

Analyzing Undergraduate Statistics Majors' Preparation in Communication with Non-Statisticians in the University of California, Berkeley

Yehchan Yoo

Writing statistical reports for non-statisticians is a crucial part of a statistician's job, so it is critical for undergraduate statistics courses to prepare statistics majors for such communication with non-statisticians. To check how much preparation the undergraduate statistics majors were receiving in statistical writing for non-statisticians in University of California, Berkeley, I surveyed members of the Statistics Undergraduate Student Association, gave an online questionnaire to a graduate statistics student who was an undergraduate statistics major in the university, and interviewed two statistics professors. The survey, the questionnaire, and the interviews indicated that – while undergraduate statistics majors did receive a lot of training in statistical writing in their undergraduate statistics major classes – they did not get much practice doing statistical writing with non-statisticians as their general audience from the classes. While the interviews indicated that the Department of Statistics was hesitant to enforce stricter writing requirements for statisticians due to a lack of resources and a desire to attract more students into the department's programs, there overall still existed a need for the department to devote more resources to statistical writing education.

KEY WORDS: statistical writing, statistical education, target audience, report

1 INTRODUCTION

Statistician Ken Williams (1972, p. 127) once wrote in *Journal of the Royal Statistical Society*:

Yehchan Yoo is a first-year undergraduate student at University of California, Berkeley. He is an international student from South Korea and intends to double-major in statistics and political economy.

Series D (The Statistician), “The statistician...must be continually aware of the need to communicate the results of his expertise in an intelligible manner to the customer. If he fails to do so the lack of numeracy of his reader is no justification for his failure”. Being a statistician often involves communicating the statistical meanings behind certain data to non-statisticians (“A Day in the Life of a Statistician” 2018). After all, statisticians assist in research for a variety of academic fields and industries outside of statistics, such as biology, economics, chemistry, and entertainment (“A Day in the Life of a Statistician” 2018). So, it is crucial for prospective statisticians to learn how to write statistical reports for non-statistician readers before heading into academia or industry.

Unfortunately, it was unclear whether undergraduates statistics majors in the University of California, Berkeley, were learning how to communicate statistical findings to non-statisticians through writing. Some statistics classes did force students to practice statistical communication. For instance, in Stat C8, past class assignments and assessments included problems that required students to explain the validity of their statistical analysis and knowledge (Adhikari 2016a,b; “Homework 7: Correlation, Regression, and Least Squares” 2017; “Project 2: Inference and Capital Punishment” 2018). In a past Stat 157 class, each student was required to submit a write-up on a research (Aldous 2017).

However, I was uncertain whether every statistics major took such writing-intensive statistics courses. Every statistics major in the University of California, Berkeley, currently needs to take Stat 133, Stat 134, Stat 135, Stat 140, and three classes in the Stat 15x class series to graduate with a Bachelor's degree in statistics; however, a statistics major could avoid writing-intensive classes like Stat 157 and take other classes in the Stat 15x class series

(“Statistics Major” 2018). Every statistics major who entered the university on Fall 2018 or after also may need to take Stat C8 or Stat 20; but, even here, a prospective statistics major can take Stat 20 rather than the writing-intensive Stat C8 (“SUSA Enrollment Advising for Fall 2018” 2018). From here, I questioned how much training the statistics majors in the University of California, Berkeley, were really receiving in communicating their statistical findings to non-statisticians from undergraduate statistics major courses.

2 METHODOLOGY

2.1 Survey of Undergraduate Students

As a member of the Statistics Undergraduate Student Association, I first decided to survey students in the association to get a general grasp of how much declared or intended statistics majors in the university actually learned about written statistical communication with non-statisticians. I used Google Forms to create the survey; I then distributed the survey link to all members of the association through the association’s Slack workspace’s general channel at 5:04 P.M. on April 29.

To ensure that only the answers of the Statistics Undergraduate Student Association members would be used for my research, the survey was made accessible only to people with Google accounts associated with the University of California, Berkeley, and automatically collected the university Google email addresses of all surveyees. The email addresses of the surveyees were then checked through the Statistics Undergraduate Student Association’s email address list to verify if the surveyees were in the association. The survey also allowed only one answer from each respondent to prevent anyone from providing multiple redundant responses for the survey; the survey also did not allow the respondents to edit their responses after submission, primarily to prevent the surveyees from repeatedly changing the answers during my analysis of the survey answers for the research presentation in the Chiang Research Festival on May 1. However, the surveyees were allowed to get emails of their responses for the survey if they wished.

I stopped accepting responses for the survey at 5:00 P.M. on May 3 to allow myself time for analysis of the survey responses before the research paper submission on May 7. While the survey was initially planned and announced on the survey form and on

the general Slack channel to close down at 3:00 P.M. on May 3, the lack of responses – especially from the end of April 30 to the end of May 2 – prompted me to delay the survey close time to 5:00 P.M. to allow for more responses. I announced this delay in survey close time on the general Slack channel at 12:30 P.M. on May 2.

The survey consisted of five sections: Basic Information, Academic Information, Course Experience, Writing Experience Outside the Courses, and Handling Your Information. Basic Information section contained questions on the surveyees’ pre-Berkeley background to potentially assess how much the surveyees’ pre-Berkeley education may have helped them with their statistical writing. Academic Information section contained questions on the academic disciplines the surveyees’ were studying and the surveyees’ outlook on what they planned to do after graduating from the university to assess the amount of motivation the surveyees might have in learning to do statistical writing. Course Experience section contained questions on the statistics major courses the surveyees might have taken in the university to assess the types and target audiences of the writing the surveyees did in their statistics courses; Writing Experience Outside the Courses section contained questions on the surveyees’ experiences or outlook on statistical writing outside the university statistics courses.

At the end of the survey, surveyees answered questions in the Handling Your Information section regarding whether they were fine with their names and/or individual responses being featured in the research paper submission. Surveyees who submitted the survey before 12:00 A.M. on May 30 were also asked in this section a question on whether they were fine with their names and/or individual responses being featured on the research presentation in the Chiang Research Festival on May 1; I removed this question at 12:00 A.M. on May 3 to allow some time for analysis of the survey answers before the research presentation.

2.2 Questionnaire for a Graduate Student

I also sent a questionnaire to a first-year graduate statistics student at the University of California, Berkeley, who also was an undergraduate statistics student at the university. I used Google Forms to create the questionnaire; I then sent the questionnaire link to the graduate student’s university email address at 2:02 P.M. on May 2.

The questionnaire was similar to the survey given to the undergraduate students. To ensure that only the response of the graduate student would be used in research, I made the survey only accessible by people with Google accounts associated with the University of California, Berkeley. I also had the survey automatically collect the university Google email addresses of all respondents. The questionnaire also allowed only one answer from each respondent to prevent anyone from providing multiple redundant responses for the questionnaire; the questionnaire also did not allow any respondent to edit his/her responses after submission.

Like the survey, the questionnaire consisted of five sections: Basic Information, Undergraduate Academics, Graduate Academics, Writing Experience Outside the Courses, and Handling Your Information. Basic Information section contained questions on the respondent's pre-Berkeley background to potentially assess how much his pre-Berkeley education may have helped him with statistical writing. Undergraduate Academics section contained questions on the academic disciplines he was studying (along with statistics) and the writing he did in the undergraduate statistics major courses he took. Graduate Academics section contained questions on the respondent's purpose for his graduate education along with questions on the writing he did in the graduate statistics courses he took. The section also contained questions on the writing the respondent may have done for his graduate thesis and/or comprehensive exams, since getting a graduate statistics degree in the University of California, Berkeley, requires taking a comprehensive exam or writing academic theses ("The MA Program in Statistics" 2018; "The Statistics PhD Program" 2018). Writing Experience Outside the Courses section contained questions on his experiences or outlook on statistical writing outside the university statistics courses. Handling Your Information section contained questions on whether he was fine with his name being featured in the research paper submission. I stopped accepting responses for the questionnaire at 5:00 P.M. on May 3.

2.3 Interviews of Statistics Faculty Members

I requested four statistics professors at the University of California, Berkeley, to hold a ten to fifteen-minute interview with me for this research paper. Out of these four professors, Professor Ani Adhikari and Professor David Aldous consented

to be interviewed for this research paper. Both of these professors taught undergraduate statistics major courses in the past. Professor Adhikari taught Stat C8, Stat 134, and Stat 140; Professor Aldous taught Stat 134, Stat 150, and the previously mentioned Stat 157 class (Adhikari 2018; Aldous 2018). Since these two professors had experience teaching prospective or declared undergraduate statistics majors in the past, I thought interviewing these professors would provide helpful input for my research paper.

I told each of these two professors, at the beginning of the interview, that there would be no audiovisual recording of the interview but would be a digitally written note of the interview; I also told each of these two professors the purpose of my research paper and of the class the paper is meant for (D. Aldous, personal communication, May 3, 2018; A. Adhikari, personal communication, May 3, 2018). I also told these professors that the project submission can potentially be uploaded online to the course webpage in the near future (D. Aldous, personal communication, May 3, 2018; A. Adhikari, personal communication, May 3, 2018). Then, I asked each professor for consent on having his/her name and interview responses being mentioned in the project submission and thanked each professor for setting aside his/her busy schedule for the interview; the interviews began right after the professors gave their explicit consent (D. Aldous, personal communication, May 3, 2018; A. Adhikari, personal communication, May 3, 2018)

Throughout the interviews, I asked the two professors about their own experiences in doing statistical writing, their course materials, and their thoughts on the state of undergraduate writing education for statistics majors at the University of California, Berkeley (D. Aldous, personal communication, May 3, 2018; A. Adhikari, personal communication, May 3, 2018). At the end of each interview, I thanked the professor again and told the professor that a typed summary of the interview would be sent to him/her by Monday, May 7 (D. Aldous, personal communication, May 3, 2018; A. Adhikari, personal communication, May 3, 2018).

3 DISCUSSION

When I stopped accepting responses for the undergraduate survey, nine people from the Statistics Undergraduate Student Association responded to my survey. The graduate respondent also responded

to the questionnaire by the time I stopped accepting responses for the questionnaire. Looking at the surveyees' responses, the surveyees generally reported Stat 15x series classes to be rather writing-intensive, especially compared to other statistics major courses. Out of five surveyees who did not take any Stat 15x classes, four stated that their statistics classes were not very writing-intensive when asked which statistics courses they took were the most writing-intensive and which ones were least writing-intensive; all of these five surveyees also reported only doing short answer responses and mathematical proofs. On the other hand, all four surveyees who took Stat 15x courses, along with the graduate student who took three Stat 15x series courses, reported doing research reports or comprehensive project/analysis write-ups. Isaac Pariser – a surveyee who took Stat 153 and Stat 155 along with Stat 13x series courses – explicitly noted that Stat 15x series courses seemed to be more writing-intensive than Stat 13x series courses. Most interestingly, out of four students who stated to have done statistical writing outside of university courses, the three students who thought their statistics courses helped with their writing all took Stat 15x series courses. So, it seemed that Stat 15x series courses gave undergraduate statistics major a lot of practice in real-world statistical writing, especially compared to lower-division statistics courses and Stat 13x series courses.

However, most undergraduate surveyees did not seem to be targeting their statistics course writings for non-statisticians; seven of the nine surveyees stated that they were their statistics course writings only towards their classmates and course staff. Even among four surveyees who took Stat 15x series courses, only two of them targeted their statistics course writings only towards their classmates and course staff. The graduate surveyee also only targeted his course writings towards their classmates and course staff for his undergraduate statistics courses. Hence, while undergraduate statistics majors may have been getting practice in doing statistical writing, they seemed to not be getting much education in doing statistical writing for non-statisticians.

Undergraduate surveyees generally thought their statistical major courses did not help or will not help with their statistical writing outside of university courses. Five surveyees thought their undergraduate statistics courses helped or will help; two surveyees were unsure; two surveys thought their undergraduate statistics courses did not or will not

help. The graduate respondent was also unsure if his undergraduate statistics courses helped with his statistical writing in his graduate courses and outside of university courses. Interestingly, most undergraduate surveyees who did statistical writing outside of university courses and the graduate respondent scored themselves four or five out of five on how skilled they were in statistical writing outside of undergraduate university courses. All of these undergraduate surveyees and the graduate respondent even reported targeting their non-course statistical writing towards their non-statistician colleagues. Even so, the graduate respondent and these surveyees were split on whether their undergraduate statistics courses helped them in doing statistical writing outside of university courses or not (with two saying yes, two saying maybe, and one saying no), so it could be that some of them were getting a lot of outside assistance in learning to do statistical writing towards non-statisticians.

Looking at faculty interviews, the professors seemed more focused on making their students practice statistical writing than making the students deeply think about whom they are communicating their statistical data to. Professor Aldous discussed how he wanted his students to get some practice in doing statistical writing, while Professor Adhikari put a lot of focus on making her students get over the inhibition of doing statistical writing (D. Aldous, personal communication, May 3, 2018; A. Adhikari, personal communication, May 3, 2018). However, both professors did not hold particularly high standards for grading their students' writing with Professor Aldous saying that he only gave "common-sense" feedback and Professor Adhikari saying that she could not hold her students to higher standards due to many of her students not speaking English as their primary language (D. Aldous, personal communication, May 3, 2018; A. Adhikari, personal communication, May 3, 2018). Such focus on writing practice, rather than the quality of writing, may explain the surveyees reporting a high amount of writing in some statistics major courses, but mostly not targeting their course writing for non-statisticians.

Professor Aldous noted that, while the department should do more in providing statistical writing education, statistics majors would eventually master statistical writing in graduate years (D. Aldous, personal communication, May 3, 2018). However, considering that only one out of nine undergraduate surveyees showed consideration for going to graduate

school for a Master's or Philosophy in Doctor degree, most undergraduate surveyees and perhaps most undergraduate statistics majors would not get proper statistical writing training in graduate school. Not only that, while the graduate respondent's answers point to graduate statistics courses being incredibly writing-intensive, the graduate respondent also mentioned that his course writings were targeted towards fellow classmates and course staff, not non-statisticians. So, contrary to Professor Aldous's belief, going to graduate school might not help current undergraduate statistics majors learn how to do statistical writing for non-statisticians.

When asked whether the amount of preparation that the undergraduate statistics majors in the University of California, Berkeley, were getting in statistical writing was justified, Professor Adhikari listed out many reasons for impracticality of enforcing a stricter writing education for statistics majors, citing lack of resources and potential dissatisfaction of statistics majors at the idea of a writing requirement (A. Adhikari, personal communication, May 3, 2018). Her response made me realize that the Department of Statistics perhaps did not want to institute stricter writing requirements partially due to lack of resources and partially due to the department's effort to attract more students to its programs. As an intended statistics major, I often heard in statistics courses and among my friends in Statistics Undergraduate Student Association that the Department of Statistics was underfunded. Looking at graduate costs, a Master's degree in statistics also came with a much higher tuition cost than most other graduate degrees, rivaling the cost of a Doctor of Optometry degree ("Tuition Costs & Fees" 2017). Such high cost of a Master's degree in statistics may signify the statistics department's need for more money, since higher tuition could be used to better fund the department. Professor Adhikari's attempt to accommodate students whose primary languages were not English by lowering writing standards also may signify the department's need for more money, since being accommodating to such students can bring in more nonresident international students to the department's programs, which – since nonresidents pay more for their degrees in the university than residents – could boost the department's funding.

Professor Aldous did once casually mention in the interview that the Department of Statistics was competing with the engineering departments for prospective data science students (D. Aldous,

personal communication, May 3, 2018). Such competition may explain why Professor Adhikari was hesitant about the idea of instituting stricter writing standards for statistics majors, since such standards may make prospective data science-oriented students move from statistics programs to competing programs like engineering, computer science, data science, and applied mathematics programs. Considering that most undergraduate surveyees – despite being intended or declared statistics majors – were uncertain about working in statistics in the future and were majoring in other subjects along with statistics, the Department of Statistics may not be wrong that a stricter writing requirement would push a lot of prospective students away from the department programs.

The graduate respondent and both professors did mention Professor Nolan's writing classes as classes that undergraduate statistics majors can take to learn a lot about statistical writing (D. Aldous, personal communication, May 3, 2018; A. Adhikari, personal communication, May 3, 2018). However, searching up Professor Nolan's classes from Fall 2016 to Fall 2018 (including summer sessions), only two statistical writing classes showed up. One of her classes was a Stat 15x series course in Fall 2017, which a statistics major must take three of to obtain his or her Bachelor's degree in statistics; however, a statistics major could take other 15x series courses to fulfill such requirement ("2017 FALL STAT 157 002 SEM 002" 2018, "Statistics Major" 2018). Her other class was a Stat 198 course in Spring 2018, a course not required for statistics majors to take ("2018 SPRING STAT 198 009 GRP 009" 2018). Also, her 15x series writing course also had only 17 spots and required explicit enrollment permission from the professor herself; meanwhile, her Stat 198 course had only 15 spots with 5 spots reserved for those with enrollment permission ("2017 FALL STAT 157 002 SEM 002" 2018; "2018 SPRING STAT 198 009 GRP 009" 2018). Such low number of available spots and the difficulty of getting these spots, combined with the low frequency of such classes, indicated that the Department of Statistics may not have been very open to teaching its major students about statistical writing. Also, only 19 students were enrolled or waitlisted in her Stat 15x class, while only 8 students enrolled in her Stat 198 class; such low number of students in her classes may signify a lack of interest in statistical writing among undergraduate statistics majors ("2017 FALL STAT 157 002 SEM 002" 2018; "2018 SPRING STAT 198 009 GRP 009" 2018).

4 CONCLUSION

Overall, it seemed that – while lower-division statistics major classes and Stat 13x/Stat 140 classes were not too writing-intensive – Stat 15x series classes were fairly writing-intensive and gave undergraduate statistics majors a lot of practice in doing statistical writing. However, the classes seemed to grade students’ statistical writing rather leniently, and students who majored in, were majoring in, or planned to major in statistics in the University of California, Berkeley, were generally not certain that their statistics courses helped with their statistical writing outside of the courses. While Professor Nolan’s writing classes provided academic forums for statistical writing for non-statisticians, her writing classes were small and not exactly easy to get into. Also, with the Department of Statistics being desperate for more resources and more student enrollment, the department did not seem to be very supportive of instituting stricter writing requirements for statistics students.

I must note that there were a lot of limitations to my methodologies. While it would have been ideal to survey everyone or a random sample of the population of all declared or intended undergraduate statistics majors in the university, the practical difficulty of identifying the members of this population – especially intended undergraduate statistics majors, potentially taking lower-level statistics courses – meant that surveying the members of the Statistics Undergraduate Student Association remained the most practical method to gather undergraduate input for this research paper. Also, only 9 out of 115 current members of the Statistical Student Association responded to the survey, so having more members answer the survey may have led to a more accurate analysis of the state of writing education for undergraduate statistics majors. Interviewing more professors, especially Professor Nolan, would also have helped as well.

However, as Professor Aldous mentioned, the Department of Statistics could do better in providing statistical writing education to undergraduate statistics majors (D. Aldous, personal communication, May 3, 2018). The department could do so by instituting writing components in more statistics major courses to prevent statistics majors from picking classes that are not writing-intensive. However, if the department faces practical problems with instituting a stricter writing requirement, the department could at least expand the size of its writing classes and

advertise its statistical writing classes more. Also, while there exist Reading and Composition classes that are related to certain majors such as legal studies and sociology, there is no Reading and Composition class related to statistics (“Reading and Composition Requirement” 2018). The Department of Statistics could work with administrators to create statistics reading classes that lowerclassmen could take. Considering the importance of writing in a statistician’s day-to-day tasks, putting some effort in to improve undergraduate statistical writing education in the University of California, Berkeley, would go a long way in improving the quality of the university’s undergraduate statistics education as a whole.

References

- Adhikari, A. (2016), “Data 8, Fall 2016: Midterm Exam,” *Data 8: The Foundations at Data Science* [online]. Available at <http://data8.org/data8assets/exam/data8-fa16-midterm.pdf>.
- Adhikari, A. (2016), “Data 8, Fall 2016: Final Exam,” *Data 8: The Foundations at Data Science* [online]. Available at <http://data8.org/data8assets/exam/data8-fa16-final.pdf>.
- Adhikari, A. [cited May 7, 2018], “Ani Adhikari, Teaching Professor,” *the University of California, Berkeley Department of Statistics* [online]. Available at <https://www.stat.berkeley.edu/~ani/>.
- Aldous, D. (2017), “Stat 157 Home Page,” *University of California, Berkeley Department of Statistics* [online]. Available at <https://www.stat.berkeley.edu/~aldous/157/index.html>.
- Aldous, D. [cited May 7, 2018], “David Aldous: courses taught since 2001,” *the University of California, Berkeley Department of Statistics* [online]. Available at <https://www.stat.berkeley.edu/~aldous/old-teach.html>.
- Williams, K. (1972), “Statistics and Communication,” *Journal of the Royal Statistical Society. Series D (The Statistician)* [online], 21, 127–128. Available at <http://www.jstor.org/stable/2987323>.
- (2017), “Homework 7: Correlation, Regression, and Least Squares,” *DataHub* [online]. Available at <http://datahub.berkeley.edu/user-redirect/interact?repo=data8assets&path=materials/sp17/hw/hw07>

- (2017), “Tuition Costs & Fees,” *Berkeley Graduate Division* [online]. Available at <http://grad.berkeley.edu/admissions/costs-fees/>.
- (2018), “Project 2: Inference and Capital Punishment,” *DataHub* [online]. Available at <http://datahub.berkeley.edu/hub/user-redirect/git-sync?repo=https://github.com/data-8/materials-sp18&subPath=materials/sp18/project/project2/project2.ipynb>.
- (2018), “SUSA Enrollment Advising for Fall 2018,” *Google Drive* [online]. Available at https://docs.google.com/presentation/d/1IjXXgcNw6B1UFirzjXN_OAkOm8b-2i0AkrO4l7og3I/edit?usp=sharing.
- [cited April 10, 2018], “Reading and Composition Requirement,” *Berkeley Academic Guide* [online]. Available at <http://guide.berkeley.edu/undergraduate/colleges-schools/letters-science/reading-composition-requirement/>.
- [cited April 23, 2018], “A Day in the Life of a Statistician,” *The Princeton Review* [online]. Available at <https://www.princetonreview.com/careers/149/statistician>.
- [cited April 23, 2018], “Statistics Major,” *the University of California, Berkeley Department of Statistics* [online]. Available at <http://statistics.berkeley.edu/programs/undergrad/major>.
- [cited May 7, 2018], “2017 FALL STAT 157 002 SEM 002,” *Berkeley Academic Guide* [online]. Available at <http://classes.berkeley.edu/content/2017-fall-stat-157-002-sem-002>.
- [cited May 7, 2018], “2018 SPRING STAT 198 009 GRP 009,” *Berkeley Academic Guide* [online]. Available at <http://classes.berkeley.edu/content/2018-spring-stat-198-009-grp-009>.
- [cited May 7, 2018], “The MA Program in Statistics,” *the University of California, Berkeley Department of Statistics* [online]. Available at <https://statistics.berkeley.edu/programs/graduate/masters>.
- [cited May 7, 2018], “The Statistics PhD Program,” *the University of California, Berkeley Department of Statistics* [online]. Available at <https://statistics.berkeley.edu/programs/graduate/phd>.