Data Citations and Reproducibility  
in the Undergraduate Curriculum

Diego Mendez-Carbajo, Ph.D.

Alejandro Dellachiesa, Ph.D.

Data citations are the foundation of reproducibility. To develop reproducibility skills among undergraduate students we must start with basic data literacy skills such as citing data consistently.

**Disclaimer**: The views expressed in this article are those of the authors and don't necessarily reflect the position of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

## Introduction

The scholarship of teaching and learning in economics documents multiple efforts to bring the quantitative dimension of our professional work closer to the undergraduate college curriculum.

Economics educators describing data-focused assignments and projects (Wolfe, 2020; Halliday, 2019; Wuthisatian and Thanetsunthorn, 2019; Marshall and Underwood, 2019; Mendez-Carbajo, 2015 & 2019) highlight the data-finding step of these projects. Even when the datasets are directly provided to the students, (e.g., Easton, 2020) the instructors emphasize the broader literacy dimensions of the assignments. However, there is neither professional consensus about how to build data-literacy skills (Wuthisatian and Thanetsunthorn, 2019) or much actual research on their mastery among economics students (Halliday, 2019).

In this chapter, we document baseline proficiency levels among undergraduate college students related to identifying data series and their sources. We also put forward an accessible pedagogical strategy to develop basic reproducibility skills.

We argue reproducibility should be part of the undergraduate curriculum in economics because it is a valuable professional skill to be developed throughout the curriculum by consistently citing the data sources used in economic arguments. We must instill the practice leading by example and enrolling the help of librarians

## Expected Proficiencies

There is a natural overlap regarding the development of data-literacy skills between economics and library science: both disciplines value it and contribute to its development.

The two seminal descriptions of data literacy expected proficiencies among undergraduate students are provided by Hansen (2012) and Pothier and Condon (2019). The first of the seven broad competencies of economics majors named by Hansen directly address data provenance. It states: “Access existing knowledge: […] Track down economic data and data sources. Find information about the generation, construction, and meaning of economic data.”

The library science perspective provided by Pothier and Condon is articulated through seven expected data competencies of economics and business majors. The last one states: “Data ethics: The principles of data ethics are built on data ownership, intellectual property rights, appropriate attribution and citation, and confidentiality and privacy issues involving human subjects.”

The utilitarian and ethical aspects of data reproducibility outlined above are bridged by the American Economic Association’s (AEA) (2020) Data and Code Availability Policy, which clearly states “All source data used in the paper shall be cited, following the AEA Sample References.” However, the scholarship documenting the collaboration in this area between instructional economics faculty and librarians is limited. Neither the calls by economics instructors (McGrath and Tiemann, 1985; Li and Simonson, 2016; Mendez-Carbajo, 2016) nor the experiences documented by librarians (Wheatley, 2020; Wilhelm, 2021; Waggoner and Yates Habich, 2020) appear to have broad impact.

## Evidence of Broad Data Literacy Skills

Mendez-Carbajo (2020) documents baseline levels of data literacy competency in several areas key to the accurate and ethical use of data for communication and decision-making among high school and college students.

In the online economic education module produced by the Federal Reserve Bank of St. Louis “FRED Interactive: Information Literacy”, two separate groups of high school students (*N*= 450) and college students (*N*= 912) answer seven pre-test questions. The questions are mapped to the data literacy competencies described by both Pothier and Condon (2019) and Hansen (2012).

The analysis finds effectively identical levels of average baseline data literacy competency between high school and college students. However, it also documents much higher levels of perceived self-efficacy among college students than among high school students. In other words, college students are no more knowledgeable or skilled than high school students but are significantly more confident in their work. This finding highlights a major challenge for instructors working to develop the expected proficiencies identified in the literature: the average college student is unduly comfortable in their limited understanding of the primary sources of economic data.

## Evidence of Narrow Reproducibility Skills

During the fall semester of 2020, we distributed a short online assignment to all 854 students enrolled in two different upper-division economics courses offered by a large public university in the United States.

On average, the students are slightly above 20 years of age, 49% identify themselves as female, 21% identify as non-White racial or ethnic minorities, and 92% report English is their native language. Academically, 87% of students are business, economics, or finance majors and hold a grade point average of 3.41. Also, 68% of students are currently enrolled in a statistics course required by their program and, on average, have previously completed more than one and a half economics courses.

The assignment had three sections:

* First, the students were directed to read a brief, 900-word, essay on [how to create data citations with FRED®](https://research.stlouisfed.org/publications/page1-econ/2020/10/21/data-citations-with-fred). This essay provided background on the value of good data citations for practitioners of economics and could be used as reference material for the next two sections of the assignment.
* Second, the students were directed to read two short --under 600 words, economic essays. See them [here](https://files.stlouisfed.org/files/htdocs/publications/es/04/ES0427.pdf) and [here](https://files.stlouisfed.org/research/publications/economic-synopses/2020/10/08/renewable-sources-of-electricity-where-excess-capacity-is-built-in.pdf). Each included a line graph of economic data. In the text, the authors referenced the data series and their sources while interpreting the quantitative information presented in the graph.
* Third, the students were asked to complete three tasks: identify the data series discussed in the essay; identify the sources of the data series discussed in the essay; and identify the missing elements of a data citation provided in the essay.

The assignment was completed in its entirety by 501 students. Table 1 reports our findings.

**Table 1**. Data Literacy Skills

|  |  |  |
| --- | --- | --- |
| **Scores, Misconceptions and Errors** | **Essay A** | **Essay B** |
| Identifies Series Correctly | 0.57 | 0.47 |
| Identifies Sources Correctly | 0.21 | 0.03 |
| Identifies Incomplete Citation | 0.18 | -0.04 |
| Can’t Identify Sources | 0.05 | 0.12 |
| Confuses Source with Distributor | 0.72 | 0.73 |
| Considers Citation to be Complete | 0.25 | 0.40 |

Note: Data Literacy Scores: 𝑆𝑐𝑜𝑟𝑒= (#𝐶𝑜𝑟𝑟𝑒𝑐𝑡 𝐴𝑛𝑠𝑤𝑒𝑟𝑠 − #𝐼𝑛𝑐𝑜𝑟𝑟𝑒𝑐𝑡 𝐴𝑛𝑠𝑤𝑒𝑟𝑠) / (#𝐶𝑜𝑟𝑟𝑒𝑐𝑡 𝐴𝑛𝑠𝑤𝑒𝑟𝑠)

We document very weak student data literacy competencies associated with narrow reproducibility skills. Data literacy scores related to correctly identifying the sources of the data or recognizing an incomplete data citation are very low. Moreover, we document a frequent misconception of confusing the data source with the distributor.

These findings have practical implications for instructors, whether they are librarians or economic educators. Our work suggests there is a substantial instructional opportunity to help students develop the ability to recognize data series and their sources. In that regard, disambiguating the roles of data distributors and data sources can potentially yield large benefits to students, who would be able to acquire a more sophisticated understanding of how data are created and made available.

## Proposed Instructional Intervention

We propose a broad instructional intervention for economics instructors reflecting the fact that correctly citing the data is a foundational literacy skill.

* Lead students by example and consistently name the sources of all data referenced or used in your teaching.
* Embed this practice in all your teaching, regardless of the type or subject of the course.
* Enroll the help of librarians by leveraging their ongoing instructional outreach on information literacy to include data citations.

Proficiency in identifying data sources is foundational to the development of reproducibility skills. The earlier and the more frequently students are exposed to best practices in data citations, the more effortlessly they will be able to adopt sophisticated professional replicability practices.

## Conclusion

Reproducibility should be part of the undergraduate curriculum in economics:

* It is a valuable professional skill that shows the background work that goes into doing economic research. Citing the sources of the data makes research work more thorough.
* This skill should be developed throughout the curriculum. This skill is not particular or exclusive to econometrics or statistics courses.
* The first step is to consistently cite the data sources used in economic arguments. This includes data tables, plots, and in-text references.
* We must instill the practice by leading by example. Economics educators should enroll the help of librarians in developing this skill among students.

## References

American Economic Association. (2020). Data and Code Availability Policy. https://www.aeaweb.org/journals/data/data-code-policy.

Easton, T. (2020). Teaching econometrics with data on coworker salaries and job satisfaction. *International Review of Economics Education*, 34, 100178. DOI 10.1016/j.iree.2020.100178.

Halliday, S. D. (2019). Data literacy in economic development. *The Journal of Economic Education*, 50 (3), 284-298, DOI: 10.1080/00220485.2019.1618762

Hansen, W. L. (2012). An expected proficiencies approach to the economics major. In *International handbook of teaching and learning economics*, ed. G. Hoyt and K. McGoldrick, 188–94. Cheltenham, UK and Northampton, MA: Edward Elgar.

Li, I., and Simonson, R. D. (2016) The value of a redesigned program and capstone course in economics. *International Review of Economics Education*, 22, 48-58, DOI: 10.1016/j.iree.2016.05.001.

Marshall, E. C., and Underwood, A. (2019). Writing in the discipline and reproducible methods: A process-oriented approach to teaching empirical undergraduate economics research. *The Journal of Economic Education*, 50 (1), 17-32. DOI: 10.1080/00220485.2018.1551100

McGrath, E. L., and Tiemann, T. K. (1985). Introducing empirical exercises into principles of economics. *The Journal of Economic Education*, 16 (2), 121-127. DOI: 10.1080/00220485.1985.10845107

Mendez-Carbajo, D. (2015). Visualizing data and the online FRED database. *The Journal of Economic Education*, 46 (4), 420-429. https://doi.org/10.1080/00220485.2015.1071222

Mendez-Carbajo, D. (2016). Quantitative reasoning and information literacy in economics. In *Information Literacy: Research and Collaboration across Disciplines* (pp. 305-322), Barbara D’Angelo, Sandra Jamieson, Barry Maid, and J anice R. Walker (editors). Perspectives on Writing. Fort Collins, Colorado: WAC Clearinghouse and University of Colorado Press. https://wac.colostate.edu/books/infolit/chapter15.pdf

Mendez-Carbajo, D. (2019). Experiential learning in macroeconomics through FREDcast. *International Review of Economic Education*, 30 (1). DOI: 10.1016/j.iree.2018.05.004.

Mendez-Carbajo, D. (2020). Baseline competency and student self-efficacy in data literacy: Evidence from an online module. *Journal of Business & Finance Librarianship*, 25:3-4, 230-243. DOI: 10.1080/08963568.2020.1847551

Pothier, W., and Condon, P. (2019). Towards data literacy competencies: Business students, workforce needs, and the role of the librarian. *Journal of Business and Finance Librarianship* 25:3-4, 123-146. DOI: 10.1080/08963568.2019.1680189

Waggoner, D., and Yates Habich, B. (2020). Collaboration is the key: faculty, librarian and Career Center professional unite for marketing class success. *Journal of Business & Finance Librarianship*, 25:1-2, 82-91. DOI: 10.1080/08963568.2020.1784658

Wilhelm, J. (2021). Joint venture: An exploratory case study of academic libraries’ collaborations with career centers. *Journal of Business & Finance Librarianship*, 26:1-2, 16-31. DOI: 10.1080/08963568.2021.1893962

Wheatley, A., Chandler, M., and McKinnon, D. (2020). Collaborating with faculty on data awareness: A case study. *Journal of Business & Finance Librarianship*, 25:3-4, 281-290. DOI: 10.1080/08963568.2020.1847553

Wolfe, M. (2020). Integrating data analysis into an introductory macroeconomics course. *International Review of Economics Education*, 33, DOI: 10.1016/j.iree.2020100176

Wuthisatian, R., and Thanetsunthorn, N. (2019). Teaching macroeconomics with data: Materials for enhancing students’ quantitative skills. *International Review of Economics Education*, 30, 100151. DOI 10.1016/j.iree.2018.11.001.