Practical Training in Reproducible Methods for Undergraduates: Scaling the AEA Reproducibility Verification into the Classroom

Lars Vilhuber, Takshil Sachdev, Ilanith Nizard
Cornell University

Abstract

For the past 8 years, the LDI Replication Lab has verified the **computational reproducibility** of papers appearing in AEA journals. Over **200 undergraduate** students have participated in this activity. They received intensive training at the start of their participation and gained experience throughout. This poster will identify how to scale this particular way of teaching reproducibility into a classroom-based, semester-long activity. We believe, based on feedback from "graduates" of the Lab, that the experience of hands-on engagement with data and code for conditionally accepted manuscripts is highly beneficial to the education and professional development of the students employed as Replication Interns (see also Vilhuber et al, 2022; Ball, 2023).

Introduction

Richard Ball asked the (rhetorical) question "Is it feasible to include reproducible research methods in undergraduate training?" (Ball 2023), answering that question with "Yes we can!" Exposure to and training in well-structured programming will benefit students working both in academia and the private sector. While this type of training does not turn students into active programmers (without further coursework), they will become **resourceful in debugging** (inevitably) imperfect code submitted to journals. They are also taught how to **communicate** with other people on technical problems, in an objective and constructive manner. They practice this while preparing reports for authors. Such skills are useful in academic and private sector settings for both technical and non-technical feedback mechanisms.

Background

Over the past 8 years, we have trained **200+ students**, in 3 training sessions per year, who have subsequently helped produce **4,470 reports on 2,791 articles** (Vilhuber, 2025).
The majority of undergraduate students were recruited at **Cornell University**, from across STEM and social science fields, mostly sophomores and juniors. We have expanded the search beyond Cornell in the past three summers, running an internship program with 20 non-Cornell students from various universities and colleges (**Haverford, Bryn Mawr, Wellesley, Notre Dame, U Colorado Boulder**). Through this process, we have identified the necessary pre-requisites and technical tooling to conduct reproducibility verification with students.

Learning Objectives and Tasks

Our extended curriculum introduces students to core components of reproducible research, including command-line skills, integrating text and code, high-performance computing, data ethics, and working with confidential data. Additional modules cover archiving, grant writing, and preparing referee-style evaluations. The classroom version of the program focuses on the following practical skills:

Data acumen: Students analyze data provenance as described by authors, must assess the feasibility of accessing all data; may be asked to obtain data from a variety of databases/ request protocols; learn to work with data that has distribution restrictions (see Figure 1).

Technical evaluation: Students must verify their ability, within the range of the technical means at their disposal, to computationally reproduce results. They must also recognize when they are not the most qualified person to do the task.

Skeptically follow instructions: Students attempt reproduction as per instructions (README) but must also understand when not to follow instructions (b/c incorrect/inefficient).

Run and debug code in multiple software environments: Students are selected based on prior exposure to at least 1 software environment, but must run and debug code in others as well. Learn about commonality in software design approaches; critically evaluate sources of debugging information (help files, internet, colleagues). Working with unfamiliar (HPC, Windows!) environments

Objective and empathetic technical writing: Students must prepare report, communicating what went right and what went wrong; understanding that they are "communicating up" to a more senior individual; provide constructive criticism where appropriate.

Lars Vilhuber, Takshil Sachdev, Ilanith Nizard
(Cornell University)
Corresponding author:
lars.vilhuber@cornell.edu **or** lv39@cornell.edu

Challenges

1)Obtaining **timely but realistic manuscripts** and replication materials. Initial training can happen on recently published articles. Manuscripts from journals without data editors might provide a realistic juxtaposition. For data acumen and assessment, published articles are sufficient, and anecdotally, may be easier to use from journals **with** data editors.
2)Using a **sufficiently large cross-section of articles** in a classroom setting. In our experience, students acquire many of the needed skills within about 5 articles. Sampling might not yield the necessary diversity in difficulty; curation of selected articles is time-consuming. It may be possible to use input from various data editors.

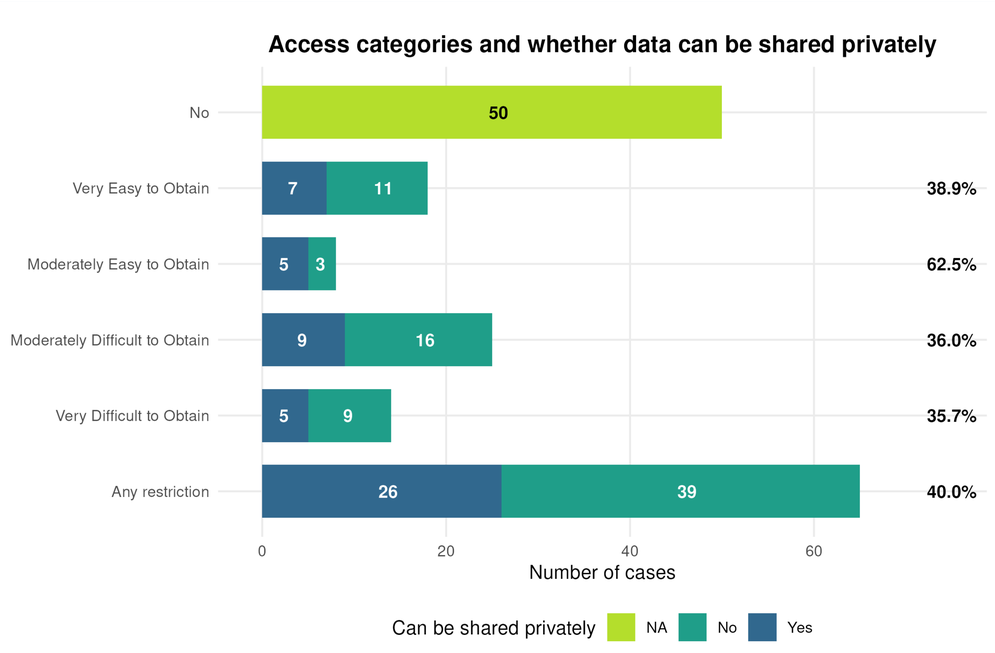

Figure 1. Distribution of Data Access in AER articles in 2024.

Discussion

We think it is possible to organize a **"pool" of articles** to which authors, wishing to submit to journals, can submit their code. Classes would draw from this pool, and prepare a realistic response within 2-3 months. A pool of "editors", which need not be the instructor themselves, can handle reviewing the real reports that get sent back to real authors. This is similar to internal efforts at various institutions (Arguillas et al, 2022; Jones, 2024), and similar to various CS and DS classes that work with real external clients (f.i., Paloian, 2022).

Some outcomes

2020 sociology graduate working for a nonprofit research organization: "[I received] overwhelmingly positive feedback on my documentation method in code reviews."	**2021 Economics graduate, pursuing a Ph.D. in Political Economics** "… I feel like [LDI Replicator position] was the single most important thing … to prepare myself to succeed in [predoctoral fellowship]…."	**2024 Economics intern** "For every issue I did solve, there was a gratifying moment, knowing I've explored something new and that the authors would read and heed my documented solution."

Conclusions

We believe that such a classroom activity can be implemented in many colleges and universities at the undergraduate level, and hope to discuss with attendees of the poster session.

References

1. Vilhuber, Lars, Hyuk Harry Son, Meredith Welch, David N. Wasser, and Michael Darisse. 2022. "Teaching for Large-Scale Reproducibility Verification." Journal of Statistics and Data Science Education 30 (3): 274–81. https://doi.org/10.1080/26939169.2022.2074582.
2. Ball, Richard. 2023. "'Yes We Can!': A Practical Approach to Teaching Reproducibility to Undergraduates." Harvard Data Science Review 5 (3). https://doi.org/10.1162/99608f92.9e002f7b.
3. Vilhuber, Lars. 2025. "Impacts of the AEA Verification Process". https://aeadataeditor.github.io/aea-cumulative-summary/impacts_of_aea_data_editing.html, accessed on 7 Dec 2025.
4. Arguillas, Florio, Thu-Mai Christian, Mandy Gooch, Tom Honeyman, Limor Peer, and CURE-FAIR WG. 2022. 10 Things for Curating Reproducible and FAIR Research. June 27. https://doi.org/10.15497/RDA00074.
5. Jones, Maria. 2024. "Introducing Reproducible Research Standards at the World Bank." Harvard Data Science Review 6 (4). https://doi.org/10.1162/99608f92.21328ce3.
6. Paloian, Sierra, Kirsten Doehler, and Alexandra Lahetta. 2022. "Implementing a Senior Statistics Practicum: Lessons and Feedback from Multiple Offerings." Journal of Statistics and Data Science Education 30 (2): 114–26. doi:10.1080/26939169.2022.2044943.