TRAINING GRADUATE STUDENTS TO TEACH STATISTICS AND DATA SCIENCE FROM A DISTANCE

Wendy Rummerfield, Federica Zoe Ricci, and Mine Dogucu
Department of Statistics, University of California, Irvine USA
wrummerf@uci.edu

Enrollment in undergraduate statistics and data science courses has rapidly increased in just the last decade, resulting in an increased reliance on graduate teaching assistants (GTAs) and graduate instructors of record (GRIs). In the age of the COVID-19 pandemic, teaching from a distance has become a necessity. Many instructors, including GTAs and GRIs, need to adapt to new technologies and reconsider pedagogical decisions. This paper presents our experiences from a graduate teaching fellowship program created because of the pandemic. The program had two major components: 1) pedagogical workshops attended by teaching fellows from multiple disciplines across the university and 2) one-on-one mentoring by a faculty member from the fellow's primary discipline. Here, we provide a unique look at graduate training from both the perspective of the mentor and the mentee. We share a sample training curriculum and propose recommendations for those interested in implementing teaching training opportunities for graduate students.

INTRODUCTION

More and more graduate programs worldwide are hiring graduate student instructors (GSIs) and graduate teaching assistants (GTAs) to, respectively, teach or assist in teaching introductory statistics courses. In fact, Blair, Kirkman, & Maxwell (2015) reported that across all levels of statistics departments in the United States alone, GSIs teach about 31% of introductory statistics courses. This may be due to increasing financial burdens on institutions or departments due to the uptick of enrollment in statistics and data science courses over the last few decades. In our department, GTAs are assigned to teach discussion sessions that complement lectures taught by the instructor. GSIs are hired as the instructor of record typically during the summer terms.

Literature on GTAs/GSIs in the last decade is relatively sparse, especially as it concerns distance learning. Still, a few studies on the subject reveal that current training is lacking, specifically in discipline-specific content knowledge and recommended pedagogical practices (Blair, Kirkman, & Maxwell, 2015; Justice, Zieffler, & Garfield, 2017). Further, as technology and statistical software advance, preparation for these graduate students must remain current. In light of the surge in remote learning caused by the coronavirus pandemic, the importance of up-to-date preparation for (statistics) GTAs/GSIs has become even more apparent. With remote classes came a heavy reliance on technological tools, such as learning management systems (LMS) and synchronous online video communication software. With this new set of software (and hardware) came a novel set of innovative pedagogical methods.

In response to the switch to online learning, the Graduate Division at the University of California, Irvine (UCI) created the Division of Teaching Excellence and Innovation (DTEI) Graduate Summer Fellowship. Approximately 300 graduate students, from 13 different schools, participated in the 2020 DTEI summer fellowship; four of them were from the Statistics department. This fellowship provided graduate students with some summer funding as there were fewer available summer internship opportunities due to the state-mandated quarantine, and many of these students do not have summer funding. The first of its kind at any of the UC campuses, the purpose of this fellowship was to train graduate students (Masters and Ph.D.) in online tools for instruction, which could be taken back to departments to assist faculty with the switch to distance learning. In addition to helping faculty convert their materials to an online format, the program gave fellowship awardees a chance to develop many academia-oriented and practice skills. Topics discussed and focused on were: the development of course spaces on the LMS; lecture video construction; transitioning assessments to a remote format; forming activities to facilitate student/instructor interaction; student engagement, and many other pedagogical concepts.

Typically, each school in the university offers a mandatory training for new graduate students called the Teaching Assistant Professional Development Program (TAPDP) and requires completion of a 4-credit course for teaching assistants. While this course covers skills needed for an in-person class, the DTEI fellowship was explicitly designed for training in teaching from a distance (amidst a

In: R Helenius, E Falck (Eds.), *Statistics Education in the Era of Data Science*
Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),
Aug-Sept 2021, Online conference. iase-web.org/Conference_Proceedings.php

global pandemic). This included lessons on the features of various technological tools as well as sensitivity and diversity training. Combined and shared amongst the individual departments of the fellows, these competencies helped provide students (and faculty) with a sense of normalcy and structure during a time filled with turmoil. The goal of mastering the new features and nuances inherent in the virtual classroom, was to provide learners with the same standard of education on which the university prides itself.

The program included two components: discipline-agnostic and discipline-specific training. DTEI facilitated the former, which fellows from different disciplines across campus attended. The latter was directed under the supervision of an instructor (mentor), typically working in the same department as the fellow (mentee). Also unique to this program at UCI was the continued communication and assistance provided by the DTEI fellowship teaching team after the discipline-agnostic training concluded. Fellowship recipients and their mentors were encouraged to reach out to the DTEI teaching team in the second half of the training. They conducted additional workshops and provided a platform for GTAs/GSIs from across the campus to share ideas and ask questions.

In this paper, the authors - two graduate student fellows and their program adviser from the statistics department at UCI - discuss the efficacy of a university-wide training designed in the wake of the pandemic and comment on lessons learned and recommendations for the future.

DISCIPLINE-AGNOSTIC TRAINING

DTEI provided the discipline-agnostic training in two ways: 1) a one-week remote course taught during the first week of the program (June 22-26), and 2) a series of three one-hour workshops held throughout the summer in 2020.

The week-long training delivered at the program's onset consisted of one-hour daily workshops (Monday to Friday), complemented by additional materials for review and assignments to be submitted on Canvas - the LMS used at our institution. DTEI taught the workshops synchronously online via a streaming platform, and fellows were required to either attend them live or to watch the video recordings by the end of the week. Online assignments accompanying each module were also due at the end of the week. These aspects of flexibility were essential features that ensured the training would be accessible to its body of diverse fellows - graduate students from all departments, at any year in their programs, and possibly living in different time zones. For instance, one of the mentees authoring the current paper had qualifying exams during the same week as the training, and these accommodations allowed her to manage both more easily.

Similarly, DTEI taught the three workshops held throughout the summer synchronously, but the recordings were available for trainees to watch asynchronously. Required and optional activities available on the LMS complemented these workshops as well.

DTEI appeared to have designed the training with both short- and long-term goals in mind. In the short term, the program aimed to help GTAs improve their discussion sessions, especially aspects that concerned remote instruction. A longer-term objective was to train graduate students to be the future instructors of record.

The material covered during the discipline-agnostic part of the training had pedagogical and technological elements. The pedagogical training covered many topics such as course design (e.g., backward design), classroom activities (e.g., setting clear goals, learning objectives, and learning outcomes), and techniques for active learning. The instructors also lead discussions on diversity and shared concrete examples of what makes a learning environment inclusive. On the technical side, the workshop discussed how to use certain features of the LMS to help simplify course site navigation and create online learning activities. The ultimate goal was to train faculty and other GTAs/GSIs on how to move existing in-person educational activities to the online space. Additionally, the workshop covered available software and hardware that instructors could use to increase engagement during synchronous lectures and tools (some of which were available through the university) for asynchronous lecture recording and editing.

DISCIPLINE-SPECIFIC TRAINING

During the application and recruitment phase of the fellowship program, DTEI asked for graduate students to find a mentor and list a specific course that they would assist in preparing. In our case, one mentee chose the lower-level Introduction to Data Science (https://www.introdata.science/)

undergraduate course, while the other mentee chose the upper-level Introduction to Bayesian Data Analysis (https://www.stats115.com/) undergraduate course. Before the pandemic, both mentees had experience working as course readers, GTAs, and one mentee as a GSI. They were already proficient in using R and R Markdown, the school's online grading system (Gradescope), and the LMS (Canvas). Both mentees were also familiar with class management and had a strong knowledge of course content. Considering the mentees' prior experiences, we focused on discipline-specific training that would be more relevant to teaching during the pandemic.

We considered three main discipline-specific components: technological, pedagogical, and content. For the technological element of the training, we focused on creating our educational resource web pages using HTML and CSS. As the switch to remote learning eradicated on-paper handouts and assignments, it was imperative to take into account that many students did not have access to printers and, therefore, make as much content available online as possible. This also included using Git and GitHub for version control and collaboration to create web pages and educational tools. Perhaps, the most substantial part of the training was recording and editing videos. The mentor and mentees learned this together. Some of the highlights from what we learned about video production include attention to lighting and sound and, of course, editing software. In the spirit of learning-by-doing, we worked both separately and together to experiment with various editing techniques. We shared our discoveries with each other and later with other faculty and GTAs/GSIs via an instructional video in a blog post.

The pedagogical element of the training focused on student engagement, which includes behavioral, emotional, and cognitive components. We have done an extensive literature review to learn about best practices and how to incorporate these into our teaching. Through this learning opportunity, we have included short polls in our zoom sessions to make learning be more interactive and engaging. We again provided a summary of these practices and how to implement these changes in a blog post.

We have seen the outcomes of different components (pedagogy, technology, content) of the program reflected in our teaching during the academic year. For instance, while learning about the technical tools on video recording and editing, we also learned about how to keep students focused as they watch the video by keeping videos short and concise and supplementing the videos with questions. Most importantly, the training helped us have a better understanding of students' challenges with the pandemic and its reflections on their learning.

Finally, we worked to create activities with captivating material that would be most beneficial to the isolated learner. Each mentee prepared an activity on a selected topic, chosen among those that students typically struggle to understand in introductory statistics classes. One of the activities focused on maximum likelihood estimation and the other on variance-covariance matrices. The educational materials developed were prepared in a format (e.g., online quizzes) that would be suitable for use in the online courses that the mentees would assist with during the coming 2020-2021 academic year. In all activities, we integrated an ample number of exercises, diagrams, and figures to break up any lengthy text.

There are many different forms of training programs for GTAs/GSIs to consider (e.g., seminars, courses, workshops, etc.) that can be accessible to all students. What set this fellowship program apart was that the selected group of fellows who received the broader training then tuned the teachings to be more applicable to a specific program. Fellows then shared this knowledge with peers and faculty members in the department. This was especially important as the shift to remote teaching incited a need for discipline-specific training for most students and faculty in the department. We chose to do this through a series of blog posts and an instructional video. These were shared with GTAs and instructors of our department, school, and the general public. This material is publicly available on https://www.datapedagogy.com/#category:remote_teaching.

DISCUSSION

Overall, in our opinion, the DTEI Summer Fellowship program was a huge success - even more so considering the short amount of time the DTEI and mentors had to prepare for the training. It is important to remember that this quickly put-together training took place near the onset of the COVID-19 pandemic. Instructors and GTAs/GSIs were forced into the world of distance learning with little experience, whereas mentees in future programs will have accumulated much more knowledge of

remote instruction. Nonetheless, having completed the program and subsequently taught and assisted in at least three virtual classes, in this section, we will present a synopsis of the benefits and challenges of the fellowship program and share recommendations.

We first would recommend the dual training scheme adopted in the program. Based on our training and teaching experiences, we provide a list of topics in Table 1 that could potentially be adopted in such programs. The discipline-agnostic part of this format can help save time and resources and allows GTAs/GSIs from all departments to come together to learn about the more universal (online) pedagogical methods. Even more, this gives graduate students the rare opportunity to interact with colleagues from outside their field of study. Once students have been trained on the more general fundamentals of remote learning, fellows will go back to their own departments and customize the material to fit their discipline better. Unlike in workshops, the mentorship model in the discipline-specific portion allows participants to produce usable material (assignments, blog posts, etc.) and get immediate feedback from a mentor. If the discipline-specific training is not *course-specific*, however, one could use introductory course topics to start. We especially recommend topics that are often difficult for students to understand or for graduate students to teach.

As the restrictions of the COVID-19 pandemic begin to lessen and society gets accustomed to a new normal, online learning will likely have a different place in education, especially at the post-secondary level. Thus, knowing the software available to you through your institution (and/or through your own means) will be particularly helpful. This training program could also be beneficial because it pushes instructors and students to stay up to date with the latest technological tools.

Very importantly, we understand that this model may not be easily scalable at all institutions. Especially with the discipline-specific portion of the program, training such as this can be costly in many aspects. It may even be hard to find interested and available faculty or mentors to meet the demand. One option to ease some of this burden is to train a smaller group of highly motivated graduate students who can assist other students or faculty. Peer mentoring is recognized as a helpful socialization process. Trainees often look to their peers for information regarding the expectations and standard practices of GTAs in the department (Justice, Zieffler, & Garfield, 2017). The advantages of peer mentoring go beyond the efficient use of resources and community-building. It also provides excellent opportunities for peer mentors to take on leadership roles and build their curriculum vitae (fellows in our program earned certificates in remote instruction).

Even more, the benefits of a teaching training program can extend beyond graduate school. Programs like this can help graduate students prepare for teaching-focused careers, which can be quite fulfilling for statisticians (Dogucu, 2020). For students who pursue research-focused academic careers, teaching is also likely to be an essential portion of their job. Those outside academia often need to present ideas, mentor junior colleagues and communicate statistical ideas to non-statisticians. These crucial skills have many overlapping components with teaching skills.

Table 1. A Sample of Topics to Cover in GTA/GSI Training

	Discipline-Agnostic	Discipline-Specific (Statistics and Data Science)
Pedagogy	Curriculum development Backwards design Goals and SLO/o's Bloom's taxonomy Student-centered teaching Inclusivity Microaggressions Understanding Flexibility with course policies & assignment format Accessibility: font style, color, captions	Behavioral, emotional, and cognitive engagement Consistent weekly schedule Use diverse, real-world examples Active participation GAISE recommendations Emphasize conceptual understanding Use real data Active learning Utilize technology for data analysis & visualization
Technology	LMS Website organization Creating assignments	Tools for literate coding and reproducibility RStudio/R Markdown Best practices for live coding

	Discussion boards Virtual communication software Capabilities: video, audio, chat, breakout rooms, polls, attendance, virtual backgrounds, screen sharing, etc. Accessibility: recording and live captioning Video production & editing Equipment Lighting, sound & video quality, length, etc. Captions	HTML/CSS Version control Git/GitHub Online grading tools (e.g., Gradescope) Programming or mathematical assignments Remote (mathematical) writing Document-camera Tablet
Content	One day separate STEM vs. non-STEM workshops STEM course might discuss recording or sharing drawings, calculations, or writing code (a)synchronously	Ensure GTAs have a comprehensive understanding of topics, not just procedural knowledge (e.g., p-values) Focus on statistical methodology that can be helpful pedagogical tools (e.g., simulation methods for inference, big data)

Beyond the pedagogical skills, the program utilized data science tools such as R Markdown, git, and GitHub for educational purposes. These tools were used for the purposes of preparing educational materials, such as preparing tutorials and exercises using the R Markdown-based learnr package, git, and GitHub for collaboration within the teaching team as well as dissemination. Even if graduate students are trained in using these tools for conducting research, we recommend additional training to expose them to their use in statistics and data science education as they are some of the most commonly used software in statistics and data science. Beyond teaching and research, students can use these tools to make personal websites and blogs which can be helpful for early-career professionals gaining visibility in their career paths (Dogucu, 2021).

In conclusion, roughly a year after the training, we have already seen many benefits of the fellowship program focused on teaching from a distance. Though we can only speak for the mentees who are authors of this paper, we are quite confident that many of the fellows in the program learned how to utilize new software (and even hardware) for use in the classroom. We, specifically, have also started incorporating open-source software into our own teaching, but also into our research; in fact, one of the mentees transferred her entire dissertation over to a GitHub repository. Additionally, the fellowship recipients also emphasized a vital part of teaching: inclusivity and diversity. The training increased our awareness and sensitivity to the diversity and needs of the student body. This has forever impacted our teaching as we work to create a safe teaching environment where all students have an equal opportunity to receive the same education.

We expect these skills to outlast the pandemic and the graduate school experience of the mentees. Therefore, we advocate for teaching training - whether through coursework or workshops - to be part of every student's Ph.D. preparation. We recommend other institutions that implement such programs offer both discipline-agnostic and discipline-specific components. It is important to remark that what we presented on these components focuses on a specific set of resources (e.g., R); however, the overall scheme could also be implemented using many other tools (e.g., Python and Julia).

REFERENCES

Blair, R. M., Kirkman, E. E., & Maxwell, J. W. (2015). *Statistical abstract undergraduate programs in the mathematical sciences in the United States: 2015 CBMS survey.* American Mathematical Society.

Dogucu, M. (2020). Teaching Careers (for Statisticians): What You Should Know. *Amstat News, 521*, 32-34.

Dogucu, M. (2021). Contributing to Open Education: Why, How, and What I am Doing. *AMS Notices, 68*(3), 367-369.

Justice, N., Zieffler, A., & Garfield, J. (2017). Statistics graduate teaching assistants' beliefs, practices and preparation for teaching introductory statistics. *Statistics Education Research Journal, 16*(1).