

Board 17: Teaching STEM undergraduates discipline-specific writing skills: a data-driven learning approach

Dr. Ryan K Boettger, University of North Texas

Ryan K. Boettger is an associate professor and assistant chair in the Department of Technical Communication at the University of North Texas. His research areas include data-driven learning, content analysis, and technical editing. His research in STEM education is currently funded by the National Science Foundation. He can be contacted at ryan.boettger@unt.edu.

Dr. Stefanie Wulff

Teaching STEM undergraduates discipline-specific writing skills: A data-driven learning approach

Our research is funded by the NSF Division of Undergraduate Education and seeks to improve the quality of technical writing instruction for undergraduate STEM students. Specifically, we propose a data-driven learning approach for teaching STEM students writing patterns specific to their respective disciplines.

In this paper, we summarize the basic tenants of data-driven learning (or DDL) and then describe the development of six instructional DDL units, the deployment of these units in a senior-level fish ecology course, participants' perceptions of the DDL, and our current work in this area involving 140 participants in introductory and advanced technical writing courses.

Data-driven learning

As the educational marketplace expands, institutions of higher learning are experimenting with how active learning increases student success. Freeman et al.'s meta-analysis of STEM education studies found that active learning significantly increased course grades over didactic methods and was particularly effective in classes of 50 or less students. In contrast, students were 1.5 times more likely to fail a course that lacked active learning strategies [1].

The spectrum of active learning ranges from simple activities, such as writing minute papers or pausing for reflection, to more complex activities, such as hands-on technology and inquiry learning. Active learning is being promoted as a means of improving academic achievement, increasing students' retention, improving student-instructor interactions, and supporting peer relations [2].

DDL encompasses multiple active learning strategies. DDL contrasts with traditional deductive, lecture-based methods and promotes students' active engagement through computer-assisted tools and a databank of authentic writing. Instruction is typically designed to (i.) foster students' active engagement, (ii.) encourage inductive activities that allow students to explore a topic on their own terms, (iii.) promote interaction between students, and (iv.) provide students with output-focused activities to apply this new knowledge [3].

Our research specifically explores the role DDL can play in the teaching of technical writing. The undergraduate technical writing service course has become a mainstay across most institutions of higher education; however, its interdisciplinary appeal has arguably jeopardized the quality of the instruction, particularly for STEM majors [4]. Specifically, the heterogeneous student population has resulted in generic, prescriptive instruction that often contradicts how STEM practitioners communicate on the job [5]. A DDL approach to teaching technical writing has the ability to customize instruction for students as it facilitates the exploration of typical language patterns in a large sample of authentic writing that represents a variety of disciplines.

DDL unit development

The researchers developed resources for six DDL units. The first unit showed student participants how to perform basic functions in AntConc. AntConc is a free text processing tool that assists in the analysis of large data sample; it was specifically designed to facilitate the DDL [6]. The five remaining units focused on a specific grammar topic: passive voice, reporting verbs, evaluative adjectives, hedges and boosters, and transitions. We identified these grammar topics based on what instructors previously identified as vital to successful STEM writing instruction. The units emphasize data and results as well as enhance students' ability to accurately report, interpret, draw conclusions from, and explain possible errors in the data sets they access using AntConc, as well as in their own writing [7-11].

Each unit contained a lesson plan, in-class activities, an infographic fact sheet, and homework assignments (with answer keys for instructors). The in-class activities engaged students with one of three data sets:

- A student writing data set, which included 99 files of student technical and scientific writing, including abstracts, critical reviews, process explanations, progress reports, proposals, and white papers. All the texts earned a grade of "A" or "B" from the instructors of record.
- A professional writing data set, which included 240 files of published writing in cell biology, electrical engineering, mechanical engineering, applied linguistics, marketing, philosophy, and physics. All of the papers were peer-reviewed and published in the top journals in their respective fields.
- A professional speech data set, which included 152 files of transcribed speech, including lectures, classroom discussions, lab sessions, seminars, and advising sessions.

Students engaged with any of these data sets depending on how language use was emphasized in a particular unit. For example, legitimate uses of passive voice was explored in the professional writing data set because it is a structure that instructors often recommend their students avoid. However, the DDL unit allows students to explore how published authors in their discipline use passives, thereby enforcing its actual use and rhetorical function. On the other hand, student pairs explored common boosters (e.g., *really*, *totally*, *literally*) in the professional writing and professional speech data sets in order to distinguish between how we communicate in writing versus how we communicate in speech.

The homework assignments were designed to (i.) measure students' initial understanding of the grammar topic taught and (ii.) measure their ability to apply their new knowledge. Assessing the homework provides an important baseline on the overall effects of DDL.

Deployment of the DDL units

We piloted our six instructional DDL units during the fall 2018 semester with 17 undergraduates enrolled in a senior-level fish ecology course. The demographics of this particular sample were rather homogenous. In addition to the majority of participants being seniors majoring in ecology, 76.47% self-identified as female, 88.23% as white or Caucasian, and 94.12% as native English speakers.

The major writing assignment in this course was a critical review of a fish family. Over the course of the 15-week semester, students produced five drafts of their review, with each draft requiring additional content on the fish family.

Each DDL treatment lasted an average of 25 minutes and was part of the course's required lab component. For this first study, one of the researchers taught all the units rather than the instructor of record, who is a trained ecologist. (Note: our current study explores the effects of DDL when taught by instructors who are not formally trained in corpus linguistics).

Every DDL lesson followed the same sequence: students were introduced to the grammar topic and then engaged in a data-driven activity. Students then received a short lecture on how to effectively use the grammar topic before engaging in another data-driven activity with one of the three data sets. Students were then assigned homework related to the unit and given access to reference materials, including the infographic fact sheet. Informed consent granted the researchers access to the course learning management system, where all the participants' homework and writing were downloaded for analysis. In the final week of the semester, students completed a survey that collected their perceptions of the DDL units.

Preliminary results

The data obtained from the students are currently being analyzed along various dimensions. One analysis will focus on students' uptake and change of use of the illustrative language examples presented in the instructional units and accompanying materials. This will be done by creating concordances and frequency counts of all attestations of the language examples in the student data (using the statistical and programming environment *R*). Uptake and use will be tracked as a function of time in the semester, instructor, the students' major, and other available meta-data. In a second analysis, the data will be annotated for parts of speech and syntactic information (using the corresponding spaCy natural language processing tools). This will facilitate a comprehensive analysis of students' uptake and change of use of the grammar topics covered in the instructional units above and beyond the specific language examples provided in the instructional units.

In addition, students' responses to the satisfaction survey will be analyzed, and correlated with the uptake and use results obtained in the above-mentioned analyses.

Survey results

Overall, students had little initial difficulties locating the data-driven materials in their learning management system or with the technical aspects of using the AntConc text processing tool ($M = 1.67/4$). These difficulties decreased further by the end of the treatment ($M = 1.17$). These results are encouraging as previous research has identified technological difficulties as a barrier to DDL [12].

On average, participants enjoyed the data-driven learning ($M = 3.42/4$) and understood the purpose of engaging in learning about grammar as part of their fish ecology coursework ($M = 3.42$). Further, most participants believed their general writing improved as a result of the DDL ($M = 3.17$) and that they would use the related skills in their future writing ($M = 3.33$).

Participants also ranked the DDL content lessons from most enjoyable to least enjoyable. Hedges and boosters, which was the third unit taught, was the most enjoyable, followed by transitions, which was the sixth unit taught (see Table 1). Ongoing data analysis will assess if participants' enjoyment level correlates with the frequency of the language patterns in their technical writing.

Table 1. The DDL content units ranked by enjoyment level, order taught in the treatment.

Unit	Mean	Order Taught
Intro to DDL/AntConc	N/A	1
Hedges and boosters	2.33	3
Transitions	2.83	6
Reporting verbs	3.08	4
Evaluative adjectives	3.25	2
Passive voice	3.50	5

However, participants' desire to use the DDL materials was disconnected from how often they actually used the materials. On average, participants consulted the DDL materials outside of the course a little over 30% of the time ($M = 3.83/7$). Similarly, participants consulted the DDL materials for writing assignments outside of the course about 30% of the time ($M = 3.33$). However, every participant wished they had consulted the DDL materials outside of the course ($M = 4.58$) as well as for writing assignments outside the course ($M = 4.42$). These types of results align with previous findings [13] and therefore suggest that future research could investigate what would incentivize increased use of the DDL materials.

Future directions

These preliminary results motivated a larger study ongoing throughout the spring 2019 semester that includes 144 students who are enrolled across six sections of introductory or advanced technical writing courses. The instructors who teach these respective sections (and who are all untrained corpus linguists) were trained to teach the six DDL units. Over 60% of these students are enrolled in a STEM major, including biology, computer science, and mechanical and energy engineering. DDL-related homework and writing assignments are being collected and prepared for analysis, which should reveal more generalizable results on the effects of DDL in learning technical writing, which we will report at the ASEE conference.

References

- [1] S. Freeman *et al.*, "Active learning increases student performance in science, engineering, and mathematics," *Proceedings of the National Academy of Sciences*, vol. 111, no. 23, pp. 8410-8415, 2014.
- [2] A. F. Lippens, "How active learning spaces support an evolving pedagogy," in *Expert Perspective* vol. 2019, EAB, Ed., ed, 2016.
- [3] K. Chujo, L. Anthony, and K. Oghigian, "DDL for the EFL classroom: Effective uses of a Japanese-English parallel corpus and the development of a learner-friendly, online parallel concordancer," in *American Association for Corpus Linguistics*, 2009.
- [4] L. Meloncon and P. England, "The current status of contingent faculty in technical and professional communication," *College English*, vol. 73, no. 4, pp. 396-408, 2011.

- [5] J. Wolfe, "How technical communication textbooks fail engineering students," *Technical Communication Quarterly*, vol. 18, no. 4, pp. 351-375, 2009.
- [6] L. Anthony, "AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom," in *Professional Communication Conference Proceedings*, 2005, pp. 729-737.
- [7] T. F. Jaeger, "Redundancy and reduction: Speakers manage syntactic information density," *Cognitive Psychology*, vol. 61, no. 1, pp. 23-62, 2010.
- [8] U. Römer, "English in academia: does nativeness matter?," *Anglistik: International Journal of English Studies*, vol. 20, no. 2, pp. 89-100, 2009.
- [9] U. Römer, "The inseparability of lexis and grammar Corpus linguistic perspectives," *Annual Review of Cognitive Linguistics*, vol. 7, no. 1, pp. 140-162, 2009.
- [10] R. K. Boettger, "Explicitly teaching five technical genres to English first-language adults in a multi-major technical writing course," *Journal of Writing Research*, vol. 6, no. 1, pp. 300-329, 2014.
- [11] E. Finegan and D. Biber, "Register variation and social dialect variation: The register axiom," in *Style and Sociolinguistic Variation*, P. Eckert and J. R. Rickford, Eds. Cambridge, UK: Cambridge University Press, 2001, pp. 235-267.
- [12] F. Farr, "Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users," *Language Awareness*, vol. 17, no. 1, pp. 25-43, 2008.
- [13] R. K. Boettger, "Using corpus-based instruction to explore writing variation across the disciplines: A case history in a graduate-level technical editing course," *Across the Disciplines*, vol. 13, no. 1, 2016.