life question of the average speed of cars on the road. There is no single unequivocal answer. It all depends on the particulars of the way in which the question is defined (Falk, Lann, and Zamir, 2005)."

We also note that there are other examples of engaging statistics questions explicitly designed to help students understand the need for careful definition in order to avoid ambiguity or multiple answers. In one such example, Isaacson (2011) shows the counts of gold, silver and bronze medals from the 2008 Olympics for five countries and asks which country did "best"? The authors conjecture that some of the reasons identified by Silver et al. (2005) for why mathematics teachers offer only limited support for students to consider multiple pathways to a single correct answer may have some overlap or parallels to why some statistics teachers offer only limited support for students to consider multiple answers.

6. Analysis of Classroom Interaction

In this section, we analyze five excerpts from the transcript of the class' interaction on the opening day's lesson. The excerpts discussed (see Table 1) were chosen because they exemplify elements of problem-based inquiry learning as they played out in small group and whole class discussions around the two counterintuitive examples.

Table 1: Overview of Video Transcript Excerpts

Excerpt #	1	2	3	4	5
Step of Cycle (Figure 1)	2	2	3	3	3
Related Appendix	A	A	A	A	B

Our analysis highlights the nature of the interaction – how it reflects certain principles of inquiry-based learning and how it makes certain ideas available for reflection and consideration by the class. In some excerpts, we also consider what the ideas raised might indicate about the quality of students' thinking and development of their conceptual understanding. All students' names in the excerpts are pseudonyms. At the end of this section, we highlight three aspects of inquiry-oriented classrooms that stand out: varying participation structures, well-crafted/chosen problems, and instructor's role as a facilitator/co-learner.

6.1 Small-Group Discussion

In the Excerpt 1 below, a group of four female students discuss their answers to the problem in Appendix A, which asks them to compare two students' pass rates on tests taken over Fall and Spring semesters. The first question in the problem asks which of the two students did better. The second question asks students to consider what assumptions they're making in coming to an answer. The third question, which the students are considering in the first excerpt below, asks them, given their assumptions, if they can think of any other possible answers to the question of which of the two students did better.

Excerpt 1:

Brenda: I didn't really understand - I didn't know how to answer number three. ((READING ALOUD QUESTION #3, FROM THE HANDOUT THAT APPEARS IN APPENDIX A)) by making different choices or assumptions can you think of another possible answer to question number one?

Amy: I thought- well, the way I saw it is that if you just ignore everything else then see that Patricia [sic] had five, and Bruce had eight, then you would automatically think that Bruce did more. If you didn't look at the rest.

Brenda: Oh, okay.

Li-Ling: I just used the same way but I just- to test, like, who did worse the last semester, so its [sic] mean, like- because for the first question, you choose Patría?

Brenda: Mhm.

Li-Ling: Like, she performs well, so basically you can use the same way, like, who did worse. So is [sic] should be Bruce did worse the first semester.

Brenda: Yeah, the percent first.

Amy: Mhm.

Brenda: Okay. The same thing for the spring semester.

Li-Ling: Yeah. Uh-huh.

In this excerpt of the interaction, we see how the small-group participation structure enables a sort of risk-taking (i.e., Brenda's admission that she did not know how to answer the third question) and perspective-sharing (i.e., Amy and Li-Ling describe their interpretations and approaches) that might not otherwise occur in a lecture or whole-class discussion. In other words, were it not for this low-stakes type of peer-to-peer sharing of answers/solutions, Brenda might not have risked letting on that she was in doubt. Instead, Brenda's peers readily share their perspectives and approaches, doing so in a way that is both cooperative and somewhat tentative. Amy expresses her approach initially as simply her own perspective ("Well, the way I saw it was…"), and then she switches to the second person ("if you just ignore everything else…"), encouraging Brenda to see the problem from her perspective. Li-Ling then expresses her approach as applying a reverse logic to the more obvious answer to question one. Thus neither Amy nor Li-Ling suggests that their approaches are "correct" or "the right way." Rather they offer their perspectives in a spirit of collaboration/cooperation that is at the heart of peer-to-peer learning in inquiry-oriented approaches. In her subsequent turns, Brenda is able to elaborate on Li-Ling's explanation ("Yeah, the percent first" and "Okay. The same thing for the spring semester"), which suggests that Brenda accepts and understands her peers' explanations.

It is interesting to note that in her first turn in Excerpt 1, Amy is focusing initially on comparing numbers of tests passed, rather than fractions of tests passed. This illustrates how more and more sophisticated answers are possible to the intentionally ambiguous, open-ended prompt of who "did better"?

In Excerpt 2 below, Group 2 considers the same problem and Kara shares with her group an assumption she made in considering question 2 (of the sheet in Appendix A).

Excerpt 2:

Kara:	The only thing I saw where that [comparing percentages of tests passed by Bruce and Patría to determine which student "did better"][1] might be a problem is if each test was weighted differently.
Alberto:	Each test is what?
Kara:	Like, if each test was weighted differently. Like, if one test counted for, like, ten percent, and the other one counted for twenty. If something like that happened in the semester, then the percentage wouldn't tell you anything.
Alberto:	Mhm.
Marta:	Mhm.
Kara:	So, my assumption was that each test has the same weight. So five out of six means you have to have the same weight for each test. So you can say by percentage who did better. Does that make sense?
Alberto:	Yes. Yeah. You mean five out of six and eight out of ten, right?

In Excerpt 2, Kara explains to her group an assumption she made that each test has the same weight, a creative logical leap that is not explicitly prompted by the wording of the problem. It is interesting to see that this group, like the group in the previous excerpt, is also using a collaborative tone in their conversation. After explaining her thinking, Kara checks to see if her group members follow and agree with her. Although both respond affirmatively to her query ("Does that make sense?"), it is not clear that either Alberto or Marta fully understand Kara's explanation because neither student responds in a substantive way to it. Longer excerpts of the transcript show that the members of this group worked together in a collaborative way, sharing answers, admitting to each other when they were confused or had made a mistake, and offering explanations of their own thought processes. However, as Excerpt 2 illustrates, just because the conversation takes a cooperative tone does not necessarily mean that all participants are always attending to the same phenomenon and actually learning the same things at the same time. In this case, there is no way to know what Alberto and Marta took away from Kara's explanation of her assumption. Nevertheless, the sort of collaborative talk visible in this excerpt of the transcript is representative of the tone and mode of the conversation during the entire lesson, showing that this group, like the other one, also used the small-group conversation to test and compare their ideas before "going on record" in front of the professor and the entire class.

---

[1] In instances such as this, the words in [square brackets] have been added to the transcript in order to clarify speakers' meanings, most often because they have used a pronoun or some other ambiguous wording to refer to parts of the conversation that occurred before the excerpted part.

6.2 Whole-Class Discussion

In Excerpt 3 below, students are sharing their answers to the problem in <u>Appendix A</u> with the whole class. The professor first asks the class which of the two students in the problem scenario did better in the fall semester (Question 1). The students concur that Patría did better than Bruce in the fall semester, so the professor asks for their reasons for saying so. One student offers her reasoning that Patría's test scores were better than Bruce's on a straight percentage comparison. Another student offers that she used the same basic reasoning, but simply compared fractions rather than computing a percentage. The professor acknowledges both answers, saying, "Okay." Then he probes for the assumptions underlying their reasoning (Question 2).

Excerpt 3:

Prof:	What assumptions do you feel like went into that? Or do you feel you made any assumptions to be able to say that?
Kara:	Well I put that, the assumption is that, for each of the tests within the semester, that they have the same weight. You don't have like the first test being 20% the second being 30 and so on. 'Cause if that's occurring then you can't really conclusively say who did better without having more information.
Prof:	Fair enough. Okay, did anyone else have a different kind of assumption or choice that they identified?
…	
Alberto:	Well, another thing is, I assumed they were not in that same classroom because one made [sic] six tests and the other one ten, right?
Prof:	Oh, because of the different numbers? Okay. That's an interesting point. So then maybe they weren't in the same classroom, or maybe they were, but it was differentiated instruction and the students could kinda go at their own pace. But that's a good point, that they might not have been in the same classroom. But even so, you could still say, but which student did better? Even if they were in different classrooms. Okay, so, but that's interesting. Maybe they weren't in the same classroom.

An important aspect of this exercise was eliciting and probing these kinds of assumptions. This was built into the written handout that the students referred to (in Questions 2 and 4), but expecting that students might make different assumptions, the professor felt it was also worthy of whole class discussion. In eliciting and probing the students' assumptions, the professor acknowledges and accepts any and all answers without evaluating them. With Alberto's response, the professor accepts and then elaborates on his answer, providing a related explanation to the one Alberto had provided. Such acceptance of and engagement with students' ideas by the professor serves to empower students and set the standard that in this classroom students' ideas are inherently valuable and serve as the starting place for the class's ongoing inquiry.

In Excerpt 4 below, the professor highlights Simpson's paradox, without naming it, by asking students to comment on the fact that in the first problem, it appeared that Patría had "done better" in both the fall and spring semesters, yet Bruce had "done better" overall.

Excerpt 4:

Prof:	Now, wait a minute. You all told me Patría did better in the fall and spring, but not for the whole year?
Tania:	Yeah.
Prof:	Did that seem weird to anyone?
Unidentified student:	Mhm.
Prof:	Did that come up in discussion? Did that kinda seem weird? Did you notice that those answers look different? Did that come up in discussion, Amy?
Amy:	Mhm. A little, yeah. We said that we automatically assumed that Patría would do better, but then when you do the numbers-
Brenda:	It's Bruce.
Prof:	Huh. So what do you make of that? Was that- Have you seen this kinda thing before?
Kara:	I don't know why it happens, but I've heard of it. I know what it is.

It is clear from this excerpt that the students recognize the paradox, even if they are not yet able to articulate a full explanation of it. Nevertheless, the problem and the discussion have served the goal of raising their awareness of the existence of such a paradox. Later in the semester (namely, in Chapter 16 of the text), the class returned to Simpson's paradox and students were reminded of the example of Patría and Bruce's test scores.

In Excerpt 5 below, we turn to the second problem (on average class size: see Appendix B) and the professor asks students for their responses. Most students based their answer on a calculation of the mean class size. However, Kara offered that she had calculated the answer in two different ways and had two responses. Rather than asking Kara to explain her calculations right off, the professor asked if anyone else could explain how Kara had arrived at her answers.

Excerpt 5:

Prof:	Now what did you say was the average class size? Was it a single number?
Kara:	Five. [the mean of {3, 3, 4, 10}]
Several stds.:	Five.
Prof:	Everybody said five?
Amy:	Yeah.
Kara:	I had two options. I said five and three.
Prof:	Five or three. Okay. Interesting. Did anybody have any other second option besides five or three? ((1.5 SECOND PAUSE)) Okay, how do you all think Kara came up with three? Can you guess?
Unidentified student:	She took the two [class-size values] that had the three? [the mode of {3, 3, 4, 10} is 3]
Prof:	Okay. And in your own words, Kara, how many- why did you pick three as the other possible answer?
Kara:	Well, I thought of average- the term average in two different ways. The mathematical average, which is the mean, which is five.
Prof:	Okay.

Kara: And then I also thought of it in terms of how you would use it in common English which means that when you say something's average, it's what's typical, what happens most often. So to me that's the mode.

Prof: Ah. The most common. Okay. All right.

In this excerpt, by asking the class to consider how Kara came to choose "three" as a possible answer, the professor demonstrates a strategy for broadening class participation: asking students to explain another student's answer. This strategy is particularly useful when one or a few students dominate class discussions or tend to offer answers more readily than others. In the beginning of a semester, such a strategy helps set the expectation that all students will follow the discussion and actively participate in developing the class's understanding of the concepts under study.

6.3 Highlighted Findings

Asking a student to explain another student's answer also disrupts a common pattern of interaction in classrooms in which instructors tend to play a dominant, omniscient role by asking all the questions, expecting specific "right" answers, ratifying or rejecting students' answers, and then explaining those answers in other words for students who might not have followed the logic. Inquiry-based instruction places students and instructors on more equal footing. As members of the classroom community, students and instructors are all "questioners" and "knowers"; everyone is responsible for participating in the classroom conversation and, as such, in the construction of the knowledge of the classroom community.

To summarize, these five excerpts illustrate several aspects of inquiry-oriented classrooms and the strategies that instructors and students employ in them that are worthy of highlighting:

1) *Varying participation structures*: The sorts of collaborative talk encouraged through **small-group interactions,** such as those in Excerpts 1 and 2 where students are engaging in discussions around well-crafted, open-ended problems, provide students with opportunities to articulate their misunderstandings and developing understandings in a low-stress, peer-to-peer environment. In this context, students may be more likely to risk being wrong, or to admit that they don't know an answer, as Brenda did in Excerpt 1, eliciting explanations of their reasoning from her peers.

   In **whole-class discussion,** the instructor has the opportunity to probe students' reasoning and to create common conceptual ground for the class to build on over time.

2) *Well-crafted problems:* While some practitioners of inquiry pedagogy argue that it is best if the problems the class investigates come from students own queries and concerns, it is also possible to anticipate some of those concerns, as well as the common conceptual misunderstandings that students might hold, and to design explorations or problems based on open-ended questions that will require them to grapple with dilemmas, confront misconceptions, and/or wonder about implications and applications. In the case of the two problems, the prompt given to these students was designed to encourage them to think broadly and entertain multiple possible solutions.

3) *Instructor's role as facilitator, co-learner:* Unlike in a traditional classroom where the instructor works to maintain control of the conversation and adopts a stance that suggests she is omniscient, in an inquiry-based classroom, the instructor is a both a facilitator and co-learner. Far from adopting a hands-off approach, however, the instructor must develop (or choose) good topics and problems and know how to engage students in them and encourage and support their investigations. Excerpts 3, 4, and 5, show the instructor challenging students to articulate their reasoning or the reasoning of their peers. The instructor is shown withholding judgment and acknowledging ideas that he hadn't considered. In short, the instructor is setting the tone for certain modes of engagement and discussion that he hopes will continue throughout the course.

7. Discussion

In this paper, we have presented a detailed case study of a single lesson – from design through implementation – for the first day of class in a graduate-level statistics education course. We have described the sorts of design decisions the instructor made in developing the module in order to engage students' interest through counterintuitive problems and set a tone for types of statistical thinking students would engage in throughout the course. We have also illustrated how that module "played out" in a particular class setting, focusing on specific aspects of the classroom conversation that unfolded and how the instructor attempted to guide students toward, if not wholly new understandings, at least the initial recognition of some cognitive conflict.

We argue that what happens on the first day of class matters – it sets the tone for the entire semester in at least two important ways, as exemplified in the case study. First, students were introduced to examples of types of problems to be grappled with in the course, and in the process, also introduced to habits of mind and questioning that are valued in the discipline (e.g., Chance, 2002). An example from our study is to be aware that it is not enough to say "find the average" without knowing which type of average and the basis unit being averaged over. More generally, this is a habit of mind to seek clear operational definitions and to be aware of how different choices or assumptions can impact a model or result. These more open-ended challenges do not occur as regularly in mathematics, where the problems encountered in school typically have one clear right answer. While the content of the explorations may be specific to this course, the strategy of introducing counterintuitive concepts or dilemmas that result in cognitive conflict in order to spur students' learning applies to any university subject or course, as does the call of Rossman (2010) to "ask good questions".

Second, students were also socialized on the first day into the forms of interaction and engagement that would characterize subsequent class sessions, which often followed a similar trajectory as the cycle in Figure 1, with the difference that Step 4 was no longer an explicit part of class time. As is expected in inquiry-based instruction, students were not left to their own devices. Rather, the instructor guided each of these phases of the class, orienting students to the concepts under consideration through the written problem and through his oral directions, questions, and probing, between and during discussion phases. We illustrated how the instructor established and maintained his stance as a facilitator of students' learning, rather than as the final arbiter of correct answers. Such tactics as asking students to explain a peer's answer and requiring students to articulate their reasoning, not simply to report their answers, could also be adopted in nearly any university classroom.

By delaying or omitting routine discussion of the syllabus, the intervention cycle and activities described in the lesson can be used in any introductory statistics course with an 80-minute meeting time (which is typical for 3-credit hour classes that meet twice a week). It can also be scaled down to fit classes whose meetings involve a 50-minute period (which is typical for 3-credit hour classes that meet three times a week) simply by using only one of the two explorations. Still further time could be saved by omitting most of the questions in Step 4 (which were there mainly to prompt students to reflect on, for example, some of the ways in which mathematics differs from statistics and how that might affect student expectations for a statistics course) or by shortening the length of time allotted for individual reflection in Step 1. Indeed, a 30-minute cycle (paralleling Steps 1-3) was put to regular effective use by a colleague of the first author in an undergraduate mathematics course for pre-service teachers (see Section 3.1 of Esquinca, 2011).

Another way to maximize usage of time is to develop scenarios that are perhaps more arithmetically streamlined, such as this one used by the professor on the first day he taught this course during the fall 2011 semester:

> "On a high-stakes test where the minimum passing score is a 7, which teacher's class did better? Mr. Jones' 5 students' scores were: 2, 3, 7, 7, 7. Ms. Gomez' 5 students' scores were 4, 5, 6, 6, 8."

Students readily recognized that Jones' class had triple the passing rate of Gomez' class, even though it had a lower average.

The instructional cycle can be readily scaled up to work with larger class sizes, with the main difference being a larger number of groups (each having 3 or 4 students). The dynamics within each group during Steps 1 and 2 would be no different as they were in this study with a smaller class. In Step 3, however, the time constraint might mean that not every group gets to share all of their insights with the entire class. In practice, the number of different approaches generated by a small number of groups may not be that different than for a larger number of groups, and after soliciting three different ideas – each from a different group in the room – it is likely that there will not be many other groups with ideas that are truly different from what has already been said. The instructor can use time efficiently by not calling on groups arbitrarily, but by calling on one group "at random" and then asking "what group came up with a different answer/approach from what we have already heard?"

The instructor does not get a second chance to make a first impression. As students enter a classroom for the first time, they notice how the room is arranged, how the instructor is positioned, whether they readily receive a greeting, and whether there is something already awaiting them – a question, an outline, a warm-up problem, or even a cartoon (Lesser and Pearl, 2008). While there are inspirational oral readings (e.g., Lesser, 2010b) and engaging interpersonal activities (e.g., Arvidson and Huston, 2008) designed to help set the tone for the first day (or first moments) of almost any kind of course, the authors argue that it is valuable to have available resources that are content-rich and tailored for the particular audience enrolled in statistics classes.

APPENDIX A
"Student Test Performance" exploration

"Step 1":

TABLE 2: Data on Classroom Tests						
	fall semester		spring semester		full school year	
	P	B	P	B	P	B
Tests passed	5	8	6	4	11	12
Tests failed	1	2	8	6	9	8
Tests taken	6	10	14	10	20	20

Above is test performance information for two students, Patría and Bruce.

1.) Explain which student did better in the fall semester (or explain why it is not possible to determine this).

2.) What choices or assumptions are you implicitly making in giving your answer in Question #1? Explain.

3.) By making different choices or assumptions, can you think of another possible answer to question #1? Explain.

4.) Explain which student did better in the spring semester (or explain why it is not possible to determine this).

5.) Explain which student did better for the overall full school year (or explain why it is not possible to determine this).

"Step 4":

6.) What did you learn from the "student test performance" exploration and if you learned more than one thing, which thing was the most important?

7.) What new questions, if any, does this exploration generate for you and what, if anything, do you feel confused about?

8.) Was the process or the result of the "student test performance" exploration consistent with how you view **mathematics**? Explain.

9.) Was the process or the result of the "student test performance" exploration consistent with how you view **statistics**? Explain.

10.) Was the process of the "student test performance" exploration consistent with your prior expectations for this particular course? Explain.

11.) Can you think of any other examples or situations where you could apply a similar process or obtain a similar result to the "student test performance" exploration? Explain (please be as specific as you can).

APPENDIX B
"Average Class Size" exploration

"Step 1":

1.) Do you consider the word "average" to be a clear-cut term with no ambiguity in its meaning? Explain.

2.) "A small school has 185 students divided among 7 classrooms.

The classroom sizes are: **20, 20, 20, 25, 30, 35,** and **35**." What would you say is the 'average class size' (please show your work and reasoning)? [*note: the simpler dataset {3, 3, 4, 10} from [Lesser (2010a)](#) was offered as a substitute if the students wished*]

3.) What choices or assumptions are you implicitly making in giving your answer to question #2?

4.) By making different choices or assumptions, can you think of other possible answers to question #2? Explain.

"Step 4":

5.) What did you learn from the "average class size" exploration and if you learned more than one thing, which thing was the most important?

6.) What new questions, if any, does this exploration generate for you and what, if anything, do you feel confused about?

7.) Was the process or the result of the "average class size" exploration consistent with how you view **mathematics**? Explain.

8.) Was the process or the result of the "average class size" exploration consistent with how you view **statistics**? Explain.

9.) Was the process of the "average class size" exploration consistent with your prior expectations for this particular course? Explain.

10.) Can you think of any other examples or situations where you could apply a similar process or obtain a similar result to the "average class size" exploration? Explain (please be as specific as you can).

APPENDIX C
Background on Counterintuitive Examples

There has been much discussion on the power of creating uncertainty or surprise with striking examples (e.g., Sowey, 2001, in statistics; Movshovits-Hadar, 1988, or Zaslavsky, 2005, in mathematics) to motivate students. Striking examples include both "intuition-building" scenarios such as analogies (e.g., Martin, 2003; Groth & Bergner, 2005), as well as examples that engage the intuition by being initially counterintuitive (e.g., Movshovitz-Hadar & Webb, 1998; Huck & Sandler, 1984; Romano & Siegel, 1986; Székely, 1986). More advanced collections of counterintuitive examples are Stoyanov (1987) and Wise and Hall (1993). For the definition of a counterintuitive example, we follow Lesser (2002), who requires "both that [most students would] have an initial expectation or primary intuition (a directional hypothesis, so to speak) and that that primary intuition with respect to a result contradicts and is, at least initially, very resistant to the normative view."

Lesser (1998, 2002) discusses competing views on the use of counterintuitive examples, but the empirical evidence collected so far indicates strong potential for their judicious use. Lesser (1998) conducted survey research of university introductory statistics students that found a highly significant positive correlation ($r = .666$, $n = 97$, $p < .001$) between level of interest and level of surprise for each item on a list of true statistics statements, though the survey instrument's construction may not have definitively ruled out the possible role of contextual variables. Movshovitz-Hadar and Hadass (1990) found relevance to cognitive conflict, motivation, misconceptions, and constructivism in a naturalistic study of 52 pre-service secondary mathematics teachers encountering a fallacious proof of the irrationality of a particular number. The case study of Wilensky (1995) found that engagement with paradox can motivate learners to overcome conceptual and epistemological obstacles to learning probability.

Also, Lesser (1999a) found from a year-long case study using "typical case" sampling that introductory statistics college students seemed to enjoy the experiences in which their intuition was surprised, characterizing it as a refreshing change from the routine predictability of their usual mathematics/statistics experiences. By staying within the recommendation of Lesser (1998, p. 12) "to limit examples to those that actually occur in real life (this eliminates contrived probability paradoxes, but still leaves plenty of examples to choose from) and that can be readily explained or explored by means other than analytic mathematics alone," it did not seem unreasonable to conjecture that any impact of such counterintuitive examples would be to help not hinder motivation.

Other research involving counterintuitive scenarios did not find or look for an attitudinal difference, but reported learning successes that one might reasonably imagine were supported by or accompanied by some respectable level of motivation. For example, Castro (1998, p. 252) found high school instruction from a conceptual change perspective "significantly improves subjects' intuitive probability reasoning, compared to the traditional instructional model." Noting that students find certain results counterintuitive because of their reliance on heuristics such as representativeness and availability, Shaughnessy (1977) found that course sections randomly selected to use an active, small-group, problem-solving, model-building, experimental approach that worked through scenarios (e.g., disjunctive events such as the 'birthday problem')

were more successful than course sections randomly selected to use a lecture format in changing student misconceptions (especially those stemming from reliance on representativeness). One particular success Shaughnessy noted was the near elimination of the use of a representative multiplier when working with the probability of disjunctive events (e.g., the classical birthday problem).

Support for counterintuitive examples is also consistent with the observation of von Glasersfeld (1995, p. 68):

> "The learning theory that emerges from Piaget's work can be summarized by saying that cognitive change and learning in a specific direction take place when a scheme, instead of producing the expected result, leads to perturbation, and perturbation, in turn, to an accommodation that maintains or reestablishes equilibrium."

The idea that cognitive conflict generated from an unexpected result could accelerate learning or facilitate deeper engagement has been discussed by many educators in many forms, including conceptual change, conflict teaching, and structured academic controversy. Several reasons why this approach may be effective are offered by del Mas and Bart (1989, pp. 42-43):

> "when subjects encounter a situation which they believe can be assimilated into their existing schemas but which are not resolved when the schemas are applied, the contra-dictions prepare the subjects to restructure or accommodate their schemas.….Second, a contradictory situation helps to highlight conflicts between a subject's present strategy and the correct strategy…which can aid better recall of the new strategy. Finally, the contradictory situation helps a subject focus on key relationships among variables and to disregard the variables which may lead to misjudgments or misinterpretations of causal effects."

It should be noted that what is classified as counterintuitive can vary over the centuries. For example, Northrop (1944, p. 171) notes that the item "In two tosses of a single coin, what is the probability that heads will appear at least once?" actually was solved incorrectly by the first-rate eighteenth-century French mathematician D'Alembert. However, the number of heads in two tosses is now considered one of the simplest ways to introduce a probability distribution that is nonuniform, and appears as a basic example in many introductory textbooks (e.g., McClave & Sincich, 2000, p. 167; COMAP, 2009, p. 250).

It should also be noted that not all counterintuitive examples are well-suited for the intervention in this study. For example, the birthday problem (e.g., Lesser, 1999b) would not do well because (1) it's more probability than statistics, (2) full understanding requires the ability to work with probability principles/rules that students might not have as they walk into the first day of class, and (3) the "punchline" is something that students may have heard before and there is a danger that this could ruin the experience for other students.

Acknowledgements

References

Agresti, A. & Finlay, B. (1997). *Statistical methods for the social sciences* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

American Statistical Association (2010). *Guidelines for assessment and instruction in statistics education* (GAISE) *college report*. Alexandria, VA: ASA [www.amstat.org/education/gaise/]

American Statistical Association (1994). G. Burrill (Ed.), *Teaching statistics: Guidelines for elementary to high school*. Palo Alto, CA: Dale Seymour.

Arvidson, P. S. & Huston, T. A. (2008). Transparent teaching. *Currents in Teaching and Learning*, *1*(1), 4-16.

Bennett, K. L. (2004). How to start teaching a tough course: Dry organization versus excitement on the first day of class. *College Teaching*, *52*(3), 106.

Bransford, J. D., Brown, A. L. & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brodie, K. (2007). Teaching with conversations: Beginnings and endings. *For the Learning of Mathematics, 27*(1), 17-23.

Brooks, D. W. (1985). My first day of class—is a typical class day. *Journal of Chemical Education*, *62*(7), 604-605.

Brouillette, J. R. & Turner, R. E. (1992). Creating the sociological imagination on the first day of class: The social construction of deviance. *Teaching Sociology*, *20*(4), 276-279.

Brown, K. (2009). Don't waste the first day. *The Teaching Professor*, *23*(9), 3.

Castro, C. S. (1998). Teaching probability for conceptual change. *Educational Studies in Mathematics*, *35*(3), 233-254.

Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, *10*(3), www.amstat.org/publications/jse/v10n3/chance.html

Chazan, D. & Ball, D. (1999). Beyond being told not to tell. *For the Learning of Mathematics, 19*(2), 2-10.

Consortium for Mathematics and its Applications (2009). *For all practical purposes: Mathematical literacy in today's world* (8th ed.). New York: Freeman.

Cowan, M. M. & Piepgrass, K. W. (1997). The first day of class on a two-year campus: What students really want us to know about their science classroom experiences. *Journal of College Science Teaching*, *27*(2), 104-106.

delMas, R. C. & Bart, W. M. (1989). The role of an evaluation exercise in the resolution of misconceptions of probability. *Focus on Learning Problems in Mathematics*, *11*(3), 39-54.

DeVaney, T. A. (2010). Anxiety and attitude of graduate students in on-campus vs. online statistics courses. *Journal of Statistics Education*, *18*(1), 1-15.
www.amstat.org/publications/jse/v18n1/devaney.pdf

Dorn, D. S. (1987). The first day of class: Problems and strategies. *Teaching Sociology*, *15*(1), 61-72.

Esquinca, A. (2011). Bilingual college writers' collaborative writing of word problems. *Linguistics and Education*, *22*(2), 150-167.

Falk, R., Lann, A. & Zamir, S. (2005). Average speed bumps: Four perspectives on averaging speeds. *Chance*, *18*(1), 25-32.

Falk, R. & Lann, A. L. (in press). Weighted means through the looking glass. *Teaching Statistics*. [published online on April 28, 2011]

Fischer, J. & Perez, R. (2008). Understanding English through mathematics: A research-based ELL approach to teaching all students. *Research Monograph of TODOS*, *1*, 43-58.

Gee, J. P. (2005). Language in the science classroom: Academic social languages as the heart of school-based literacies. In R. Yerrick and W-M. Roth (Eds.), *Establishing scientific classroom discourse communities: Multiple voices of teaching and learning research* (pp. 19-44). Mahwah, NJ: Lawrence Erlbaum Associates.

Gibbons, P. (1998). Classroom talk and the learning of new registers in a second language. *Language and Education*, *12*(2), 99-118.

Glaser, B. G. & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Publishing Company.

Gordon, M. (1991). Counterintuitive instances encourage mathematical thinking. *Mathematics Teacher*, *84*(7), 511-515.

Groth, R. E. & Bergner, J. A. (2005). Pre-service elementary school teachers' metaphors for the concept of statistical sample. *Statistics Education Research Journal*, *4*(2), 27-42, www.stat.auckland.ac.nz/~iase/serj/SERJ4(2)_groth_bergner.pdf

Harper, S. R. & Edwards, M. T. (2011). A new recipe: No more cookbook lessons. *Mathematics Teacher*, *105*(3), 180-188.

Hemenway, D. (1982). Why your classes are larger than 'average.' *Mathematics Magazine*, *55*(3), 162-164.

Henslee, A. M., Burgess, D. R. & Buskist, W. (2006). Students' preferences for first day of class activities. *Teaching of Psychology*, *33*(3), 189-191.

Hermann, A. D. & Foster, D. A. (2008). Fostering approachability and classroom participation during the first day of class: Evidence for a reciprocal activity. *Active Learning in Higher Education*, *9*(2), 139-151.

Huck, S. W. & Sandler, H. M. (1984). *Statistical illusions: Problems* [companion volume: *Statistical illusions: Solutions*]. New York: Harper and Row.

Hulsizer, M. R. & Woolf, L. M. (2009). A guide to teaching statistics: Innovations and best practices. [web content to supplement the Wiley-Blackwell book of the same title] www.webster.edu/teachstats/chapter4.htm#firstclass

Isaacson, M. (2011). Multiple choice: Olympic success. Paper presented at 4th United States Conference on Teaching Statistics, Cary, NC. www.statlit.org/pdf/2011IsaacsonUSCOTS6up.pdf

Kadane, J. B. (2008). Load factors, crowdedness, and the Cauchy-Schwarz inequality. *Chance*, *21*(1), 33-34.

Keeler, C. M. & Steinhorst, R. K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field. *Journal of Statistics Education*, *3*(2), www.amstat.org/publications/jse/v3n2/keeler.html

Lann, A. & Falk, R. (2005). A closer look at a relatively neglected mean. *Teaching Statistics*, *27*(3), 76-80.

Lehrer, R. & Schauble, L. (2006). Scientific thinking and science literacy. In W. Damon, R. Lerner, K. A. Renninger, and I. E. Sigel (Eds.), *Handbook of child psychology,* 6th ed. *Vol. 4: Child psychology in practice* (pp. 153-196). Hoboken, NJ: John Wiley & Sons.

Lesser, L. (1998). Countering indifference using counterintuitive examples. *Teaching Statistics*, *20*(1), 10-12. www.rsscse-edu.org.uk/tsj/wp-content/uploads/2011/03/lesser.pdf

Lesser, L. (1999a). An interpretivist investigation of undergraduate statistics students' learning: Lessons learned about surprise. Paper presented at first national meeting of Association for Research in Undergraduate Mathematics Education, San Antonio, TX.

Lesser, L. (1999b). Exploring the birthday problem with spreadsheets. *Mathematics Teacher*, *92*(5), 407-411.

Lesser, L. (2001). Representations of reversal: An exploration of Simpson's paradox. In Albert A. Cuoco and Frances R. Curcio (Eds.), *The roles of representation in school mathematics* (pp. 129-145). Reston, VA: National Council of Teachers of Mathematics. Also available at: www.statlit.org/pdf/2001LesserNCTM.pdf

Lesser, L. M. (2002). Letter to the editor. *Journal of Statistics Education* [Online], *10*(1). www.amstat.org/publications/jse/v10n1/lesser_letter.html

Lesser, L. M. & Pearl, D. K. (2008). Functional fun in statistics teaching: Resources, research and recommendations. *Journal of Statistics Education*, *16*(3), 1-10. www.amstat.org/publications/jse/v16n3/lesser.pdf

Lesser, L. (2009). Sizing up class size: A deeper classroom investigation of central tendency. *Mathematics Teacher, 103*(5), 376-380.

Lesser, L. (2010a). Sizing up class size: Additional insights. *Mathematics Teacher*, *104*(2), 86-87.

Lesser, L. (2010b). Opening intentions. *The Teaching Professor*, *24*(9), 4.

Lyman, F. T. (1981). The responsive classroom discussion: The inclusion of all students. In A. S. Anderson (Ed.), *Mainstreaming Digest* (pp. 109-113). College Park, MD: University of Maryland.

Martin, M. A. (2003). 'It's like ... you know': The use of analogies and heuristics in teaching introductory statistical methods. *Journal of Statistics Education* [Online], *11*(2). www.amstat.org/publications/jse/v11n2/martin.html

McClain, K. & Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics, 45*(1/3), 103–129.

McClave, J. T. & Sincich, T. (2000). *Statistics* (8th ed.). Upper Saddle River, NJ: Prentice Hall.

Minich, B. (2010). *Liberating structures for academic communities: Discovering inclusiveness through dialogue and action*. Philadelphia: Drexel University College of Medicine. www.drexelmed.edu/Portals/0/2010_ForumReport.pdf

Movshovitz-Hadar, N. & Hadass, R. (1990). Preservice education of math teachers using paradoxes. *Educational Studies in Mathematics*, *21*(3), 265-287.

Movshovitz-Hadar, N. & Webb, J. (1998). *One equals zero and other mathematical surprises: Paradoxes, fallacies, and mind bogglers*. Berkeley, CA: Key Curriculum Press.

Movshovitz-Hadar, N. (1988). School mathematics theorems – an endless source of surprise. *For the Learning of Mathematics*, *8*(3), 34-40.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Northrop, E. P. (1944). *Riddles in mathematics*. New York: Litton Educational Publishing.

Perlman, B. & McCann, L. I. (1999). Student perspectives on the first day of class. *Teaching of Psychology*, *26*(4), 277-279.

Romano, J. P. & Siegel, A. F. (1986). *Counterexamples in probability and statistics*. Monterey, CA: Wadsworth.

Rosebery, A. S., Warren, B. & Conant, F. R. (1992). Appropriating scientific discourse: Findings from language minority classrooms. *The Journal of the Learning Sciences*, *2*(1), 61-94.

Rossman, A. (2010). Ask good questions. Paper presented at the 8th International Conference on Teaching Statistics. Ljubljana, Slovenia. http://statweb.calpoly.edu/arossman/AskGoodQuestionsICOTS.ppt

Schwenk, A. (2006). Distortion of average class size: The Lake Wobegon effect. *College Mathematics Journal*, *37*(4), 293-296.

Shaughnessy, M. (1977). Misconceptions of probability: An experiment with a small-group, activity-based, model building approach to introductory probability at the college level. *Educational Studies in Mathematics*, *8*(3), 295-316.

Silver, E. A., Ghousseini, H., Gosen, D., Charalambous, C. & Strawhun, B. T. F. (2005). Moving from rhetoric to praxis: Issues faced by teachers in having students consider multiple solutions for problems in the mathematics classroom. *Journal of Mathematical Behavior*, *24*(3/4), 287-301.

Sowey, E. R. (2001). Striking demonstrations in teaching statistics. *Journal of Statistics Education* [Online], *9*(1). www.amstat.org/publications/jse/v9n1/sowey.html

Stavy, R. & Tirosh, D. (2000). *How students (mis-)understand science and mathematics: Intuitive rules*. New York: Teachers College Press.

Steen, L. A. (2001). *Mathematics and democracy: The case for quantitative literacy*. Princeton, NJ: Woodrow Wilson National Fellowship Foundation.

Stoyanov, J. M. (1987). *Counterexamples in probability*. New York: Wiley.

Székely, G. J. (1986). *Paradoxes in probability theory and mathematical statistics*. Dordrecht: Reidel.

Terwilliger, J. & Schield, M. (2004). Frequency of Simpson's paradox in NAEP data. Paper presented at American Educational Research Association (22 pp.). www.statlit.org/pdf/2004TerwilligerSchieldAERA.pdf

Vogt, W. P. (2007). *Quantitative research methods for professionals*. Boston: Pearson.

von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning*. New York: Routledge Falmer.

Wagner, C. H. (2009). Average perceived class size and average perceived population density. *College Mathematics Journal*, *40*(4), 284-287.

Wilensky, U. (1995). Paradox, programming and learning probability: A case study in a connected mathematics framework. *Journal of Mathematical Behavior*, *14*(2), 253-280.

Williams, A. S. (2010). Statistics anxiety and instructor immediacy. *Journal of Statistics Education*, *18*(2), 1-18. www.amstat.org/publications/jse/v18n2/williams.pdf

Wilson, J. H., Stadler, J. R., Schwartz, B. M. & Goff, D. M. (2009). Touching your students: The impact of a handshake on the first day of class. *Journal of the Scholarship of Teaching and Learning*, *9*(1), 108-117.

Wilson, J. H. & Wilson, S. B. (2007). The first day of class affects student motivation: An experimental study. *Teaching of Psychology*, *34*(4), 226-230.

Wise, G. L. & Hall, E. B. (1993). *Counterexamples in probability and real analysis*. Oxford: Oxford University Press.

Zaslavsky, O. (2005). Seizing the opportunity to create uncertainty in learning mathematics. *Educational Studies in Mathematics*, *60*(3), 297-321.

Lawrence M. Lesser
Department of Mathematical Sciences
The University of Texas at El Paso
500 W. University Avenue
El Paso, TX 79968-0514
mailto:Lesser@utep.edu

Kerrie Kephart
Department of Human Centered Design and Engineering
Box 352315
University of Washington
Seattle, WA 98195
mailto:Kkephart@uw.edu