V for Voice: Strategies for Bolstering Communication Skills in Statistics

Davit Khachatryan and Nathaniel Karst
Mathematics and Science, Babson College, Babson Park, Wellesley, MA

ABSTRACT
With the ease and automation of data collection and plummeting storage costs, organizations are faced with massive amounts of data that present two pressing challenges: technical analysis of the data themselves and communication of the analytics process and its products. Although a plethora of academic and practitioner literature have focused on the former call, the latter challenge has received less attention. Here, we present strategies that we have found effective for bolstering communication skills of both undergraduate students and masters candidates at a business school. These approaches are based on the case studies that provide ample opportunity for oral communication among students with strong backgrounds in mathematics and introductory statistics, and solid training in written communication among student populations with less homogeneous preparation. We provide detailed discussions motivated by concrete examples executed across three courses.

1. Introduction
In an important and much cited report on statistics pedagogy, Cobb (1992) presented a need for increased group discussion and written as well as oral presentations in classroom (among numerous other improvements to statistics education). According to the report, with the advances in statistical software it was now possible to deviate from the traditional formula-based instruction and instead dedicate time to practicing writing, a convenient pathway to better understanding (Cobb 1992). Support for fostering written communication skills in statistics education came also from Sharpe (1991), citing among numerous added values of writing the impact it has on internalization and conceptualization of material and its influence on the quality of communication of methods and conclusions (Sharpe 1991). Regarding the latter, the author states that the qualitative interpretation generated by the future statistician should be communicated intuitively to clients: “Statistical clients might include social scientists, admissions officers, surgeons, psychiatrists, and art historians, each with a varying degree of statistical understanding.” Garfield (1993) argued in favor of practicing communication as part of a collaborative learning framework; through communication students can observe misconceptions that they thought they had already mastered. Smith (1998) stressed the preparation of written reports and delivery of oral presentations as some of the notable benefits of active learning. Further, Garfield and Gal (1999) added that communication skills are imperative not only for conveying students’ own views but also for helping students intelligently argue about others’ use and interpretation of data. Rumsey (2002) considered communication and interpretation skills essential for what the author calls “statistical competence,” noting that communication skills are among the most important traits that employers are looking for among new graduates. The importance of bolstering communication skills in statistics education has been stressed by these authors in these and in a number of other academic publications, perhaps culminating with the recent ASA Guidelines (2014), which state that “Students need to be able to communicate complex statistical methods in basic terms to managers and other audiences and to visualize results in an accessible manner.” The importance of communication in the workforce has been stressed by employers themselves, as in a recent white paper by PwC (2015). We, the statistics educators need to rise to this challenge.

In spite of the evidence outlined above, not much attention has been dedicated in the literature to the presentation and discussion of actual strategies and examples that we, statistics educators can implement in classroom to that end. Moreover, there is a lack of discussion as to how the existing methods to improve communication should adapt in the current age of Big Data. The purpose of this article is to provide a well-rounded and comprehensive discourse on approaches that we have found effective for fostering communication skills when teaching our undergraduate and master’s students courses that in one way or another relate to Big Data. With our pedagogic approach we strive to provide an opportunity for students to enhance their verbal and written communication skills of technical material. Although the core objectives of the courses outlined in this article are of technical/quantitative nature, bolstering of communication skills can only help achieve those learning objectives.

The remainder of the article is organized as follows: the next section provides the contextual setting as well as the conceptual...
framework for the development of communication skills discussed throughout the rest of the article. Section 3 presents a discussion of approaches for bolstering written communication skills in a group of graduate students with heterogeneous prior preparation. The section that follows discusses strategies for improving verbal communication skills in undergraduate programs where the technical background and knowledge of introductory statistics of students is much more homogenous and, on the whole, stronger than that of the graduate audience. Section 5 concludes the article.

2. The Four V’s?

Thinking about communication without regard to the rapidly growing complexity of the professional environment in which the communicator will find herself upon graduation is a futile exercise. In the context of quantitative professions, the present period of our informational age is characterized by the oceanic amount of data, which gets generated and stored each second. To put into perspective, in the year 2000, an organization hosting a terabyte of data was at the vanguard of technological development (Franks 2012). Fast forward into the present, and the sizes of many corporate data warehouses are measured in petabytes, that is, thousands of terabytes. Based on 2012 statistics, there are 2.5 exabytes (or equivalently 2.5 million terabytes) of data created on a daily basis, and this volume doubles roughly every 40 months (McAfee and Brynjolfsson 2012). Further, as of 2012 the internet saw more traffic each second than the amount of information in the entire internet in 1992 (McAfee and Brynjolfsson 2012). The ongoing exponential expansion of data has not only been fueled by the ease of data collection, but also by the plummeting cost of data storage. Indeed, a gigabyte of data storage that used to cost roughly 19 dollars back in 2000 cost roughly 7 cents as of the third quarter of 2012 (Ames 2012). The sheer size and complexity of data available to organizations as a result of these technological innovations, as well as the pressing organizational needs to generate readily actionable insights from such data under aggressive and competing deadlines, makes the communication skills of newly graduated analytical professionals fundamentally important.

Laney (2001) discussed the complexity of data in the context of electronic (e-) commerce. In particular, the increased depth and breadth of information that can be collected on each e-commerce transaction together with the plummeting storage costs had led to massive amounts of stored transaction data. These copious data, coupled with both the increasing speed at which data arrived at the point of sale, as well as the diversity of data characterized by a “variety of incompatible data formats, non-aligned data structures, and inconsistent data semantics” (Laney 2001) called for both improved data processing techniques and for a realignment of managerial thinking in order to rise to these challenges. Laney (2001) went on to suggest a conceptual framework in which Big Data is characterized by the increasingly large size (“volume”) of data, the speed at which that information arrives (“velocity”), as well as the diversity in data types (“variety”), now colloquially referred to as the “3V’s of Big Data. Related to those challenges, Big Data is commonly thought of as cumbersome to handle with the use of modern data processing and analytic capabilities and technologies (Chen and Zhang 2014).

With the avalanche of data descending upon for-profit and nonprofit organizations alike, today the teaching of communication skills should, therefore, be viewed—to the extent possible in classroom – in the context of Big Data. In this article, we propose a pedagogical model that integrates the voice into the existing “3V” framework (Figure 1). By carefully controlling for the data velocity, volume, and variety in the activities and case reports to be described throughout the remainder of this article, we both imparted the core concepts of data mining and helped our students develop their ability to give voice to analytics process and its products. As seen from Figure 1, voice serves not only as the byproduct of analytics but also as its catalyst.

2.1. Interconnectedness of “Voice” and Thought

Contrary to the often accepted belief that effective communication skills in quantitative disciplines can be cultivated only after a solid and deep understanding of the technical material, we strongly believe that communication skills—the voice—can and should be developed in tandem with critical thinking skills in data mining and data management. Continually practicing effective communication, be it in the form of written delivers, discussion, or presentations, can serve as a powerful medium to improve the very understanding and mastery of quantitative material. Thus, voice does not just follow as one end product of analytics education but instead continually supports the learning process, as a result improving itself. Put differently, understanding is deepened by effective communication, and this additional depth, in turn, results in better communication. The various pedagogical strategies that we describe in this article can be thought of as examples to implement the conceptual model outlined in Figure 1.

Support for the interconnectedness between knowledge and communication has been known in educational psychology and philosophy for decades. According to Dewey (1910), a verbal sign affects individual meanings in that it selects a meaning
There are many other sources that support the use of writing in analytics and pedagogical contexts. For example, the work of Lev Vygotsky (1987; Smagorinsky 1998) further supports the use of writing in analytics and pedagogical contexts. Vygotsky stresses the unity of thought and speech, explaining how these two complement each other by continuously blending together. Indeed, he argues that “Speech does not merely serve as the expression of developed thought. Thought is restructured as it is transformed into speech. It is not expressed but completed in the word. Therefore, precisely because of their contrasting directions of movement, the development of the internal and external aspects of speech forms a true unity” (Vygotsky 1987). Along those lines, according to Smagorinsky (1998), “The processes of rendering thinking into speech is not simply a matter of memory retrieval, but a process through which thinking reaches a new level of articulation.” Finally, according to Ericsson and Simon (1998), “Although verbal descriptions and explanations may not reflect spontaneous thinking with complete accuracy, such verbalizations present a genuine educational opportunity to make students’ reasoning more coherent and reflective.”

Similar evidence in pedagogical literature exists in favor of using writing as a tool for learning. For example, Bracewell and Breuleux (1994) note that “the problem itself is not usually fully defined beforehand …”. Rather, the process of problem definition is in part carried out through the activity of text production as the writer organizes, reorganizes, and elaborates knowledge in the course of writing” (Bracewell and Breuleux, 1994; Smagorinsky 1998). This once again reinforces our conceptual model illustrated in Figure 1; the process of writing, similar to that of speech, not only is an end product resulting from a deep understanding of technical material but itself can serve as a tool for social purposes make it “a conscious tool of conveying knowledge and assisting thought.” It is important to notice that with this Dewey sees the improved oral and written speech as a tool not just to convey knowledge but to aid the thinking process itself, hoping that those skills will ultimately serve as “intellectual instruments.” Further support comes from Lev Vygotsky, who in his seminal and much discussed Thinking and Speech posits that “the development of the child’s thinking depends on his mastery of the social means of thinking, that is, on his mastery of speech” (Vygotsky 1987; Smagorinsky 1998). Further, Vygotsky stresses the unity of thought and speech, explaining how these two complement each other by continuously blending together. Indeed, he argues that “Speech does not merely serve as the expression of developed thought. Thought is restructured as it is transformed into speech. It is not expressed but completed in the word. Therefore, precisely because of their contrasting directions of movement, the development of the internal and external aspects of speech forms a true unity” (Vygotsky 1987). Along those lines, according to Smagorinsky (1998), “The processes of rendering thinking into speech is not simply a matter of memory retrieval, but a process through which thinking reaches a new level of articulation.” Finally, according to Ericsson and Simon (1998), “Although verbal descriptions and explanations may not reflect spontaneous thinking with complete accuracy, such verbalizations present a genuine educational opportunity to make students’ reasoning more coherent and reflective.”

From what otherwise is vague and abstract; registers the meaning in one’s memory; and applies the acquired meaning in the future, for understanding within alternative contexts and situations. More importantly, Dewey sees the objective of improving students’ oral and written speech in—that students, who had been accustomed to using speech mainly for practical and social purposes make it “a conscious tool of conveying knowledge and assisting thought.” It is important to notice that with this Dewey sees the improved oral and written speech as a tool not just to convey knowledge but to aid the thinking process itself, hoping that those skills will ultimately serve as “intellectual instruments.” Further support comes from Lev Vygotsky, who in his seminal and much discussed Thinking and Speech posits that “the development of the child’s thinking depends on his mastery of the social means of thinking, that is, on his mastery of speech” (Vygotsky 1987; Smagorinsky 1998). Further, Vygotsky stresses the unity of thought and speech, explaining how these two complement each other by continuously blending together. Indeed, he argues that “Speech does not merely serve as the expression of developed thought. Thought is restructured as it is transformed into speech. It is not expressed but completed in the word. Therefore, precisely because of their contrasting directions of movement, the development of the internal and external aspects of speech forms a true unity” (Vygotsky 1987). Along those lines, according to Smagorinsky (1998), “The processes of rendering thinking into speech is not simply a matter of memory retrieval, but a process through which thinking reaches a new level of articulation.” Finally, according to Ericsson and Simon (1998), “Although verbal descriptions and explanations may not reflect spontaneous thinking with complete accuracy, such verbalizations present a genuine educational opportunity to make students’ reasoning more coherent and reflective.”

From what otherwise is vague and abstract; registers the meaning in one’s memory; and applies the acquired meaning in the future, for understanding within alternative contexts and situations. More importantly, Dewey sees the objective of improving students’ oral and written speech in—that students, who had been accustomed to using speech mainly for practical and social purposes make it “a conscious tool of conveying knowledge and assisting thought.” It is important to notice that with this Dewey sees the improved oral and written speech as a tool not just to convey knowledge but to aid the thinking process itself, hoping that those skills will ultimately serve as “intellectual instruments.” Further support comes from Lev Vygotsky, who in his seminal and much discussed Thinking and Speech posits that “the development of the child’s thinking depends on his mastery of the social means of thinking, that is, on his mastery of speech” (Vygotsky 1987; Smagorinsky 1998). Further, Vygotsky stresses the unity of thought and speech, explaining how these two complement each other by continuously blending together. Indeed, he argues that “Speech does not merely serve as the expression of developed thought. Thought is restructured as it is transformed into speech. It is not expressed but completed in the word. Therefore, precisely because of their contrasting directions of movement, the development of the internal and external aspects of speech forms a true unity” (Vygotsky 1987). Along those lines, according to Smagorinsky (1998), “The processes of rendering thinking into speech is not simply a matter of memory retrieval, but a process through which thinking reaches a new level of articulation.” Finally, according to Ericsson and Simon (1998), “Although verbal descriptions and explanations may not reflect spontaneous thinking with complete accuracy, such verbalizations present a genuine educational opportunity to make students’ reasoning more coherent and reflective.”

Similar evidence in pedagogical literature exists in favor of using writing as a tool for learning. For example, Bracewell and Breuleux (1994) note that “the problem itself is not usually fully defined beforehand …”. Rather, the process of problem definition is in part carried out through the activity of text production as the writer organizes, reorganizes, and elaborates knowledge in the course of writing” (Bracewell and Breuleux, 1994; Smagorinsky 1998). This once again reinforces our conceptual model illustrated in Figure 1; the process of writing, similar to that of speech, not only is an end product resulting from a deep understanding of technical material but itself can serve as a tool for better understanding. According to Ericsson, Krampe, and Tesch-Römer (1993), viewed as a deliberate practice, the act of writing can aid the development of ideas among scientists and serve as a “knowledge-transforming” process rather than just a unidimensional “knowledge-telling” process. Beins (1993) found that as students gain experience in writing they improve their skills in interpreting statistical results, as well as their abilities with statistical computations. Similarly, Bean (2011) argued that writing helps develop the dialogic thinking skills among students and that going through multiple drafts while revising fosters the culture to “formulate, develop, complicate, and clarify their own idea.” Writing is investigative in nature, and the process itself cultivates the character of discovery and exploration, in line with what Foucault poses with his question “If you knew when you began a book what you would say at the end, do you think that you would have the courage to write it?” (Rux 1988). See also Applebee (1981), as well as Wells and Chang-Wells (1992) for support of the use of writing as a tool for exploration.

3. Strategies and Examples for Graduate Students

At the graduate level, Babson College offers both an MBA and several more specialized master’s degrees, including the newly launched MS in Entrepreneurial Leadership (MSEL) that aims to impart skills related to ideating, contextualizing, and launching new ventures to students with diverse undergraduate training (including liberal arts, engineering, and business) and little work experience. This program is delivered over nine months, and this tight timeline puts pressure on courses to be both individually succinct and tightly integrated with one another. The Big Data and Business Analytics course, for instance, took place in a 2-credit format meeting once a week and hosted explicit linkages to disciplines like marketing, operations, and information technology. The course featured both individual and group assessments. For the group-based case reports discussed in this section, students were randomly assigned to groups of four or five.

Developing voice in the analytics process was a major learning goal for the course, and the class’s assessment structure reflected this. The student teams were assigned a series of three case reports, each based around a specific dataset and a particularly analytics context. In the first, students were given a well-known and often used dataset (“Titanic: Machine Learning from Disaster”) relating demographic traits of passengers on the Titanic to whether they lived or died in the subsequent tragedy, and asked to use predictive analytics to tell a story about societal gender and class norms at the time. In the second, students were given a dataset (“Yelp Dataset Challenge”) detailing the characteristics of users of the online recommendation site Yelp, and asked to help a fictive medium-sized restaurant chain in the Phoenix area develop a coherent marketing strategy that would increase its overall revenue. Finally, students were given a collection of data (“Forest Cover Type Prediction”) relating various physical aspects of 30 m by 30 m parcels of land in national forests in the western United States to the dominant species of tree in each parcel, and asked to tell a story about how different physical conditions support different ecosystems and how ecosystems change after events like wildfire. In each case report, students were expected to present a clear rationale for their data pre-processing strategy, a non-technical explanation of the quantitative methods they used, a detailed comparison of their various models in terms of quality metrics of their choosing, a justification of the validity of these metrics, and one-page executive summary tying together the most salient points related to these aspects of the analytics process. Teams typically submitted 10–20 page case reports, together with supplementary tables and figures. In all cases, students were asked to write with an audience of non-technical managers in mind. Given our population of business-oriented professional students, this design decision was quite natural. That being said, asking students to write for an intelligent but otherwise technically ignorant audience can be quite pedagogically useful.
for any student group. By asking students to clearly explain in general terms “the problem, why it matters, potential solutions and the benefits of fixing it,” we are in effect asking our students to maximize the potential impact of their work by couching both their analytics process and its final products in a way that increases overall engagement in a broad audience (Kwok 2013).

We opted for written deliverables instead of oral presentations for several reasons. First and most important, we wanted to provide opportunities for students to develop their voice across the analytics process, and oral presentations in some sense felt too limited in this regard—there is only so much one can talk about in a short presentation. Second, in addition to encouraging students to be more thorough, writing requires students to be more precise—a major change for students with undergraduate degrees in more qualitative disciplines. And third, we knew that students would have a chance to develop their presentation skills and “elevator pitches” at a number of other places in the MSEL curriculum, and so emphasizing writing skills seemed to better fit into the overall mission of the program.

### 3.1. Case Studies

There exists evidence, often from the beginning of the 1990s supporting the need for more active and experiential instruction as part of statistics education, often emphasizing communication as either a natural byproduct or a catalyst for experiential learning. For example, Hogg (1991) voiced disapproval regarding the format of statistics pedagogy which, often being taught (unjustifiably) as a branch of mathematics, lacked what the author called the “plan-do-study-act” iterative cycle that is at the heart of applied statistics. And the scarcity of effective projects in statistics education only fueled the fire. Garfield (1993) arguing in favor of collaborative learning as part of statistics pedagogy, noted that such a collective medium enhances motivation toward the problem being learned. According to the author the synergistic effect of collaborative learning requires the “positive interdependence” where group members “encourage and facilitate each other’s efforts” (Garfield 1993; Johnson 1991). Hakeem (2001) found empirical evidence that participation in semester-long projects improves student performance on examinations, and attributed that to the added value of experiential learning for deepening understanding of statistical inference.

It was clear from the outset that while some students were quite comfortable with the relatively under-defined nature of these projects, others were quite unnerved by the lack of direction and the amount of ambiguity inherent in these problems. This sequence of deliverables was deliberately designed to introduce ambiguity slowly. It begins with the Titanic; every student, regardless of background, knew the saying “women and children first” and that passengers in third class died in much higher numbers than passengers in first class. The challenge, then, was for students to search for the quantitative signatures of these gender and class norms in their predictive model, and then see how well these norms actually predicted the fates of passengers located in the test/validation partition. By beginning with a project in which students already know the conclusion of the story, in which students are universally and equally well-versed, and in which the dataset was of low velocity, volume, and variety, we aimed to both equal the playing field between well-prepared and under-prepared students and to give all students regardless of background a relatively straightforward way to begin developing their voice. Even here, though, we noticed issues related to the professionalism, clarity, and technical correctness in the submissions. Thus, the feedback on this first case report was quite important in setting norms for the remaining two deliverables.

The second project was centered on developing a market segmentation of Yelp users to help a medium-sized restaurant grow its business. The main pedagogical goal of this activity was to introduce a relatively large amount of ambiguity in a context in which students felt comfortable. Uncertainty arose from several places. First, the case prompt was deliberately unclear as to how the restaurant would define a successful outcome. Students were faced with questions such as: “Should the marketing team focus on simply growing the number of customers, or rather increase spending per customer?” or “How much should brand loyalty factor in to an approach?” Students were given only the instruction that any strategy they put forward had to be well-thought out and self-consistent, and then left to their own devices to determine the most pertinent course of action to achieve their goals. This approach forced students both to take ownership of their own learning and to critically examine for a higher-level perspective what makes a Big Data outcome “good” or “bad” in a particular context. The dataset itself also introduced some amount of ambiguity. The data here are somewhat less clean than those in the Titanic dataset, containing both outliers and variables that might better capture user behavior if they were transformed. For instance, instead of considering “Number of Reviews” and “Months Active,” some groups chose to instead include only the ratio as a measure of user activity. This dataset was substantially larger in volume than the one used in the first case report, and the possibility of atypical values and derived fields incrementally increased data variety, as well. As students gained confidence in their ability to give voice to their findings, we as instructors could increase the difficulty of the other aspects of the projects.

The final project was by far the most difficult for students, both in terms of the data at hand and the overall context. It was in some sense the most accurate depiction of a consulting application of data analytics that students would see in the course. The dataset for this case report consisted of a collection of physical attributes (e.g., elevation, distance to water, and amount of shade at certain times) and the dominant tree species for a number of 30 m by 30 m parcels of land in forests in the western United States. Even to students with quantitative backgrounds, the meanings of many of the fields were initially quite opaque. For instance, the dataset contains observations of 40 different soil types, each of which a layperson would be very hard-pressed to understand in any depth. Most of the cohort, especially international students, was not familiar with the tree species they were trying to predict. Even students who were somewhat familiar with these trees did not know much about their preferred habitats. In sum, teams entered the analytics process with much less of an idea where it would take them than in previous case reports. Most teams earnestly engaged
with the material, learning enough of the background material to make sense of the results they were producing. Moreover, by this part of the semester, most teams were able to use the voice they had developed throughout the course to tell a coherent story that wove together both the contextual information related to the problem and the process and product of their analysis, despite the fact that the dataset was both very large in terms of volume and high in variety as compared to those used in previous projects.

3.2. Assessment Methodologies

The traditional method for giving students feedback on written deliverables is through the written comments. While these are undoubtedly useful, we felt that other strategies might provide more and better formative feedback for students as they learned to give voice to the data analytics process. All case reports were graded using a rubric that independently evaluated the following categories:

- visualization: did the figures and/or tables in the case report clearly convey information?
- writing style: was the writing appropriately professional without being overly technical?
- process: were quantitative tools appropriately described and deployed?
- analysis: were the results and the conclusions of the analytics process correctly and concisely interpreted?
- recommendations and context: were the conclusions of the analytics process made operational through the use of clear and actionable recommendations and/or contextual realizations?
- structure: was the overall organization of the document effective in conveying content?

In a preliminary read of each case report, the instructor made brief written comments on each of these areas in the margins of the document. After completing this initial phase, the instructor then recorded audio feedback that both more fully explained each of these comments, and tied comments together into overarching themes, focusing on evaluation areas in which the team did well as well as areas in which the team could improve in the future. These audio recordings, which were each roughly 10–20 min in length, were then distributed to all students via E-mail. Students responded quite positively to this strategy, indicating anecdotally that the audio comments gave a much more complete picture of the team’s overall progress than the written comments alone provided.

Of particular importance to the development of voice were comments related to writing style, recommendations, and structure. Overall, students were relatively comfortable laying out the overall flow of the document, and so comments related to structure were relatively minor. Writing style was a much bigger concern. Writing professionally for a non-technical audience is inherently challenging, and teams erred in exactly the ways one might expect: some groups, usually those in which students had ample prior technical training, produced dense, inapproachable documents, whereas other groups produced documents whose style was not precise enough to accurately convey their findings. Instructor comments focused on helping students find this line between these two extremes by citing individual sentences throughout the document when commenting on writing style. Indeed, this ability to pull from throughout the text when giving overall feedback is one of the major strengths of the oral recording format as opposed to in-line written comments. Many teams also struggled putting their quantitative results in context by making clear and concrete recommendations. Again, we were able to point out when giving feedback individual instances in which teams did not take their quantitative analysis to its conclusion, as well as suggest some possible recommendations that they might have made given the predictive models they had created.

There was a clear progression in quality over the deliverable schedule of the course, with most teams quickly improving from C quality work to B quality work on the writing style component, for instance, between the first and second case reports. We observed similar trends in structure. Students did not make as much progress in their recommendations, primarily because the application area of each case report was different; students might learn how to better couch their cluster analysis in terms of market share in one case report, but this did not help them tie their findings about forest cover type to timber management strategies. These hints at instructor-side trade-off: presenting a diverse collection of application areas comes at the cost of allowing students to develop depth in any one particular area. Overall, we believe that the oral feedback given to students regarding different aspects of their voice allowed for a more nuanced understanding of their strengths and weaknesses and a quicker quality improvement rooted in this understanding.

In the first iteration of the course, case reports were evaluated solely on a group basis; there was no individual assessment associated with the deliverable. Students indicated that this format created free rider issues, in which some group members did not contribute their fair share. To correct for this undesirable side effect of the case report format, we implemented a peer assessment scheme in the second iteration of the course. In hopes of encouraging strong, universal participation in the group activity throughout the semester, students were informed at the beginning of the course that they would be evaluated by their peers twice during the semester. In these feedback surveys, students were asked to both rate each of their team members on numerical scale (an absolute measure) and allocate a total number of points across all team members (a relative measure). In addition to these numerical scores, students were asked to contribute several short responses detailing what each of their teammates and what they themselves contributed to the final product. These responses documented the authoring of specific sections, ownership or certain aspects of the modeling process, and miscellaneous contributions like group organization and copy-editing. In constructing each individual student’s final case report grades, the instructor adjusted the initial group grade upward if a particular student outperformed the norm on both the absolute and relative measures and adjusted the group grade downward if a particular student underperformed both measures. In practice, very few of these adjustments were needed, as the mere presence of peer evaluation seemed to eliminate many of the free rider issues observed in the first iteration of the course.
We have anecdotal evidence from students that practicing their written communication skills helped in their oral communication: students reported discussing these case reports in depth during job interviews and indicated that being able to talk about predictive problems, the context surrounding them, and the analytic solutions they implemented was an excellent way to show content mastery. The course as it has been described here is no longer taught, having been replaced with a joint information technology and Big Data course. As such, we are unable to perform a more controlled assessment of overall student learning outcomes; our conclusion here will unfortunately remain solely anecdotal.

In addition to better data collection, we can suggest several other process improvements to the case report format documented here. If course enrollment is sufficiently small, we might advise assigning a short, individual case report toward the beginning of the term. This assessment would be formative, with the primary objective of providing feedback to individual students across the entire analytics lifecycle before they go on to work in larger teams. This sort of early expectation of individual accountability might help even workload across teams in later projects and further reduce free rider issues. Alternatively, if one wanted a more summative individual assessment near the end of the term, one could substitute a final, individual case report to measure learning outcomes. We might also recommend a more formal peer evaluation process throughout the case report sequence in which teams formally review one or more deliverables from other group or groups. Using this methodology, best practices might diffuse through the cohort organically, and instructors could use peer feedback as a starting point for in-class discussions of these best practices.

At the start of the semester, most students would have been incapable of tackling an open-ended analytics project centered on a large dataset with a diverse collection of fields, even if they had a good grasp of the technical tools they would be using. By allowing students to gradually find their voice over a sequence of case reports that only gradually introduce data volume and variety, we have given a heterogeneous group of students the tools they need to both meaningfully engage in the data analytics process and communicate their findings.

4. Strategies and Examples for Undergraduate Students

4.1. Background on Courses and Audience

Undergraduate students at Babson College who concentrate in analytics are required to take a case-based course covering concepts in and providing hands-on experience with data mining. The syllabus for the course consists of topics in visualization and exploratory data analysis, unsupervised learning techniques such as market basket analysis and clustering, and supervised methods, including k-nearest neighbors (KNN), logistic regression, classification and regression trees, and naïve Bayes. The course is optional for students not concentrating in analytics but is quite popular, with roughly 250 enrollments per year. Another advanced course offered at our institution is the Time Series and Forecasting course that covers concepts such as autocorrelation, smoothing methods, autoregressive integrated moving average (ARIMA) models, time series transformations, and transfer function models. While many of the students taking these courses in fact concentrate in analytics, any student enrolled in these courses is required to have taken an introductory statistics course that includes concepts of basic probability, confidence intervals, and hypothesis testing, as well as a second course in statistics that includes multiple-linear regression.

The majority of the undergraduate students at our institution, by the time of taking the above mentioned courses, have a solid working knowledge of Microsoft Excel, some even are proficient with functionalities such as pivot tables and v-lookups. Classes in our undergraduate program are relatively small with a typical cap of roughly 30 students. Both of the above courses meet twice per week and each class lasts for 95 min. Students are generally extremely motivated, take many challenges arising throughout the course positively, and are ready and willing to go the extra mile to be successful in the course.

The strategies and examples aimed at cultivating communication skills among our undergraduate students were based on mini-cases and were delivered as part of the two courses mentioned above. Although before taking our course students generally had not had prior experience with data pre-processing, their working knowledge of Excel and at times R/RStudio was a significant asset in tackling the data-cleaning challenges that they faced. Consequently, here we ventured to provide datasets with a number of issues in need of resolution. As such the data for those mini-cases had varying degrees of volume, velocity, and variety, and, unlike the strategies described in the previous section, here we had a substantial flexibility to expose students to a relatively high degree of data-related complexity without the hazard of “losing” them during that process.

The examples described below to cultivate the voice among undergraduate students come, respectively, in the form of a short-term intensive and a month-long format. The group-work facet of these assignments deserves a special note. According to survey results provided by Brookfield (2015), an effective way to cultivate critical learning skills is through small group engagements in which group members help evaluate previously unchallenged assumptions and consider new lines of thought. Lemke (1990) adds further support for this social (as opposed to individualistic) aspect of learning by stressing the importance of students explaining and discussing concepts introduced by the instructors in an unfamiliar manner. Furthermore, Bean (2011) notes that when a student explains a concept to a peer she avoids the “status difference” hindrances that would be present were she explaining it to a teacher. Small group engagements also provide practice for public speaking in a more “local” setting before delivering to a larger audience, and the added value is larger for students whose primary language is not English (Johnson 1991; Motley 1998; Bowers 1986; Neer 1987).

4.2. Designing a Spam Filter: A Short-Term Intensive Activity

The size of a dataset is often associated with the number of rows and columns that it has or the amount of physical memory that it occupies. However, often such considerations presuppose that the data under consideration are stored in a single table rather than made up of several (or many) individual
pieces. In many practical situations, the data that organizations provide for client engagements are not compactly represented by a single table but are more often than not scattered over several files or data repositories. The constituents can be large or small, by any standard definition of size. However, regardless of the number of rows and columns that each constituent table has, developing the strategy to effectively bring the constituents together presents additional challenge when working with such scattered information. One may consider the property of data residing in multiple disparate sources as another characteristic of size—again regardless of the numbers of rows and columns of individual tables.

The mini-case to be discussed in this section was part of the case studies in Business Analytics course. Students were faced with the challenge of working with data that, unlike many datasets that students had worked with by the time of the assignment, were not compactly stored in a single table. Instead, different parts of the data for this mini-case were provided in four different sources. The original data used here were adapted from the “Spambase” data donated by George Forman and hosted at the University of California Irvine’s Machine Learning Repository (Blake, Keogh and Merz 1998). The original data, which consisted of 4,601 E-mails represented by 57 attributes, part of which were spam and part legitimate, were stored in a single source which, together with the data dictionary and further information regarding the data can be found at https://archive.ics.uci.edu/ml/datasets/Spambase. Among the 57 attributes provided as the columns of the original dataset was the outcome indicator variable denoting the status of the E-mail (0 for legitimate and 1 for spam). As predictors, there were 48 variables denoting the incidence of individual words in the E-mail (e.g., the variable “Internet” denotes the percentage of words that are “internet”), six variables represent the incidence of individual symbols that may appear in an E-mail (e.g., variable “Dollar Sign” showing the percentage of dollar signs in an E-mail). Additionally, there were three variables denoting, respectively, the average length of an uninterrupted sequence of capital letters, the total length of uninterrupted capital letters, and the length of the longest uninterrupted sequence of capital letters in an E-mail.

For the purposes of introducing variety in this mini-case, prior to the assignment the original data were modified as follows. The data were first broken down to four constituent pieces: the first consisting only of the variables pertaining to the word frequencies, the second containing variables representing character frequencies, the third with variables related to the sequences of capital letters, and finally the fourth table listing the status of each E-mail. In each table five additional variables were created, namely: the name of the sender, the name of the recipient, the hour, minute, and the second when the E-mail reached the recipient’s mailbox. The combination of those five fields represented a primary key for each table. Note that the date stamp is intentionally left out from each E-mail record to introduce yet another subtlety to be exploited. These details behind the preparation of the data were not disclosed to students; it was among students’ responsibilities to ferret out these facts during discussion that would take place in class (discussed below). Tables A1–A4 in the appendix (available in the supplementary materials) provide extracts from each of the resulting four datasets.

Students were required to break into groups and work on the case during a 95-min class. After the completion of the class time students had a day and a half to continue working in groups before submitting a 2- to 5-page report. Students were notified that for the report to be successful it should be: technically sound; intuitive in its explanation of nontechnical material, while clear when describing technical details; and make up a meaningful story as judged by its logical and intuitive “flow.” Further, students were informed that each report should satisfy the following criteria:

- contain an informative background on spam for an audience, who is not familiar with the concept of spam, is not technically sophisticated, but wants to learn;
- clearly explain how and why the data were pre-processed by the team the way it were;
- articulate the details behind predictive modeling of the data;
- include appropriate measures of predictive modeling performance; and
- provide a summary, in the form of a managerial report, of what the analyses illuminate and what the most pronounced findings are regarding the incidence of spam.

Following is a brief description of the typical format of the 95-min class dedicated to this engagement. During the 95-min class time, a thoughtful conversation takes place both within student teams and with the instructor regarding data pre-processing. Instructor engages in a discussion with each team by moderating the discussion related to: the way in which students are thinking of merging the disparate data sources together (e.g., what type of Merge is possible and using which variables(s) as a key); what predictive models is the team thinking of using and why; how is the predictive performance going to be measured and what strategy will the team undertake to select the model that provides the best predictive power among the competing models. During these discussions students have the opportunity to brainstorm, and explain to each other (and to the instructor) particular strategies that they envision to be relevant or useful for the problem at hand. For example, students start by discussing the possibility of joining datasets based on a single field, such as the full name of the E-mail sender. However, through the feedback provided by the instructor, which often is in a form of a follow-up question to pinpoint issues arising from such an approach (e.g., “Do you think the full name of the sender uniquely identifies an E-mail?”), teams discover that such an approach to join would not effectively link records across datasets. After more brainstorming and verbal communication, typically students discover that the only combination of fields that can accomplish the merge of their choice is the combination of the five supplemental fields (mentioned earlier) that were added to each of the datasets.

Notice that what is at play in this process is the interconnectedness of analytics and voice (Figure 1). Indeed, on one hand it is the verbal communication of technical material that gets practiced through the ongoing discussions. Just by talking about data merging or modeling strategies that a student has thought of, and occasionally (and naturally) reformulating a
suggestion or a question to make it more accessible to fellow team members or to instructor, students’ verbal communication skills are expected to improve. It is important to note however, that such a verbal communication in turn leads to a refinement of the technical data management/mining approach, ultimately paving the way for an effective solution. Indeed, a student having a preliminary approach to merging or modeling the data, after having explained that approach to the rest of the team (or to the instructor), soon envisions the possibility of a better, more effective alternative. As such, verbal communication serves not only as one end product of the activity, but also as a powerful catalyst that leads to more acute analytic thinking, and ultimately to the solution.

The feedback provided by the instructor during this discussion exercise is either in the form of follow-up questions intended to guide the student to the right solution path without giving away the solution itself or, in case an argument or suggestion made by a student is not well-formulated, the instructor may ask the student to reformulate the question. Alternatively, the instructor may open a discussion and seek input from the rest of the team members regarding their teammate’s suggestion. Given the discursive structure of the activity, genuine feedback for a purposeful conversation is meant to counter or reinforce a verbal argument through logical means, or to seek further clarification through carefully thought follow-up questions. In an example provided earlier, if a student suggests to join datasets in a way which is in fact ineffective for carrying out the task at hand, one piece of feedback that would contribute to student understanding by promoting further discussion would be for an instructor to ask the follow-up question regarding why/how a student thought that such an approach would accomplish the desired objective. Alternatively, the instructor could open up the question for a group discussion.

Once the instructor receives the written deliverable reports, student teams are notified to prepare for an oral presentation of their work during the following class. Clear guidelines are provided about the format and the breakdown of responsibilities for the presentation (see the appendix). There is a heavy emphasis on clarity of exposition and intuitiveness of the explanations of both technical and non-technical material. Following the team presentation instructor asks each team member questions regarding the implementation of their parts. Often those questions lead to a discussion. For instance, a question may be asked regarding the interpretation as well as the intuition behind a positive and statistically significant parameter estimate for the variable denoting the frequency of the word “free” or a negative and statistically significant parameter estimate for the suffix “edu” in a parameter estimation summary corresponding to a logistic regression to predict the spam status. For an answer to be considered successful for such questions it should be accurate and succinct but intuitive, preferably devoid of statistical jargon. A student answering the question correctly but using statistical terminology, or attempting to answer the question by unnecessarily complicating his or her response, is asked to reformulate the answer and provide a more intuitive explanation, preferably in layman’s terms. At the end of the presentation the instructor provides verbal feedback regarding how well the presenters did, what strategies could be employed to improve the flow of the presentation, and in general which areas should a team concentrate on to bolster their presentational skills prior to the final month-long project. For example, presentations often include visualizations, such as Receiver Operating Characteristic (ROC) curve or a confusion matrix without much reference to them during the presentation. At times the presentation slides are cluttered, have inappropriate font size, and in general are hard to follow. At other times the presenters may be implementing a certain predictive model, for example, a logistic regression, but completely disregarding the insights that could be cultivated from the emerging output, such as the parameter estimation summary. All such issues are clearly communicated to the team during the presentation. In addition, teams are also provided back their written deliverables with detailed feedback by the instructor, including but not limited to the writing style, the organization of the report, the flow and intuitiveness of the writing, the clarity of the business background, as well as the overall soundness of the technical approach. In case there is a paper that significantly exceeds instructor’s expectations, upon getting permission from the respective team it is usually shared with the rest of the class to provide further clarification as to what is considered as excellent writing.

The feedback received from students regarding the effectiveness of this activity is overall very positive. All of the anonymously surveyed students in a class of 17 agreed (in majority of cases strongly) that the mini-project provided an opportunity to improve their communication skills, including verbal communication/discussion with a professor or fellow team members about quantitative material. Similarly, all of the students at the minimum agreed that the mini-project provided them an opportunity to improve their skills of presenting and writing about technical material. When asked whether they would want to see more similar projects in the curriculum, vast majority of students answered affirmatively. The written feedback that was provided in those responses included: appreciation of the opportunity to communicate the analytical thought-process during the preparation and presentation; recognition of the fact that explaining material to peers and attending to their explanations improves understanding of technical material; preference toward a smaller group format; preference that the presentations might better take place outside of class; and interest in having their presentations graded. Students also expressed appreciation of the opportunity that the mini-project provided in terms of preparing them for the larger (final) project. Some students expressed the need to have more time to work on the project.

Further details about this mini-case are provided in the appendix. It is important to note that this mini-case has a preparatory function for the month-long final project that kicks off right after the completion of this mini-case. The feedback provided by the instructor on written deliverables, during in-class discussions, and presentations provides an opportunity for students to observe through concrete examples where their technical and communication skills could be enhanced in preparation of the extensive, month-long final project. Also note that while in the mini-case discussed here the technical
complexities were related primarily to data “volume,” the case discussed in the next section was intended to develop the voice working with data of relatively high “variety.”

4.3. Jennie Maze Limited: A Month-Long Activity

This mini-case was assigned as a final group project toward the end of the Time Series and Forecasting course. By the time students received the project, they had already demonstrated a good working knowledge of time series concepts outlined earlier. This mini-case was part of the three-case series published at the Entrepreneurial Leader Collection. See also Khachatryan (2015), for further details on the use of case-method in undergraduate time series education. The abstract for this case is provided in the appendix. The interested readers may visit the Case Centre (http://www.thecasecentre.org/) for the full description of this and similar cases accompanied by all the pertinent data.

Students are provided the description of a business situation regarding the fictional company Jennie Maze Limited, a student loan lender that is considering strategies to improve profitability and stay reputable in a fiercely competitive business landscape. The company has been experiencing pronounced issues with the management of its call centers. The key performance indicators (KPIs) of Jennie Maze’s call centers based on recent data showed substantial Average Time on Hold and Call Abandonment Rates as is illustrated in Figure 2.

Due to this, Jennie Maze is planning to develop a “demand estimator” that it hopes to use for effectively “guessing” the incoming call volume at its call centers. If successful, the company can leverage the information generated by the demand estimator when scheduling the number and distribution of operators to be working at any given day during the year. Due to a lack of analytics professionals at Jennie Maze the company decided to outsource the work. The student team working on this case is informed that per their submission of a request for proposal (RFP) for a pilot study with Jennie Maze, they have been awarded the contract to build the desired predictive tool.

The team is then provided monthly data on the historical incoming call volumes from one of firm’s call centers, where the company decided to first pilot the desired demand estimator. These data are provided in two Microsoft Excel worksheets. The first worksheet contains monthly data on both answered and unanswered calls made to the call center from January 2002–July 2013 and contains 139 observations. These contain a number of missing values, outliers, and inconsistencies such as text entries, where a numerical entry is expected or an unreasonably large numerical entry in a numeric column. The second worksheet contains 260,000 rows of data on the last month available. These data, however, are not summarized and are provided on the call-level, with each entry being a call ID (text).

The particular characteristic of Big Data being exploited here for pedagogical purposes is the “variety” of data, due to the presence of missing and outlying observations as well as the presence of text (call ID) data in addition to the numerical aggregated call volume information. In addition to this relatively rich variety, the volume of data is also relatively large compared to data that students see before being engaged in this exercise. Receiving the case, the team is required to come up with an effective solution to the problem operating in this data-rich environment. Upon receiving the case, students perform preliminary business research and find out, for instance, that the business is seasonal with most of the quote inquiries done around the spring months. That provides appreciation of the importance of effective call center management in the student loan lending business. At the conclusion of the project the teams present their work to the rest of the class and turn in a written report. Both the presentation and the report are expected to be constructed for the general (layman) audience, except for the explicitly technical areas that need to be delivered in their full technical rigor, such as residual diagnostics resulting from an ARIMA model or time series transformations.

One medium for the development of voice is through the periodic status update meetings that the instructor schedules with students. These are mandatory discussion sessions where students are required to explain to the instructor their work on the project, namely:

- talk about the problem the team is trying to solve;
- explain the way that the company/client would benefit from such a solution;
- illustrate the exploratory data analyses and explain what those illuminate about the data;
- describe the issues with the data and explain how the team plans to resolve those;
- discuss with the instructor the predictive models that were tentatively entertained;

Figure 2. Average time on hold and call abandonment rates for six call centers operated by Jennie Maze. The red vertical lines correspond to fictive (national) industry averages.
• explain what the results obtained from the analyses indicate.

In addition, during the periodic status update meetings team members are required to find out through well-formulated and thoughtful questions the bits and pieces of information that were left out from the case. The format of these discussions closely resembles the verbal communication that often takes place as part of consulting engagements; deeper understanding of data and effective communication of questions go hand in hand. Indeed, after a preliminary appreciation of data a carefully framed question leads to uncovering of further details that in the absence of the inquiry would stay unknown, leading to a better understanding, which yields further questions. This cycle continues until an effective solution emerges. Again, it is worthwhile noting that this approach is anchored in the interconnectedness among analytic thinking and voice as illustrated in Figure 1.

For example, when students inquire about the format of call-level data provided as text, they are explained that the text IDs that contain a given combination of letters and numbers correspond to the calls that are in fact answered by the call center rather than abandoned. What students infer from this information is that if they are somehow able to identify all such values in the column containing 260,000 text entries then they would be able to count the number of such entries toward the aggregate answered call volume for a given month. That in turn leads to an inquiry regarding text functions that would effectively help students scan hundreds of thousands of text entries and check if each contains the desired combination of characters. After students make this follow-up inquiry, they are told by the instructor about the text functions in Microsoft Excel (if none of student proposes that solution) such as “left,” “right,” “mid,” and “len” that can help them accomplish the desired goal. If an inquiry made by a student is not articulate or not well-formulated, instructor notifies the student that the question being asked is not clear, and that the student needs to clarify the inquiry. Alternatively, the instructor may open a discussion seeking further input from team members.

Leveraging the discovered information students proceed to data pre-processing and the development of the demand estimator sought by Jennie Maze. We hasten to stress that each team member receives the same grade for the final written report (unlike the oral presentations that are graded on an individual basis), and at the end of the engagement students submit peer evaluations assessing the effectiveness of their team members throughout the project. It is in each team member’s best interest to take full ownership and be on top of all developments taking place within a team. This often translates into constant communication within the team during when the work is in progress throughout the month.

Of course, the status-update meetings are only intermediate steps and are simply meant to provide ample opportunities for students to become comfortable with communicating intuitively about material that is overly technical in its nature. Students face a larger audience during the group presentations at the conclusion of the project. Each student gets around 5 min to pitch her/his results and receives several questions from the instructor. Although the questions should first be answered by the student to whom the questions are directed, the rest of the team members are encouraged to jump in and provide their perspective on the question being asked. The grade that each student receives for the presentation is based on the clarity of her/his presentation, verbal and non-verbal skills, ability to accurately answer the questions, as well as the punctuality to fit the presentation into the allocated 5-min timeframe. The format of the discussion and the types of questions being asked by the instructor are similar to the ones discussed in the previous section. Students are encouraged to reflect on and incorporate in their written deliverable report their considerations that emerge from feedback provided to them by the instructor during the presentation.

Although no formal assessment was conducted yet, the feedback received from students on this and similar projects in Time Series and Forecasting course were generally very positive. Individual students expressed overall satisfaction with the status update meetings, reporting that such meetings helped them understand data structure and mold the plan of action to be used for modeling. Students also appreciated the multiple moving parts that were part of the project, reporting that those helped them connect the dots and see “the big picture” behind the analytics process involved in time series. Students seemed willing and even excited to participate in meaningful conversations regarding the data and modeling throughout the project, and their reported reaction was a preliminary confirmation that the project was overall well-received. We hope that an appropriately designed, controlled evaluation of learning outcomes will illustrate further benefits of our teaching approach in the near future.

5. Conclusion

John Tukey is thought to have once confessed: “The best thing about being a statistician is that you get to play in everyone’s backyard” (Leonhardt 2000). This demand for statisticians is no doubt quickly rising, fueled by technological innovations that allow organizations to collect massive amounts of data around the clock and to store these data more cost-effectively than ever before. While today being in high demand can still count as one of the advantages of being a statistician, having the technical skillset to model or explore data is only a necessary, not a sufficient skill for a quantitative professional in today’s data-rich era. Analytics inherently presupposes the ability to make sense of data as objectively as possible. But the insights cultivated from data should make sense not only to the statistician working with data (often for many hours), but also to the stakeholders. The latter more often than not do not speak the language of statistics but nevertheless have an urgent need to know the insights emerging from data (often relatively quickly). This in turn requires well-developed and effective communication skills that allow data analysts to richly convey the meaning they have made from the data to any and all stakeholders.

The need for communication skills has motivated our work in devising strategies that could effectively develop students’ voice, both in oral and written communication. Due to previously unimaginable magnitude of data that is at the fingertips of business organizations where students find themselves upon graduation, we found it useful to anchor the development of
communication in the context of Big Data. To do this most effectively, we tailored each case to the student audience at hand, introducing volume, variety and velocity in proportion to students’ prior preparation. Working with graduate students with very heterogeneous backgrounds, one needed to take care in engaging under-prepared students. With their more consistent prior training, undergraduates could be uniformly exposed to varying degrees of volume, variety, and velocity.

Regardless of the background or course/program, we treated the development of voice not as something that would naturally and instantaneously emerge in students as a result of deep understanding of analytics, but instead as a skill that can and should be developed in tandem with analytical thinking. Given the overwhelming literature from educational psychology and pedagogy supporting the unity of thinking and communication, we proposed a pedagogical model where voice and analytics reinforce and flow into each other during the learning process. We hope that the approach to developing communication skills and the strategies we have presented serve as a starting point for other educators as the community works toward better training of our students to learn, inform, and act in an increasingly quantitative professional landscape.

Acknowledgments

The authors thank Richard Cleary and Xiao-Li Meng for the kind invitation to present at a special session on Big Demand for Big Data during the 2016 Joint Mathematical Meetings held in Seattle, WA. A number of ideas underlying the current work were inspired by and motivated during the venue. In addition, authors thank three anonymous referees for the valuable observations and suggestions during the review of this work.

Supplementary Materials

The supplementary material consists of the appendix of this article, which can be accessed on the publisher’s website.

References

Ames, M. (2012), “Tips for Harnessing Hadoop the Right Way,” Sascom Magazine, 3rd Quarter, 6–8.
Applebee, A. N. (1981), “Writing in the Secondary School: English and the Content Areas,” (NCTE Research Report No. 21). Urbana, IL: National Council of Teachers of English.
Bean, J. C. (2011), Engaging Ideas: The Professor’s Guide to Integrating Writing, Critical Thinking, and Active Learning in the Classroom, Hoboken, NJ: Wiley.
Beins, B. C. (1993), “Writing Assignments in Statistics Classes Encourage Students to Learn Interpretation,” Teaching of Psychology, 20, 161–164.
Blake, C., Keogh, E., and Merz, C. J. (1998), “UCI Repository of Machine Learning Databases,” Department of Information and Computer Science, University of California, Irvine.
Bowers, J. W. (1986), “Classroom Communication Apprehension: A Survey,” Communication Education, 35, 372–378.
Bracewell, R. J., and Breuleux, A. (1994), “Substance and Romance in Analyzing Think-aloud Protocols,” in Speaking about Writing: Reflections on Research Methodology, ed. Smagorinsky, Thousand Oaks, CA: Sage Publishing.
Brookfield, S. D. (2015), The Skillful Teacher: On Technique, Trust, and Responsiveness in the Classroom, Wiley.
Chang-Wells, G. L. (1992), Constructing Knowledge Together: Classrooms as Centers of Inquiry and Literacy, Heinemann Educational Books.
Chen, C. P., and Zhang, C. Y. (2014), “Data-intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data,” Information Sciences, 275, 314–334.
Cobb, G. (1992), “Teaching Statistics,” Heeding the Call for Change: Suggestions for Curricular Action, 22, 3–43.
Curriculum Guidelines for Undergraduate Programs in Statistical Science, ASA (2014) http://www.amstat.org/education/pdfs/guidelines2014-11-15.pdf.
Dewey, J. (1910). How We Think, Boston, MA: DC Heath.
Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993), “The Role of Deliberate Practice in the Acquisition of Expert Performance,” Psychological Review, 100(3), 363.
Ericsson, K. A., and Simon, H. A. (1998), “How to Study Thinking in Everyday Life: Contrasting Think-aloud Protocols with Descriptions and Explanations of Thinking,” Mind, Culture, and Activity, 5(3), 178–186.
Franka, B. (2012), Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics, 56, Wiley.
Garfield, J. (1993), “Teaching Statistics using Small-group Cooperative Learning,” Journal of Statistics Education, 1(1), 1–9.
Garfield, J. B., and Gal, I. (1999), “Assessment and Statistics Education: Current Challenges and Directions,” International Statistical Review, 67, 1–12.
Hakeem, S. A. (2001), “Effect of Experiential Learning in Business Statistics,” Journal of Education for Business, 77(2), 95–98.
Hogg, R. V. (1991), Statistical Education: Improvements are Badly Needed, The American Statistician, 45(4), 342–343.
Johnson, D. W. (1991), “Cooperative Learning: Increasing College Faculty Instructional Productivity,” ASHE-ERIC Higher Education Report No. 4. Khachatryan, D. (2015), “Incorporating Statistical Consulting Case Studies in Introductory Time Series Courses,” The American Statistician, 69, 387–396.
Kwok, R. (2013), “Communication: Two Minutes to Impress,” Nature, 494, 137–138.
Lane, D. (2001), “3D data management: Controlling Data Volume, Velocity and Variety,” META Group Research Note, 6, 70. Chicago.
Lemke, J. L. (1990), Talking Science: Language, Learning, and Values, Ablex Publishing Corporation.
Leonhardt, D. (2000), “John Tukey, 85, Statistician; Credited the Word ‘Software’,” NY Times, http://www.nytimes.com/2000/07/28/us/john-tukey-85-statistician-credited-the-word-software.html
McAfee, A., and Brynjolfsson, E. (2012), “Big Data: The Management Revolution,” Harvard Business Review, 90, 60–68.
Motley, M. T. (1988), “Taking the Terror out of Talk,” Psychology Today, 22, 46–49.
Neer, M. R. (1987), “The Development of an Instrument to Measure Classroom Apprehension,” Communication Education, 36, 154–166.
PwC (2015), “Data Driven: What Students Need to Succeed in a Rapidly Changing Business World,” https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf
Radke-Sharpe, N. (1991), “Writing as a Component of Statistics Education,” The American Statistician, 45, 292–293.
Rumsey, D. J. (2002), “Statistical Literacy as a Goal for Introductory Statistics Courses,” Journal of Statistics Education, 10, 6–13.
Rux, M. (1988), “Truth, Power, Self: An Interview with Michel Foucault,” in Technologies of the Self, eds. Martin, Gutman, and Hutton, Amherst, MA: University of Massachusetts Press.
Smagorinsky, P. (1998), “Thinking and Speech and Protocol Analysis,” Mind, Culture, and Activity, 5(3), 157–177.
Smith, G. (1998), “Learning Statistics by Doing Statistics,” Journal of Statistics Education, 6, 1–10.
Vygotsky, L. S. (1987), “Thinking and Speech,” in Collected Works, Eds. R. Rieber, A. Carton, and L. S. Vygotsky (Vol. 1, pp. 39–285). New York: Plenum.