DO YOU HAVE EXPERIENCE? INCORPORATING EXPERIENTIAL LEARNING OPPORTUNITIES INTO STATISTICS EDUCATION IS MESSY BUT IMPORTANT

Nathan Taback
University of Toronto, Canada
nathan.taback@utoronto.ca

"If data analysis is to be helpful and useful, it must be practiced." (Tukey, 1962) Ideally all students would have experiential learning opportunities during their statistics degrees. Incorporating experiential learning opportunities into a course is messy. Experiential learning in data analysis often involves bringing collaborators from industry or other academic problems into the classroom. Aligning learning outcomes of a course and managing a collaborator's expectations presents unique challenges for an instructor. Contests, such as the American Statistical Association (ASA) DataFest, offer an opportunity for a large number of students to have meaningful co-curricular experiential learning opportunities. I will discuss principles of experiential learning in statistics education, and experiences working with industry and academia to bring experiential learning opportunities into the classroom.

INTRODUCTION

Statistical education is undergoing a dramatic change. Among the recommendations on what and how to teach introductory statistics the GAISE report (GAISE College Report ASA Revision Committee, 2016) recommends integrating real data with a context and purpose and active learning. "Our job is to help [students] use data to answer a question that matters … This may be the biggest exogenous challenge to our profession, the least explored in our undergraduate curriculum, and the most promising for rethinking what we teach." (Cobb, 2015) To inculcate statistical thinking Brown and Kass (2009) recommend "real-world problem solving". Incorporating experiential learning opportunities into all levels of a statistics curriculum can help train students to use data to answer questions. The focus of this paper will be to explore several different ways to incorporate experiential learning into statistics education.

EXPERIENTIAL LEARNING IN STATISTICS EDUCATION

There are many definitions of experiential learning. Experiential learning can be defined as "… learning from experience or learning by doing" (Lewis, 1994). Experiential teaching and learning in statistics education involves using data to answer questions, which in turn involves students making connections between questions, data, deciding on and implementing analytical strategies, and communicating findings. Teaching data analysis involves "admit[ing] that it uses judgement … we must teach an understanding of why certain sorts of techniques are indeed useful." (Tukey, 1962). The instructors' main role in experiential learning is creating a safe environment for students to develop judgement in making connections between questions, data, analytical strategies, and communicating findings. One difficulty is that "… data analysis is a highly iterative and non-linear process … in which information is learned at each step, which then informs whether (and how) to refine, and redo the step that was just performed, or whether (and how) to proceed to the next step." (Peng, 2017) This will involve teaching content with multiple styles.

Experiential education is not simply "learning by doing", and "simple participation in a prescribed set of learning experiences does not make something experiential." (Chapman, 1992) A student running a script and examining statistical output is not experiential learning. "Good experiential learning combines direct experience that is meaningful to the student with guided reflection and analysis … it is the learning and teaching process that defines whether a learning experience is experiential…". (Chapman, 1992)

EXPERIENTIAL LEARNING PRINCIPLES FOR STATISTICS EDUCATION

What constitutes experiential learning in statistics education is difficult to define. Chapman (1992) recommends that a good experiential learning and teaching process in statistics

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

education should adhere to principles. Chapman's principles have been adapted for statistical education.

Content

1. Mixture of statistical theory and experience in the application of theory.
2. Problems posed are outside of students' perceived comfort zones.
3. Content that is meaningful for students. There should be meaning for the student that is personally relevant. Data should be used from areas that interest students.

Instructor

4. Absence of excessive instructor judgement. The instructor needs to create a safe environment where students can take responsibility for discovering solutions.
5. Teaching statistics with multiple learning styles. For example, active experimentation (e.g., simulation experiments); concrete experience (e.g., real data); and reflective observation. (Kolb, 1984)

Student

6. Students are given an opportunity to reflect. Content, experience, and reflection are mixed. Students are given an opportunity to reflect on their own learning, bringing "the theory to life" and gaining insight into themselves and their interactions with the world.

Learning Environment

7. The learning environment has opportunities for students to experience statistics working with content or people in other areas.
8. The learning environment creates opportunities for students to become independently motivated. Motivations are not based on what they have to do because someone tells them they must.
9. The learning environment creates opportunities for meaningful relationships. These include student to self, student to teacher, student to learning environment, and student to student.

CREATING EXPERIENTIAL LEARNING OPPORTUNITIES IN STATISTICS

Experiential learning is often divided into co-curricular and curricular. Examples of both are described in this section.

Data Science Competition – an example of co-curricular experiential learning

The framework provided by data science competitions is the 'secret sauce' of the popularity and usefulness of machine learning (Donoho, 2015). The ingredients for data science competitions are a data set provided by a data donor and questions that can be answered with the data.

Data science competitions that focus on prediction follow the Common Task Framework (CTF) (Liberman, 2015). The CTF has these ingredients (Donoho, 2015):

(a) A training dataset that consists of observations with a list of features and a class label for the observation.
(b) Competitors whose common task is to develop a prediction rule from the training data.
(c) Competitors run their prediction rule against a separate test dataset that they don't have access to. The prediction accuracy is automatically calculated by the submitted prediction rule.

All competitors share the *common task* of developing a prediction rule that will have a high prediction accuracy (Donoho, 2015). In prediction problems, the CTF allows for efficient and unemotional judging of winners. Donoho (2015) discusses Liberman's telling of the interesting history of CTF, and how it paved the way for many automatic processes (e.g., Google translate) that we now take for granted.

When questions involve exploratory data analysis instead of prediction then the exploratory Common Task Framework (*e*CTF) is used. The *e*CTF has the following characteristics:

(a) At least one large (>1GB) dataset that consists of observations with a list of explanatory and response variables, although the choice of which is explanatory and which is response is often left open. External data is often allowed to be combined with the provided data.
(b) Competitors whose common task is to use the data to answer a question related to a general theme.
(c) Competitors submit their analyses to a panel of judges or give a short presentation of their analyses.

The common task in the *e*CTF framework is developing a custom problem and answering it using the data sets provided. Competitors have the common task of specifying specific problems, developing answering using the data, and communicating the results.

Data Science competitions can be used to create experiential learning opportunities in statistical education. The American Statistical Association (ASA) DataFest @ UofT [part of ASA DataFest] requires interested students to organize themselves into teams and sign up to participate at the end of the Spring semester. Teams spend a weekend working on a problem that is part of an *e*CTF. Mentors are available throughout the weekend to speak with students and students are given an opportunity to present their solutions at the end of the weekend.

Some challenges in implementing data science competitions are:

- Finding a data donor that is willing to let their data into the public domain.
- Obtaining funding.
- Recruiting judges and mentors.
- Organizing the event (e.g., location, purchasing prizes, food).
- Competitions using the *e*CTF tend to rely on subjective scoring. Analyses are often not reproduced to verify results.

Experiential Learning Within Statistics Courses (i.e., curricular experiential learning)

In experiential classrooms, “students can process real-life scenarios, experiment with new behaviors, and receive feedback in a safe environment. Experiential learning assignments help students relate theory to practice and analyze real-life situations in light of course material” (Lewis, 1994)

There are many ways to incorporate experiential learning opportunities into statistics courses. Many course assignments are ripe for experiential learning and teaching. William Hunter (Hunter, 1975) advocated letting “… students experience first-hand all the steps involved in an experimental investigation-thinking of the problem, deciding what experiments might shed light on the problem, planning the runs to be made, carrying them out, analyzing the results, and writing a report summarizing the work.” In a third-year course on the Design of Scientific Studies at the University of Toronto one of the four class assignments is to conduct a factorial experiment either at home or in a lab. Students are told that they can study anything that interests them, and are advised to pick something they are curious about. This is often the first-time statistics majors experience data collection, analysis, and communication. Even after students have learned the theory many are initially confused when they try to identify the difference between experimental factors and the response. Topics usually involve personal interests or hobbies, and have included baking, exercise, and video games.

Experience with Collaborators

Experiential learning opportunities that involve real-world problem solving require industry or academic collaborators from other disciplines. In the Department of Statistical Sciences (DoSS) at the University of Toronto there is currently an undergraduate capstone course and a graduate level course in statistical consulting where these opportunities are part of the course curriculum.

Some important considerations when working with industry or academic collaborators from other disciplines are outlined below.

The Student Novice. Students’ get the opportunity to play the role of a collaborator with statistical expertise on a multidisciplinary team even though their practical expertise is often closer to a novice at the beginning of the course. During initial meetings with collaborators students usually don't ask questions and are under the impression that they have to "solve the problem" on the spot. It can take several months to teach students to listen, ask questions, then return to the problem, and then return to the problem again. In other words, the process is iterative.

Collaborator Expectations. When engaging with potential collaborators it's important that expectations of students' quality and quantity of work is discussed before engagement with a collaborator. Companies and organizations often want "free" statistical help. The reality is that the statistical help is not "free", and the "cost" is time spent engaging with students. Intellectual property might also be an important consideration.

Intellectual Property Considerations. Intellectual property should be considered early in the process. In academia, there may be authorship considerations, while industry might have concerns around ownership of students' work.

Data Confidentiality. There are many interesting projects, especially in medicine, that involve confidential data. Will the collaborator allow students to work on the data outside of their institution or is the expectation that students will work on the data at computers within their institution? Students working on data at an outside institution, even if it's physically close to campus, can be challenging to incorporate into a course.

Variety of Topics. Obtaining a variety of projects that meet students' interests and have appropriate statistical depth can be challenging. Recruiting external collaborators and vetting projects is time-consuming and may not lead to projects that are meaningful for all students.

Recruiting Collaborators. Advertise within the university including affiliated research institutions and hospitals; respond to companies contacting the department that want to start a collaborative relationship with faculty members; word of mouth.

CONCLUDING REMARKS

Despite the challenges in developing experiential learning opportunities it is rewarding for students and instructors, and an important part of a modern statistics curriculum. A principled approach to experiential learning should be used to design curricular and co-curricular opportunities. Teaching students how to use data to solve real-world problems is a hallmark of experiential learning in statistics education.

REFERENCES

Brown, E. N. (2009). What is statistics? *The American Statistician, 63*(2), 105-110.

Chapman, S. M. (1992). What is experiential education? *Journal of Experiential Education, 15*(2), 16-23.

Cobb, G. (2015). Cobb, G. (2015). Mere renovation is too little too late: We need to rethink our undergraduate curriculum from the ground up. *The American Statistician, 69*(4), 266-282.

Donoho, D. (2015). *50 years of Data Science.* Retrieved from http://courses.csail.mit.edu/18.337/2015/docs/50YearsDataScience.pdf

GAISE College Report ASA Revision Committee. (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report 2016.* American Statistical Association. Guidelines for Assessment and Instruction in Statistics Education College Report 2016.

Hunter, W. (1975). *101 Ways To Design an Experiment, or Some Ideas About Teaching Design of Experiments.* The University of Wisconsin-Madison.

Kolb, D. (1984). *Experiential learning: Experience as the source of learning and development.* FT press.

Lewis, L. H. (1994). Experiential learning: Past and present. *New directions for adult and continuing education, 62*, 5-16.

Liberman, M. (2011). *Lessons for Reproducible Science from DARPA's Progams in Human Language Technology.* Retrieved from https://web.stanford.edu/~vcs/AAAS2011/AAAS2011Liberman.pdf

Liberman, M. (2015, April 1). Reproducible Research and the Common Task Method. New York, NY.

Peng, R. D. (2017). *The Art of Data Science. A Guide for Anyone Who Works with Data.* Skybrude Consulting.

Tukey, J. W. (1962). Tukey, John W. "The future of data analysis. *The annals of mathematical statistics, 33*(1), 1-67.
University of Toronto. (2016). *2017 ASA DataFest @ UofT*. Retrieved November 10, 2017, from ASA DataFest @ University of Toronto: https://utorontodatafest.wordpress.com