

READING TO WRITE, WRITING TO READ, AND THE DEVELOPMENT OF SKILLS IN DATA TRANSLATION

Alison L. Gibbs

Department of Statistical Sciences, University of Toronto, Ontario, Canada

alison.gibbs@utoronto.ca

There have been many discussions about the nature of data science and the training of data scientists. Common to these discussions is the recognized demand for data scientists who can communicate effectively to stakeholders. More recently, data translator has been identified as a needed role on data science teams for which it is difficult to find appropriately skilled people. Students in our undergraduate programs in statistics and data science typically enroll in these programs because of their interest and aptitude in mathematical and computational problem solving. How do we improve their communication skills to help fill the need for data translators? We describe ways we have integrated reading and writing in an undergraduate capstone course with the goal of developing the required skills.

INTRODUCTION

Recently, “data translator” has emerged as a popular expression to describe someone who might fill a pressing need in data science work (Maynard-Atem & Ludford, 2020). A data translator bridges the technical aspects of data science projects with the business side, translating business problems into technical and data specifications and connecting the output of technical work to the business problem by crafting a comprehensible story. Having a data translator role fill this gap between technical and business sides of projects allows data scientists to focus on the technical work and can make up for deficiencies in the communication skills of data scientists. The skills required for this translation role have been described as those of the proverbial unicorn, having technical and business understanding, and the ability to communicate in both worlds. Does this newly branded role reflect a new imperative to spend more attention on developing communication skills in statistics and data science students?

Recent recommendations for data science training curricula have included communication as a foundational skill but give far less detail about what these communication skills entail when compared to details given for technical skills. Moreover, suggestions on how to develop communication skills can differ in their approaches. For example, De Veaux et al. (2017) advocate for integration of communication throughout all courses and the National Academies of Sciences, Engineering, and Medicine (2018) espouses the benefits of a writing intensive course.

In this paper, we consider intentional integration of reading and writing as an approach to producing graduates who are better prepared for the challenges of data translation. Our focus is the communication of results with stakeholders and not areas of communication that also deserve attention, such as asking questions to formulate the problem (or creating “shared understanding” as in Vance et al., 2022), nor technical communication such as what is required for collaboration with technical team members and communication with one’s future self through the creation of well-documented, reproducible analyses. Our approach is closer to initiatives designed to help students find their statistical voices (Khachatryan & Karst, 2017; MacGillivray, 2005).

We now describe our approach to the integration of reading and writing and give examples from a capstone course for statistics students.

READING-TO-WRITE AND WRITING-TO-READ

In his memoir *On Writing* (2000), author Stephen King gives the following advice for aspiring writers: “If you want to be a writer, you must do two things above all others: read a lot and write a lot” (p. 145). This sentiment is common to many writers. For several other examples, see the International Literacy Association (2020).

Many have written about the value of using writing in the teaching of statistics, including its value for assessing statistical thinking, reducing anxiety, and developing deeper conceptual understanding (see, for example, Holmes, 2011). Fewer have written about the benefits of reading. However, we can take inspiration from the research about teaching children, where the recognition of

the importance of reading to develop writing skills and writing to develop reading skills has received significant study. (See International Literacy Association, 2020, and the references therein.) For elementary school students, it has been shown that having students write about what they have read improves their understanding and results in writing that is longer, more descriptive, and with fewer errors. Reading requires the ability to interpret meaning from written language, and guided reading can make explicit how writers express ideas and the choices they make. We believe that this also applies to undergraduate students developing skills in statistics communication, and that reading intentionally to uncover aspects of writing that do and do not work well, in tandem with writing exercises, can improve student communication skills.

READING-TO-WRITE AND WRITING-TO-READ IN A CAPSTONE COURSE

We illustrate our approach in the context of a final-year undergraduate capstone course, although elements of this approach begin in the first-year course in our statistics program of study at the University of Toronto. Equal time in the capstone course is devoted to seminars and to project work, in which students serve as the quantitative experts on a collaborative research project. The main output of the project is a written report for collaborators. Seminar time is used to prepare students for professional practice broadly, including writing the project report as well as other forms of communication. Topics that are discussed include contrasting how statistical results are presented in various venues for various intended audiences. This is typically carried out through reading and analysis of writing on important issues in statistical practice, including ethics; principles of the presentation of results of data in texts, tables, and visualizations; decisions made during an analysis and researcher degrees of freedom; and reproducibility, replicability, and false discovery. (For more on the course, see the webinar “Consultation, Communication, and Collaboration: A project course to engage statistics students in the three C’s” at <https://www.causeweb.org/cause/webinar/activity/2012-09>.)

Regardless of the topic, every seminar has the goal, sometimes explicit and sometimes stealth, of improving students’ skills in writing and reading. Preparatory reading is assigned, and questions are provided to guide students’ reading and to structure the subsequent class discussion. These questions address all aspects of the assigned reading and how it was written, including identifying the take-home message, possible implications for students’ projects, and how the authors presented their case with critical discussion about whether they were effective and why. Below we give more examples of how we have our students engage in reading and writing in an integrated manner.

Deconstructing Scientific Papers

In our discussions about writing the project report, we often discuss how to begin. Students aren’t surprised to learn that papers are not typically written in order, and it is easy to engage them in a conversation about what section of a paper they would write first and why. However, students are often surprised when we admit to them that we also do not typically read a scientific paper from start to finish. Nolan and Stoudt (2020) suggest a structure for multiple passes through a paper, starting with a scan of its structure, examining the beginning and ending elements that make the argument, and considering the evidence that is typically presented in figures and tables. We also discuss other potential orders, depending on our purpose for reading the paper, such as purposes for which we might want to focus first on the evidence. In these situations, our initial focus may be the figures and tables and their captions that provide a summary of the data from which we might begin to form our own ideas about the strength of the evidence, both in the size of observed differences and in the apparent appropriateness of the analysis. In other cases, our primary interest may be whether the work is generalizable and compelling given the greater context, as argued in the introduction and conclusion. In any of these cases, one of the goals of our class discussion is to make the invisible activity of reading be more visible.

In addition to considering scientific articles as a template for organizing a scientific argument, class discussion is often centred around parts of articles that students indicate were difficult to follow, so that students learn to value the construction of a clear and compelling argument, and to identify the nature of weaknesses in communication when they exist. To do this, we are strategic in providing readings that are both exemplars of cogent writing, and examples that we feel could be improved. We

do not ask every question below with every paper we assign, but we always ask students to consider some subset of the following:

- Can you describe the main message and why you should care in one or two sentences? If you are unsure, why?
- When you read a passage that you do not understand, where did your understanding break down? Could this be improved by writing this section differently?
- When reading the description of analyses carried out, do you feel confident that you could write code to replicate the analyses if you were given the data? If not, what information is missing?
- Do the visualizations tell a story independently of the rest of the paper? Do they reflect the key takeaways of the article that are presented in the title and abstract? What value do they add? What new questions do they raise?

The Importance of Precision in Writing

As mentioned above, when the class is assigned a reading that is a report on a scientific study that includes analysis of data, students are regularly asked to write down the model(s), sometimes as a mathematical expression, sometimes as code. Students feel confident with this exercise because this type of work connects their reading of the scientific paper to the expertise that they have developed in the methodology courses they have taken to date. Whether or not writing the mathematical expression or code can be accomplished easily and with confidence that it is correct leads to a conversation about precision in writing and opens the door to discussion of common word choices that result in ambiguity.

When reading articles in the popular press, we often encounter statistical terms used in imprecise ways. Words such as independent, significant, random, and correlated, have a precise technical meaning in probability and statistics but are used in different ways in common discourse. The effect of this lexical ambiguity on student acquisition of quantitative literacy and reasoning has been studied. For example, Kaplan and Rogness (2018) have shown that designing activities that address lexical ambiguity may lead to improved understanding of underlying concepts. Although never assigned as an exercise for course credit, students in the course have initiated the collective creation of a glossary of terms that they now recognize they need to be careful with using in their writing.

Students often encounter lexical ambiguity when they critically examine the writing of their peers, in addition to examining readings from the popular press. Careful reading of examples from peers also often highlights the need to be more expansive in their own writing, to provide an explanation of a result that is both precise and complete. For example, a typical student interpretation of some analysis might be “there is a strong effect of treatment.” Early in the course, students need prompting to see how this might be better expressed, often missing the fact that this description does not include which treatment gives better results.

Integrating Reading, Writing, and Oral Presentation

To become effective data translators, students not only need to develop skills in reading and writing but also need to develop skills in listening and speaking. In the course, listening and one-on-one oral communication skills are primarily developed in conversations with project collaborators and reflections on these conversations. Speaking to an audience in a formal presentation is a form of communication that students find particularly stressful. We now describe a popular assignment in the course, which is designed to integrate reading, writing, and oral presentation with the goal of improving students’ confidence in their presentation by requiring them to interpret and describe the argument in another person’s writing in a number of different ways. In this assignment, each student is assigned their own article, chosen from the classic series of articles in the *BMJ Statistics Notes* series (<https://www.bmj.com/specialties/statistics-notes>), the newer *BMJ* series *Endgames* (<https://www.bmj.com/education/endgames>), or the *Nature Methods: Points of Significance* series (<https://www.nature.com/collections/qghhqm/pointsofsignificance>), with the occasional short article from other sources. For their article, students do each of the following:

- Write a 250-word summary that is focussed on the key take-home message of the article.
- Write a 140-character summary of the key take-home message in plain language for a general audience.
- Give a rating (on a scale of 1 to 5), with justification, of the article's impact, with the criterion being how important the message of the article is for all students in the class to know or to learn.
- Deliver a one-slide, five-minute class presentation, with the goal that all students learn the take-home message without needing to read the article.

Each of the articles is very short (one or two pages). Many include an illustrative story, and students must decide what, if any, details of the story are important to communicate the author's key message. This assignment is given near the end of the year and connects student skills in reading with the need to re-communicate the message in new formats, both written and oral.

CONCLUSION

Sitting down with a blank page to begin writing is always a daunting prospect. For students who have chosen to study statistics because they are confident in their mathematical and computational skills, starting the task of presenting the results of their analyses in writing for a non-technical audience can feel overwhelming. We have found that integrating reading and writing, with ongoing conversations about how writing is structured, why it is structured that way, and where and why confusion arises and can be avoided, can help students develop confidence in translating their technical work for a broader audience.

REFERENCES

- De Veaux, R. D., Agarwal, M., Averett, M., Baumer, B. S., Bray, A., Bressoud, T. C., Bryant, L., Cheng, L. Z., Francis, A., Gould, R., Kim, A. Y., Kretchmar, M., Lu, Q., Moskol, A., Nolan, D., Pelayo, R., Raleigh, S., Sethi, R. J., Sondjaja, M., ... Ye, P. (2017). Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and Its Application*, 4, 15–30. <https://doi.org/10.1146/annurev-statistics-060116-053930>
- Holmes, K. Y. (2011). Tips for incorporating writing into an introductory statistics course. *Association for Psychological Science Observer*, 25(1). <https://www.psychologicalscience.org/observer/tips-for-incorporating-writing-into-an-introductory-statistics-course>
- International Literacy Association. (2020). *Teaching writing to improve reading skills*. <https://www.literacyworldwide.org/docs/default-source/where-we-stand/ila-teaching-writing-to-improve-reading-skills.pdf>
- Kaplan, J. J. & Rogness, N. (2018). Increasing statistical literacy by exploiting lexical ambiguity of technical terms. *Numeracy*, 11(1), Article 3. <https://doi.org/10.5038/1936-4660.11.1.3>
- Khachatryan, D., & Karst, N. J. (2017). V for voice: Strategies for bolstering communication skills in statistics. *Journal of Statistics Education*, 25(2), 68–78. <https://doi.org/10.1080/10691898.2017.1305261>
- King, S. (2000). *On writing: A memoir of the craft*. Scribner.
- MacGillivray, H. (2005). Helping students find their statistical voices. In *Statistics education and the communication of statistics. Proceedings of the IASE/ISI Satellite Conference*. ISI/IASE. <https://iase-web.org/documents/papers/sat2005/macgillivray.pdf?1402524993>
- Maynard-Atem, L., & Ludford, B. (2020). The rise of the data translator. *Impact*, 2020(1), 12–14. <https://doi.org/10.1080/2058802X.2020.1735794>
- National Academies of Sciences, Engineering, and Medicine. (2018). *Envisioning the data science discipline: The undergraduate perspective. Interim report*. The National Academies Press. <https://doi.org/10.17226/24886>
- Nolan, D., & Stoudt, S. (2020). Reading to write. *Significance*, 17(6), 34–37. <https://doi.org/10.1111/1740-9713.01469>
- Vance, E. A., Alzen, J. L., & Smith, H. S. (2022). Creating shared understanding in statistics and data science collaborations. *Journal of Statistics and Data Science Education*, 30(1), 54–65. <https://doi.org/10.1080/26939169.2022.2035286>