A Course on Teaching Statistics at the University Level

Andrew GELMAN

Teaching is an important part of the jobs of statistics professors and graduate students, but most of us receive little training for teaching in general or statistics teaching in particular. We have developed a course for statistics graduate teaching assistants that combines practice in statistics demonstrations and drills, discussion of teaching strategies, and feedback on classroom teaching. The goal is to get our graduate students to be more comfortable with engaging their students in active learning.

KEY WORDS: Active learning; Demonstrations; Drills; Graduate training; Statistics education.

1. THE PROBLEM

Our students do a lot of teaching during their four or five years in the Ph.D. program, typically first as graders, then teaching assistants, and finally sometimes as instructors in their own courses. The statistics department at Columbia teaches three introductory service courses (each with multiple teaching assistants) and a range of advanced undergraduate, M.A., and Ph.D. courses, almost all of which have teaching assistants. Upon graduation, many of our Ph.D. students move straight into jobs that require them to teach two or more courses per semester.

Unfortunately, doctoral students in statistics typically have very little preparation and receive little or no training in teaching. Their most common source of preparation is simply their experiences as students themselves, and their most common training is on the job, with the only feedback being their students' direct reactions and end-of-semester student evaluations. Adding to the difficulty is that our Ph.D. students generally were top performers in their classes and are very strong mathematically, so it can be difficult for them to relate to the concerns of the students they are teaching, who generally are majoring in fields such as economics, political science, and psychology, where mathematics backgrounds are mixed. In addition, most of our teaching assistants come from other countries and often have difficulty communicating fluidly in English.

2. THE NEW COURSE

To address some of these problems, we developed a new course, "The Teaching of Statistics at the University Level," to train our teaching assistants and instructors. This course, offered in the fall semester, is required of our first-year Ph.D. students, and has the following components:

- informal instruction in teaching techniques;
- group discussion of issues in teaching and learning;
- practice in conducting drills, demonstrations, lectures, and other classroom activities;
- review of research on effective teaching methods.

We use three textbooks: a short guide for graduate student teachers (Curzan and Damour 2000), a general book on college teaching (McKeachie 2002), and a book focused on class-participation activities for statistics courses (Gelman and Nolan 2002). The McKeachie book reviews research on effective teaching methods, and we go over these findings during our class discussions.

We have just finished teaching the course for the second time. Last year we met weekly for one hour, but we have increased that to two hours, combining our class with the regular meeting that was already being held for the teaching assistants of our largest introductory course. The course is team-taught by two instructors: a senior faculty member and the lead instructor for our largest introductory statistics class, many of whose teaching assistants are students in the course.

Each class period has three components: a discussion of general issues in teaching, practice in a particular technique in the teaching of introductory statistics, and a segment focused on their particular experiences in the past week and plans for next week's classes. The discussion of teaching goes through the chapters of Curzan and Damour (2000) and McKeachie (2002), covering practical advice and also reviewing research findings on teaching methods. For the statistics-specific material, we work through Part 1 of Gelman and Nolan (2002), following the sequence of topics in an introductory class. The syllabus of the course can thus be roughly derived from the tables of contents of these three books.

3. FIRST DAY OF THE TEACHING-STATISTICS CLASS

It is generally advised (see, e.g., McKeachie 2002) to begin a course with interesting material, not simply with a dry introduction, and we follow that practice ourselves. The following is an idealized description of the first day of class, omitting the inevitable slip-ups that occur in any real classroom setting (and can be valuable as opportunities for discussion).

As the students enter the classroom, we introduce ourselves to them informally and discuss what is expected from the course. We also quietly pass a sealed envelope to one of the students and ask him or her to save it for later. Then, when they have all arrived, the instructor pulls out a digital kitchen scale and a

plastic bag full of a variety candies of different shapes and sizes, and announces:

> This bag has 100 candies, and it is your job to estimate the total weight of the candies in the bag. Divide into pairs and introduce yourself to your neighbor. [At this point we pause and walk through the room to make sure that all the students pair up.] We're going to pass this bag and scale around the room. For each pair, estimate the weight of the 100 candies in the bag, as follows: Pull out a sample of 5 candies, weigh them on the scale, write down the weight, put them candies back in the bag and mix them (no, you can't eat any of them yet!), and pass the bag and scale to the next pair of students. Once you've done that, multiply the weight of your 5 candies by 20 to create an estimate for the weight of all 100 candies. Write that estimate down (silently, so as not to influence the next pair of students who are taking their sample). [As we speak, we write these instructions as bullet points on the blackboard: "Draw a sample of 5," "Weigh them," etc.]

> Your goal is to estimate the weight of the entire bag of 100 candies. Whichever pair comes closest gets to keep the bag. So choose your sample with this in mind.

We then give the bag and scale to a pair of students in the back of the room, and the demonstration continues while the class goes on.

At this point, we pass out handouts describing the course, introduce ourselves, and tell them that, in preparation for a discussion we will be having in about 15 minutes, they should each think of a story from their experiences as students or teachers that illustrates some aspect of good or bad teaching. We then spend several minutes introducing (and answering questions about) the structure and goals of the course.

At this point, the candy weighing will probably be done (for a class of 50, the weighing takes much of a class period, but our teaching class has only about 10 students and so the demonstration goes pretty fast), and we bring the class back to it by asking each pair to give their estimate of the total weight, which we write, along with their first names, on the blackboard. We also draw a histogram of the guesses, and we ask whether they think their estimates are probably too high, too low, or about right.

We then pass the candies and scale to a pair of students in front of the class and ask them to weigh the entire bag and report the result. Every time we have done this demonstration, whether with graduate students or undergraduates, the true weight is much lower than most or all of the estimates—so much lower that the students gasp or laugh in surprise. We extend the histogram on the blackboard to include the true weight, and then ask the student to open the sealed envelope, and read to the class the note we had placed inside, which says, "Your estimates are too high!"

We conclude the demonstration by leading a discussion of why their guesses were too high, how they could have done their sampling to produce more accurate guesses, and what analogies can they draw between this example and surveys of human populations. When students suggest doing a "random sample," we ask how they would actually do it, which leads to the idea of a sampling frame or list of all the items in the population.

We then step back and ask the students, as potential teachers, what statistical principles arose in the demonstration (random sampling, bias, and variance of estimation) and what teaching principles arose (learning by doing, learning by making a mistake, personal involvement, working in pairs). We then explain

that this is the sort of demonstration that we would like them to do in all their recitation/discussion sections—and, as we have illustrated, much of the demonstration can be performed in parallel with other classroom activities.

We continue the first day of class by going around the room and sharing personal stories of good and bad teaching. This allows the students to introduce themselves and gives us a sense of their concerns. We can also get useful ideas for our own teaching. For example, from one student,

> One of my favorite profs in college taught an engineering class with 300 students. On the very first day, he told us that during the lectures, he'd stop sometimes and call on students: if we sat on the left side of the room or the right side of the room, he wouldn't call on us, but for the students who sat in the middle, he'd keep a list and call on students in order—so you'd know when you were going to be called—and then you'd have to answer or you could say "pass," and the question would go on to the next student.

> At the beginning of the class, there was a lot of passing, but as the semester went on, more and more people sat in the middle third of the classroom, and people got to be pretty good about answering.

We conclude the first class period with a drill—a structured series of short problems that students should be able to solve almost automatically (once they have mastered the relevant course material). Figure 1 gives an example of a drill on confidence intervals and hypothesis testing for a proportion.

For the introductory statistics class, the drill is intended to give the students a chance to develop competence in a particular skill. For the teaching-statistics class, the purpose of the drill is to demonstrate this particular classroom activity to the graduate students. We have found that, in the absence of a detailed example of a drill, our graduate teaching assistants tend to lecture—even when we instruct them ahead of time to devote the class period to student problem-solving. This is why we include Figure 1 here, in all its detail—as an example of an entire drill which can be shown to our teachers in training.

We perform the drill for the teaching class—first having them work in pairs on the basic problem, then rapidly going through the class one at a time to answer the drill problems. Even for such simple problems, the students will make mistakes at first, but by the end of the drill they are warmed up and have the answers at their fingertips. We then step back and discuss the use of these drills in statistics recitation sections, and then send them off to their own classes.

4. THE REST OF THE COURSE

Each week, we continue the general pattern of introducing a demonstration and a drill, along with class discussion on a weekly topic chosen from the McKeachie and Curzan/Damour books. During the first few weeks, we have found it necessary to conduct the demonstrations and drills ourselves, in order to give a template for the graduate students in the class to imitate.

After the first third of the semester or so, we gradually shift the responsibility to the students. Each week, we pick a demonstration or class activity from Gelman and Nolan (2002) corresponding to the material that is about to be covered in the introductory statistics class, and we assign two of the students to prepare the demonstration for our teaching class. Having learned

Basic problem: 100 people are sampled at random and asked if they approve of George Bush's performance as President. 55 say Yes and 45 say No.

- Give a 95% confidence interval for the population proportion who approve of Bush's performance.
- Is the approval for Bush statistically significantly different than 50%?

Drill problems:

For each:

- Give a 95% confidence interval for the approval for Bush.
- Is the approval for Bush statistically significantly different than 50%?

1. 100 people are surveyed. 50 say Yes and 50 say No.
2. 100 people are surveyed. 60 say Yes and 40 say No.
3. 100 people are surveyed. 40 say Yes and 60 say No.
4. 100 people are surveyed. 30 say Yes and 70 say No.
5. 100 people are surveyed. 90 say Yes and 10 say No.
6. 100 people are surveyed. 0 say Yes and 100 say No.
7. 400 people are surveyed. 240 say Yes and 160 say No.
8. 1000 people are surveyed. 600 say Yes and 400 say No.
9. 10,000 people are surveyed. 6000 say Yes and 4000 say No.
10. 1 million people are surveyed. 600,000 say Yes and 400,000 say No.
11. 10 people are surveyed. 6 say Yes and 4 say No.

Exam-type problems:

1. 1000 people are sampled at random and asked if they approve of Bush's performance. X say Yes and $1000 - X$ say No. For what values of X will the approval for Bush be statistically significantly different than 50%?
2. A pollster will sample N people at random and ask if they approve of Bush's performance. The pollster wants the margin of error of the 95% confidence interval to be no more than 5%. How large a sample size N does the pollster need?

Discussion questions:

1. What do you do if some people don't respond to the survey or say "don't know"?
2. How would you actually take a random sample of 1000 Americans?

Figure 1. Drill questions on confidence interval and hypothesis test for a single proportion. We use this as an example of a drill for use in an introductory statistics class. Students are first given the basic problem on the blackboard and given a few minutes to work in pairs. The instructor then writes the solution on the board (ideally, adapting a solution from one of the students) and passes out this sheet. The students are given a couple minutes to digest it, and then the instructor quickly goes through the drill problems, moving through one student for each problem. The students can then be given some time to work in pairs on the exam-type problems and, if there is time, discuss the questions at the end.

from their mistakes, all the students in the class are then set to try the demonstration in their own sections. We also pick two students each week to construct a drill, based on the material to be covered in two weeks in the introductory class, and then demonstrate it in the next week's teaching class.

In addition, for each of our students who is currently teaching, we sit in on his or her section once during the semester and provide feedback. Typically, our teaching assistants spend too much time lecturing and too little time working with students in groups. We take notes during the section and then discuss our positive and negative observations during the next meeting of the teaching class. Comments range from specifics of particular lessons (e.g., demonstrations that are too hurriedly introduced, or drills in which the teaching assistant forgets to give students time to work in pairs) to more general comments (e.g., when a student is working alone, the teaching assistant should call him

or her by name and suggest another student nearby for him or her to work with).

Thus, the course naturally structures itself based on the sequence of topics in the introductory courses for which the students serve as teaching assistants. There should also be time throughout to discuss particular difficulties that arise during the semester. By the end of the course, we want our graduate students to be able to teach statistics with active student participation, and to be prepared to handle the practical difficulties that arise in teaching.

5. GROUP DISCUSSIONS

In our class, as with college teaching in general, it can be a challenge to get students to participate! We have used this struggle as a teaching opportunity and use strategies such as having students work in pairs (with at least one member from each pair required to participate in a discussion) and using students in

the class as discussion leaders. Discussion topics throughout the semester include:

- What are good and bad teaching styles?
- How to encourage student involvement and class participation.
- Testing and grading.
- Teaching to individual students.
- Particular issues in the teaching of statistics.
- Difficult situations with students.

For each topic, we ask the students to come to class prepared with anecdotes from their own experiences, and these lead to wide-ranging discussions.

We also discuss some issues that are specific to the introductory statistics course at Columbia, including our use of quizzes to motivate attendance in section and challenges in supervising student data-collection projects. Later on in the semester, we discuss issues in teaching more advanced courses and preparing one's own class.

6. PRACTICE IN TEACHING

Our students' inclination in teaching is to lecture, especially given their experiences of strongly lecture-based courses in other countries, and we tried to counteract this in the course by focusing on practice of more interactive teaching methods. As with the first day of class, discussed above, we present at least one demonstration and at least one drill each week and discuss how these fit in with that week's material in the introductory statistics course. Later on in the semester we assign pairs of students to give these presentations.

In addition, each week the students are required to perform at least one demonstration and at least one drill in their own weekly section meetings, which are problem-solving sessions with between 10 and 60 students that supplement the large introductory statistics classes. (For students in the class who are not TAs for introductory statistics, we set up a schedule in which they substitute-teach in others' sections.) We assign demonstrations from Gelman and Nolan (2002) and provide drills in handout form such as in Figure 1. We instruct them to spend the remaining time in their sections answering questions (typically about past and current homework assignments) and supervising students as they work in groups on homework problems.

Near the end of the semester, we practice introductory statistics lectures and presentations in research seminars, going over practical issues in presentations as well as general discussions of the role of teaching and research in the career of Ph.D. graduates from our department. We present teaching as a unifying theme in an academic career, comprising oral presentations and published research papers as well as classroom teaching, grading, and advising of students.

7. SUMMARY

We have so far taught The Teaching of Statistics at the University Level twice, to classes of about 10 Ph.D. students (most of whom had been in the United States less than two years), whose reactions were generally positive. We hope to do better this year by making the course goals clearer and giving examples of discussion-based section teaching, so that students can get more practice with these methods in a supportive environment with direct feedback from their colleagues.

In our own teaching, we have found class-participation demonstrations to be so convenient and useful that we include several demonstrations per week in our introductory courses. Our new teaching assistants typically resist performing demonstrations or other class-participation activities, fearing that they will not have time to cover important material. We hope that requiring the teaching assistants to practice these (from Gelman and Nolan 2002) for their actual introductory courses has the potential to relieve their anxiety and free them to include student-participation activities among their teaching tools. We also hope that a structured series of drills will make the introductory statistics recitations easier to teach and also more effective.

As in so many efforts by statisticians to improve their teaching, we offer no systematic evidence of the success of our approach, either as a before/after study or, more appropriately, compared to reasonable alternatives. We apologize for this and will think more how to evaluate our approach. But in the meantime, our positive anecdotal experience motivates us to suggest that others consider such a course for their graduate statistics programs. Certainly the literature on college teaching [summarized in McKeachie (2002) and elsewhere] supports the promotion of techniques for active learning.

REFERENCES

Curzan, A., and Damour, L. (2000), *First Day to Final Grade: A Graduate Student's Guide to Teaching*, Ann Arbor: University of Michigan Press.

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

McKeachie, W. J. (2002), *Teaching Tips* (11th ed.), Boston: Houghton Mifflin.