



Assessment-via-Teaching: Exploring an Alternative Assessment Strategy in Undergraduate Introductory Data Science Course

Lujie Karen Chen

lujiec@umbc.edu

University of Maryland, Baltimore County
Baltimore, MD, USA

Justin Thai

jthai2@umbc.edu

University of Maryland, Baltimore County
Baltimore, MD, USA

ABSTRACT

Learning-by-teaching is an active learning method that has the promise of engaging students and enhancing their learning. These benefits include the improvement of the metacognitive process, increased motivation and self-efficacy, and the opportunity to refine communication skills. These benefits are particularly valuable in data science education. Though not as widely studied, teaching assignments, where students demonstrate competency through teaching others can be used as a formative assessment tool. This paper describes an "alternative midterm" experiment conducted in an undergraduate introductory data science class. In this pilot, students were asked to demonstrate their newly learned data science skills by teaching them to individuals without data science backgrounds. The evidence of learning will be illustrated through an in-depth analysis of the students' teaching products. These products consist of two parts: (1) "Teaching Preparation" materials, such as Python Notebooks containing problems and solutions crafted by the students, and (2) "Teaching Implementation", video recordings of real-time interactions between the students and their "tutees." The paper will also present qualitative feedback from the students, highlighting promising signs of engagement and acceptance. We will conclude by discussing the challenges and opportunities of implementing this assessment strategy on a larger scale.

ACM Reference Format:

Lujie Karen Chen and Justin Thai. 2024. Assessment-via-Teaching: Exploring an Alternative Assessment Strategy in Undergraduate Introductory Data Science Course. In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2 (SIGCSE 2024)*, March 20–23, 2024, Portland, OR, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3626253.3635516>

1 INTRODUCTION

Real-world data science requires proficiency in technical skills, such as computation and quantitative reasoning, and the ability to effectively communicate with domain experts and audiences unfamiliar with data science. The question arises: how can we accurately assess these essential skills? The learning-by-teaching approach is an active learning technique known for its multi-fold benefits. It enhances the metacognitive process, motivates learners, improves

learning strategies, and provides opportunities to sharpen communication skills [1]. Assessment-via-Teaching, where students are asked to demonstrate their competency through teaching those without data science backgrounds, can be used as a formative assessment strategy, offering insights into students' multi-facet mastery, including content knowledge and communication skills. The computing education community has explored non-traditional assessment methods, such as analyzing recorded code-tracing videos [2], to gain a deep understanding of students' learning. This research contributes to the emerging literature on alternative assessment strategies for complex and rich assessment objectives involving technical skills such as programming and quantitative reasoning and soft skills to communicate to general non-technical audiences.

2 COURSE CONTEXT

This pilot took place in an introductory data science course at the undergraduate level in Fall 2022 at a 4-year Minority Serving Institute in the US. The curriculum was adapted from the Data 8 Introduction to Data Science from the University of California, University of California, Berkeley. This curriculum focuses on developing basic data science concepts in statistical inferences and computation at the same time. This course serves as the first introduction to data science, with almost no prerequisite. Students only need to have a math level of high school Algebra I, and no prior experience in programming is required. The course is open to students from all majors from across the campus. The adapted version is a three-credit course with two and a half hours of in-person instruction over 16 weeks. The first half of the course focuses on basic table functions and plotting functions, and the second half focuses on statistical inference using computational simulation. The teaching assignment described in this paper serves as the mid-term formative assessment of the material covered during the first half of the course. Enrollment is typically capped at 25 students and with the support of undergraduate teaching fellows who took the class before. Fall 2022 is the fifth offering of this course at this institute, with 25 students from a diverse range of majors across the campus and over 80% students are from minority backgrounds.

3 TEACHING ASSIGNMENT

Note: We use the following naming convention to avoid confusion. In the context of discussion of the course, the teacher of the course is called the "instructor", and the learner in the course is called the "student"; while in the context of teaching assignment, the data science student who plays the teaching role is called "tutor" and his/her "student" is called "tutee".

Figure 1 gives an overview of the timeline of the teaching assignment involving four major steps.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SIGCSE 2024, March 20–23, 2024, Portland, OR, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0424-6/24/03.
<https://doi.org/10.1145/3626253.3635516>

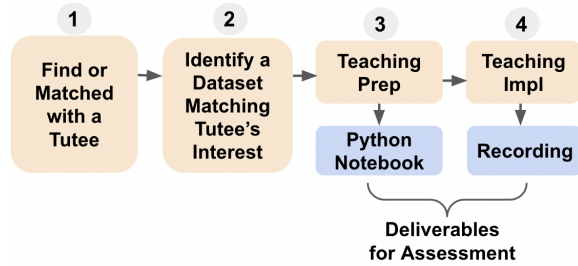


Figure 1: Overview of the Timeline of the Teaching Assignment

- (1) **Find or Matched with a Tutee.** Tutees can be anyone from the broader campus community, neighborhood, or circle of friends, regardless of profession and educational background. The only criterion is that he/she does not have prior knowledge or experience with data science or programming. Students may find a tutee on their own or be matched to one by the instructor who maintains a list of undergraduate students outside data science disciplines who express interest in this activity. The matching is mostly random, though it might consider the common interests between the tutor and the tutee if information is available.
- (2) **Identify a Dataset.** Students work with his/her tutee to identify a dataset that matches the tutee's interest. The instructor provides a list of publicly available dataset that could be used, but students are free to find dataset on their own. The instructor reviewed the selected datasets to ensure they were with appropriate content and level of complexity.
- (3) **Teaching Preparation** Students develop teaching preparation materials in Python Notebook, which is the primary programming environment for the course. The regular assignments are hosted on a private JupyterHub at the institute. We use Google Colab Python Notebook to accommodate tutees from outside the class. A tutorial was provided to students to familiarize them with the Google Colab. To scaffold the teaching preparation, students were asked to propose five "curious questions" about the dataset (encouraged to work with a tutee to get those questions) and then write code to answer those questions. A class period was devoted to demonstrating the development of a teaching notebook in Google Colab and teaching using the notebook.
- (4) **Teaching Implementation.** Students implement teaching with their tutees. The teaching modalities are flexible: students may teach online or in person. The lesson typically lasted between 30 to 45 minutes. Students are asked to record the lessons to be used for assessment.

4 ASSESSMENT STRATEGY AND FORMATIVE FEEDBACK

The assessment rubric comprises two main components: (1) teaching preparation and (2) teaching implementation, with each contributing 50% of the final grade. Under the teaching preparation category, we evaluate two key elements: the quality of the questions

posed by students regarding their interestingness, insightfulness, and usefulness and the solutions regarding their correctness, appropriateness, and the variety of table or plot functions employed. For the teaching implementation category, we review the submitted teaching video to determine if the student can effectively explain concepts and demonstrate understanding. Three teaching assistants and instructors graded the assignment. Each assignment was graded by two graders, and any significant discrepancies in evaluations were discussed and resolved. In addition to the quantitative scores assigned for each category, we also give students qualitative feedback. This includes highlighting specific segments of the teaching video where the student explained a concept well, demonstrated clear evidence of learning, or revealed misconceptions.

5 STUDENTS' REFLECTION

We analyzed the students' responses to the open-ended question about their teaching experience: "How might the teaching experience help you become a better learner?" From the analysis, several themes emerged: (1) Teaching encourages retaining material through review, practice, and reinforcement; (2) Students discovered new methods for solving problems; (3) Many gained increased confidence in their data science and/or communication skills; (4) Teaching appears to enhance metacognition. As one student pointed out, *"The process of making a lesson plan could be applied to making a learning plan. It is the same. Identify what you want to learn, find what things you would most need to know, and learn them."* (5) Teaching motivates active learning. One student noted, *"This teaching experience helped me to see the perspective of the teachers. So, I can understand how frustrated they can get, especially if the student is not paying attention or that they are not even trying to participate. So, I think I can be a better learner by attempting to participate more often and asking questions. And try to pay attention better."*

6 DISCUSSION

From this pilot study, we have gathered preliminary evidence supporting the effectiveness of this assessment method. Students' reflections indicate that this approach potentially enhances their technical skills and fosters metacognition, self-efficacy, and communication. However, we have identified several challenges. For instance, considerations must be made for students for whom English is a second language (ESL) and those who may be naturally shy. The teaching performance may depend on the background of the tutees, which are hard to control for. Additionally, we recognized potential difficulties in grading when scaling up this approach. These observations motivate the need for future work to develop more equitable and efficient assessment strategies for complex skill sets as those we attempt to evaluate.

REFERENCES

- [1] David Duran. 2017. Learning-by-teaching. Evidence and implications as a pedagogical mechanism. *Innovations in Education and Teaching International* 54, 5 (2017), 476–484.
- [2] Rachel S Lim, Joe Gibbs Politz, and Mia Minnes. 2023. Stream Your Exam to the Course Staff: Asynchronous Assessment via Student-Recorded Code Trace Videos. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1*. 144–150.