

We Need More Reproducibility Content Across the Computer Science Curriculum

Fraida Fund

ffund@nyu.edu

NYU Tandon School of Engineering
Brooklyn, New York, USA

ABSTRACT

With increasing recognition of the importance of reproducibility in computer science research, a wide range of efforts to promote reproducible research have been implemented across various sub-disciplines of computer science. These include artifact review and badging processes, and dedicated reproducibility tracks at conferences. However, these initiatives primarily engage active researchers and students already involved in research in their respective areas. In this paper, we present an argument for expanding the scope of these efforts to include a much larger audience, by introducing more reproducibility content into computer science courses. We describe various ways to integrate reproducibility content into the curriculum, drawing on our own experiences, as well as published experience reports from several sub-disciplines of computer science and computational science.

CCS CONCEPTS

• **Social and professional topics** → **Computing education**; • **General and reference** → **Experimentation**.

KEYWORDS

reproducibility, education

ACM Reference Format:

Fraida Fund. 2023. We Need More Reproducibility Content Across the Computer Science Curriculum. In *2023 ACM Conference on Reproducibility and Replicability (ACM REP '23)*, June 27–29, 2023, Santa Cruz, CA, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3589806.3600033>

1 INTRODUCTION

A wide range of initiatives to promote reproducibility have been proposed and implemented across various sub-disciplines of computer science and computational science. Many publication venues now include a process for submitting, reviewing, and badging reproducibility artifacts alongside papers that present novel research results. Similarly, flagship conferences in a variety of computer science fields now include dedicated tracks for papers that replicate or reproduce a previously published work.

Permission to make digital or hard copies of all or part 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 components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ACM REP '23, June 27–29, 2023, Santa Cruz, CA, USA

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0176-4/23/06...\$15.00
<https://doi.org/10.1145/3589806.3600033>

However, the aforementioned initiatives primarily target a narrow audience of individuals that are already part of a research community. This paper argues that there are substantial benefits associated with educating a much wider audience of computer science students about reproducibility, and actively engaging them in advancing reproducibility in computer science and computational science. These include benefits to the students, as well as potential for positive impact on the broader research community.

In this paper, we describe some benefits of teaching about, with, and for reproducibility; different ways to include reproducibility content in the computer science curriculum at various levels; and challenges and considerations that educators may take into account when adding reproducibility content to their courses. We draw on published reports describing courses at the undergraduate and graduate level that have successfully included reproducibility instruction and experiences. These include courses across several sub-disciplines including machine learning and data science [9, 17], computer networks [16], and software engineering [10], as well as computational biology [6], neuroimaging [11], and economics [15]. We also draw on our own experience teaching with reproducibility content. The author has taught or co-taught several courses involving reproducibility content at NYU Tandon School of Engineering, including (since 2017) a course on communication network modeling and analysis for PhD students and advanced MS students, and (since 2020) an introductory course in machine learning for graduate students in electrical and computer engineering.

2 WHY TEACH REPRODUCIBILITY?

Although reproducibility as a core value in computer science research is not in any doubt, it is not necessarily obvious that including reproducibility content and experiences in computer science coursework (especially at the introductory level) is worth the investment of student and instructor effort, or limited classroom time. However, we argue that this investment can yield substantial benefits. In this section, we discuss the benefits of coursework-based reproducibility experience for students who are not and will not go on to engage in research; students who eventually go on to engage in research; students who are actively engaged in research; as well as educators and the broader research community.

2.1 Students not engaged in research

The majority of students in computer science classrooms go on to industry careers that do not involve research. However, as suggested in [10], practical skills related to reproducibility can be considered preparation for industry careers where students will be asked to produce work with consideration to long-term sustainability and maintenance. Similarly, in [15], the authors argue that students who

participate in reproducibility experiences gain expertise in key dimensions including “data acumen,” “data management and curation,” and “workflow and reproducibility,” that will be of value in future careers. In [12], which describes reproducibility content addressed toward students enrolled in a professional graduate degree program, the authors even motivate the reproducibility content with case studies of consequential real-world failures due to irreproducibility.

Furthermore, for many students, the first exposure to research in the fields in which they will make their future careers is in the classroom. For students who do not go on to engage in research themselves, it is nevertheless useful to demystify the research process and help students understand how research is conducted in their area of computer science. Several of the approaches in Section 3 offer the experience of “peeking behind the curtain,” identifying the nuts and bolts of a research experiment, and critically thinking about a research experiment and its conclusions - all of which are valuable skills for professionals who are not active researchers.

2.2 Future and current research students

For students who are already enrolled in a research-focused program of study, or who will go on to such study in the future, there is even more obvious value in instruction that helps them understand how research is conducted in their field and develop a critical eye for evaluating research results. For example, in [9], which describes the use of a reproducibility assignment in a course that is part of a research-focused MSc program, feedback from students indicated that it was “eye-opening” to see how research was performed by reproducing a result, that students had come to “appreciate the critical view” of research that they gained as part of the course, and that the experience “gave good insights into the trustworthiness of research papers.” Similarly, in [17], students reported that “doing a reproduction changed how I view the scientific process” and “made me more critical of results in scientific papers.”

Students who have had instruction and experience with reproducibility as part of their coursework, who later engage in research, may also become “reproducibility ambassadors” (as illustrated in Fig. 1). These students can use best practices for reproducibility in their own research, and also promote them to their future colleagues and advisors, multiplying the impact of the original instruction.

Finally, in our own experience, as well as in reports from other educators [16], a coursework-based reproducibility experience can help launch a student’s own research efforts. Most students in our network modeling and analysis course at NYU select papers to reproduce that are aligned with existing research interests. A number of these students then build on their reproductions in meaningful, creative, ways, eventually publishing them as novel research. Anecdotally, our students report that the experience of reproducing a research result teaches necessary skills and makes them feel more confident in their ability to engage in original research.

2.3 Broader research community

Besides the benefit to direct participants, we believe that students who engage in reproducibility experiences as part of their coursework impact their entire research community, as illustrated in Fig. 1.

Students who produce and publish new reproducibility artifacts as part of their coursework provide an obvious service to the field.

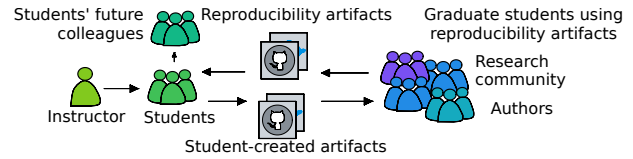


Figure 1: Incorporating reproducibility content in the classroom has the potential for big impact. Student contribute artifacts and insights back to original authors and the broader research community. Students can also directly impact their future colleagues as “reproducibility ambassadors.”

For example, in our networking course and in [16], students publish their reproduction and associated artifacts as a blog post, and are occasionally contacted later by researchers who use these artifacts. In [9], where many students submitted reports to the Machine Learning Reproducibility Challenge, the work was well received by the outside reviewers. In [17], where students were asked to contribute their reproducibility artifacts to an online repository, researchers in the field who were surveyed about this effort said that “an online reproduction attempt adds value to the original paper,” and that “I would want others to reproduce my own work.”

Students can also benefit the research community by playing a formal role in the artifact review process. In [15], undergraduate students carry out systematic computational reproducibility checks as part of the publication workflow of major economics journals.

More broadly, we believe that to incentivize the effort associated with producing reproducibility artifacts and improve their perceived value, they should be part of an ecosystem that includes many producers and consumers of reproducibility artifacts, who engage with these artifacts on an *ongoing* basis. With badging and reproducibility checks as part of the pre-publication workflow, a static reproducibility artifact is produced once by the authors, then consumed once by an artifact evaluation committee or equivalent. After this, the conversation around the reproducibility of the result ends in most cases (except for rare cases in which someone publishes a subsequent reproduction or replication).

However, reproducibility is neither as absolute nor as static as this process suggests. Reproducing a result in slightly different settings, using different implementations, or at different times can yield new insight about the result and the conditions under which the main claims remain valid. When students are tasked with either producing or consuming reproducibility artifacts as part of their coursework, they extend the conversation around the reproducibility of the result, both in time and in scope. This creates opportunities to uncover new insights and share them with the research community by having students publish a blog post, communicate with the original authors, and/or submit new reproducibility artifacts to a discipline-specific or infrastructure-connected repository.

3 HOW TO TEACH REPRODUCIBILITY?

Reproducibility content and experiences can be incorporated into undergraduate and graduate coursework in a variety of ways, separately or in combination. In this section, we describe some of these on a spectrum from “most accessible to beginners in the field” to “requires most discipline-specific expertise and capabilities.”

3.1 Include explicit instruction about reproducibility practices

Perhaps the most obvious way to include reproducibility content in the curriculum is to add explicit content about reproducibility, either as a standalone course on reproducibility, in a course on a closely related topic (e.g. scientific computation), or by describing discipline-specific standards for reproducibility in a discipline-specific course. For example, in [10], the authors suggest that reproducibility engineering be included as a part of the regular undergraduate software engineering curriculum. They propose a syllabus for such a course, which they implement in an undergraduate course offered at two universities. Similarly, in [12], the authors share their experiences teaching courses at the University of British Columbia that include explicit instruction on reproducible data science tools and workflows. As an extension of the approach described above, the authors of [11] include explicit instruction about reproducibility in a course on neuroimaging at UC Berkeley, and then asked students to complete a substantial open-ended project related to neuroimaging in which they apply these best practices for reproducibility.

In fact, many courses in computer science already incorporate instruction and expectations that are related to reproducible research practices (for example: use of version control, or careful documentation of experiment environments and results). Explicitly identifying these as best practices for *research* can potentially shape students' expectations and attitudes about reproducible research, with minimal additional effort required on the instructor's part.

3.2 Ask students to replay an already- packaged reproducibility artifact

For more experiential learning, students may be tasked with executing an already-packaged reproducibility artifact of a published result. This is an approach described in [5], for a networking course. For introductory courses, students may be asked to reproduce a foundational result that motivates the concepts emphasized in the course; for advanced courses, students may be asked to reproduce a recent impactful result that exposes students to state-of-the-art work in the field. This is especially relevant to courses that already engage students in hands-on lab exercises, where students may then engage with research artifacts in the same familiar settings.

This type of assignment may also show students how a field has advanced over time, by extending a foundational result. For example, in our introductory machine learning class, we ask students to replay an artifact [7] that reproduces a classic result from 1989, and then improves its performance by incrementally incorporating modern techniques that were discussed in our lectures.

Engaging with an already-packaged artifact can also help students learn about research methodology in their field. For example, in our course on network modeling and analysis at NYU, we ask students to execute a reproducibility artifact for a classic analytical result in queueing theory, as a simulation and as a testbed experiment. Through this assignment, students learn to recognize common pitfalls associated with these experimental methodologies.

As an addition to the above, students may be invited to extend a packaged result (for example: evaluate it under new experiment settings, add experiments to validate the proposed mechanism behind the result, or develop a novel extension of the result). An extension

that is especially valuable or insightful may be packaged by the student (potentially with the assistance of course staff) and shared as a new artifact that is descended from the original.

3.3 Ask students to create a new reproduction of a published result

For more advanced students, a course project may require students to create a new reproduction or replication of a published work. Experiences with this have been reported in a variety of sub-disciplines in computer science and computational science, including: an advanced networking course at Stanford University [16], a deep learning course at Delft University of Technology (TU Delft) [17], a course in fairness, accountability, confidentiality, and transparency in artificial intelligence (FACT-AI) at the University of Amsterdam [9], and an introductory R programming course for computational biology at Thomas Jefferson University [6]. We have similarly used this approach in graduate level networking and machine learning courses at NYU. Students may also be asked to review other students' work in a process similar to artifact evaluation.

While this type of course project may seem intimidating for anything but the most advanced courses, we have used this type of assignment in an introductory machine learning class by permitting students to participate at the level they are comfortable with. For example, students may choose to validate one claim on a small dataset, using author-provided code or even a pre-trained model. In [17], the authors similarly report that they chose to permit "light-weight reproduction" work alongside more extensive efforts .

4 KEY CONSIDERATIONS AND CHALLENGES

By building on the experiences of educators who have already experimented with teaching reproducibility, we can more easily reproduce their success. In this section, we summarize and highlight some "lessons learned" from the literature and our own experiences, and identify challenges and important considerations for introducing reproducibility in the classroom.

4.1 Scaffolding the learning process

In our own experiences and in the literature, a key enabler of success in reproducibility assignments was the materials used to scaffold the learning process. The instructors of the computational biology course at Thomas Jefferson University [6] highlighted the conventional homework assignments used throughout their course as equipping students with the skills to engage with and reproduce research results. Similarly, the neuroimaging course at UC Berkeley [11] used homework and lab sessions to prepare students for a more challenging project involving best practices for reproducibility. The FACT-AI course at University of Amsterdam [9] included several weeks of subject-specific instruction and also discipline-specific instruction about research methods and critically reading papers in the field. We follow a similar approach in our network modeling and analysis course, where students first learn about research methodologies and experiment design in this field before they are tasked with reproducing a published result. Unfortunately, developing these materials can require a substantial amount of effort on the part of the instructor. We hope that as more instructors

begin to use reproducibility in the classroom, they will create and share materials and other resources, easing this burden for others.

4.2 Selecting results to reproduce

For courses that ask students to create a new reproduction or replication of a published work, identifying suitable publications can be a major challenge. Students typically lack the expertise required to find papers that are feasible to reproduce within the scope of a course project, therefore the burden typically falls on course staff to prepare lists of papers for students to choose from. For example, in the FACT-AI course at the University of Amsterdam [9], the course staff selected 10-15 on-topic papers in advance where the dataset in the paper is available, the experiments can run on a single GPU, and the scope is appropriate for a group of 3-4 graduate students to re-implement the work in one month of full time work. The deep learning course at TU Delft [17] prepared a list of papers for which data is available, there is a clear computational result (table or figure) to reproduce, and the computational demands were reasonable. Students were also allowed to propose a paper not on the list provided by course staff. In the networking course at Stanford [16], course staff provided a list of suggested conferences and papers. The computational biology course at Thomas Jefferson University [6] took a slightly different approach, asking students to propose 6 papers with available data whose figures they would like to reproduce, and the course staff selected one of the 6 after evaluating the difficulty and reproducibility of the article's figures.

Instructors in some sub-disciplines of computer science may find it much easier to curate lists of reproducible results than in other sub-disciplines. We follow a similar approach in our own courses, where we provide a list of papers and students may propose others that we then evaluate for suitability. We found that in machine learning, which has a strong culture of reproducibility, we were able to identify many papers which had either code or pre-trained models available. In networking, this was much more difficult.

ACM badges have the potential to help with identification of suitable papers. However, "available artifacts" does not necessarily mean "useful for a replication assignment" - for example, an artifact may be data from which a figure is drawn, which would not help a student replicate the computational experiment that generated the data. Similarly, "artifacts evaluated - functional" at publication time does not always mean that artifacts are useful at the present time. In machine learning replication assignments especially, we have found that underlying frameworks and libraries change so quickly that even a few years later, replicating a result using author code can involve a substantial amount of effort to either establish the "old" environment or port code to a modern environment.

4.3 Providing support from course staff

Depending on the scope of the assignment, reproducibility experiences in the classroom can require a lot of support from course staff. For example, in the FACT-AI course at the University of Amsterdam [9], teams had either a one-hour tutorial including other teams or a private 30 minute team meeting with a teaching assistant every week. The networking course at Stanford [16] similarly had teaching assistant meet with student teams every week. Both of these courses also required students to submit an intermediate report

partway through the project, and provided feedback on this report. We have supported students engaged in reproducibility experiences with an asynchronous model (question-and-answer forum), which can be much less burdensome for course staff.

4.4 Providing access to infrastructure

For students to engage in meaningful reproducibility experiences, they will in many cases need access to research infrastructure on which to carry out experiments. Educators should consider the computational resources or infrastructure necessary to reproduce research in their field, and how they will make the necessary resources available to students. For example, the FACT-AI course at the University of Amsterdam [9] limited the scope of their reproducibility assignment to papers with experiments that can run on a single GPU, and provided this compute resource to students. The networking course at Stanford [16] provided students with compute resources on Amazon Web Service (AWS) Elastic Compute Cloud (EC2) for network emulation and simulation experiments.

In our own experience, dedicated research infrastructure including cloud computing testbeds like CloudLab [4] and Chameleon [8], testbeds for networking and distributed systems like GENI [2] and FABRIC [1], and wireless testbeds such as ORBIT [13]/COSMOS [14] and POWDER [3], have been invaluable for teaching reproducibility - we have used all of these in our own teaching. Experience with research platforms is helpful to students who may engage in research in the future. Also, for students who will package new reproducibility artifacts, making them available on an open research platform increases their reach. Some of these platforms have features explicitly tailored for reproducibility, such as the idea of a shared profile on CloudLab [4] or the Trovi portal on Chameleon [8].

4.5 Creating opportunities to contribute

Especially for courses where students produce new reproducibility artifacts, students were highly motivated to contribute their work back to the broader research community. Many educators chose to capitalize on this by deliberately providing opportunities for students to share their work as a blog post [16], by contributing to a curated repository of reproducibility artifacts [9], or both [17]. The authors of [9], for example, express a belief that students were highly motivated by the opportunity to create concrete resources and share them with the research community. Our experience has been similar; we highlight excellent student work from our reproducibility assignments on our research blog, and also help students find opportunities to share their replication or reproduction in subject-specific conferences and workshops.

5 CONCLUSIONS

This paper aims to encourage discussion around the inclusion of more reproducibility-related instruction and content in computer science classrooms. We hope this will spark new ideas, promote the creation of shared pedagogical resources, and encourage further activity and discussion around this topic.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2226408.

REFERENCES

- [1] Ilya Baldin, Anita Nikolich, James Griffioen, Indermohan Inder S Monga, Kuang-Ching Wang, Tom Lehman, and Paul Ruth. 2019. Fabric: A national-scale programmable experimental network infrastructure. *IEEE Internet Computing* 23, 6 (2019), 38–47.
- [2] Mark Berman, Jeffrey S Chase, Lawrence Landweber, Akihiro Nakao, Max Ott, Dipankar Raychaudhuri, Robert Ricci, and Ivan Seskar. 2014. GENI: A federated testbed for innovative network experiments. *Computer Networks* 61 (2014), 5–23.
- [3] Joe Breen, Andrew Buffmire, Jonathon Duerig, Kevin Dutt, Eric Eide, Mike Hibler, David Johnson, Sneha Kumar Kasera, Earl Lewis, Dustin Maas, et al. 2020. POWDER: Platform for open wireless data-driven experimental research. In *Proceedings of the 14th International Workshop on Wireless Network Testbeds, Experimental evaluation & Characterization*. Association for Computing Machinery, New York, NY, USA, 17–24.
- [4] Dmitry Duplyakin, Robert Ricci, Aleksander Maricq, Gary Wong, Jonathon Duerig, Eric Eide, Leigh Stoller, Mike Hibler, David Johnson, Kirk Webb, et al. 2019. The Design and Operation of CloudLab. In *USENIX Annual Technical Conference*. USENIX Association, USA, 1–14.
- [5] Yayu Gao, Chengwei Zhang, Xiaojun Hei, and Guohui Zhong. 2019. Learning networking by reproducing research results in an ns-3 simulation networking laboratory course. In *2019 IEEE International Conference on Engineering, Technology and Education (TALE)*. IEEE, 1–6.
- [6] Nestoras Karathanasis, Daniel Hwang, Vibol Heng, Rimal Abhimannu, Phillip Slogoff-Sevilla, Gina Buchel, Victoria Frisbie, Peiyao Li, Dafni Kryoneriti, and Isidore Rigoutsos. 2022. Reproducibility efforts as a teaching tool: A pilot study. *PLOS Computational Biology* 18, 11 (11 2022), 1–11. <https://doi.org/10.1371/journal.pcbi.1010615>
- [7] Andrej Karpathy. 2022. Deep Neural Nets: 33 years ago and 33 years from now (Invited Post). In *ICLR Blog Track*. <https://iclr-blog-track.github.io/2022/03/26/lecun1989/>
- [8] Kate Keahey, Jason Anderson, Zhuo Zhen, Pierre Riteau, Paul Ruth, Dan Stanzione, Mert Cevik, Jacob Colleran, Haryadi S. Gunawi, Cody Hammock, Joe Mambretti, Alexander Barnes, François Halbach, Alex Rocha, and Joe Stubbs. 2020. Lessons Learned from the Chameleon Testbed. In *Proceedings of the 2020 USENIX Annual Technical Conference (USENIX ATC '20)*. USENIX Association.
- [9] Ana Lucic, Maurits Bleeker, Sami Jullien, Samarth Bhargav, and Maarten de Rijke. 2022. Reproducibility as a Mechanism for Teaching Fairness, Accountability, Confidentiality, and Transparency in Artificial Intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 11 (6 2022), 12792–12800. <https://doi.org/10.1609/aaai.v36i11.21558>
- [10] Wolfgang Mauerer, Stefan Klessinger, and Stefanie Scherzinger. 2023. Beyond the Badge: Reproducibility Engineering as a Lifetime Skill. In *Proceedings of the 4th International Workshop on Software Engineering Education for the Next Generation (Pittsburgh, Pennsylvania) (SEENG '22)*. Association for Computing Machinery, New York, NY, USA, 1–4. <https://doi.org/10.1145/3528231.3528359>
- [11] K Jarrod Millman, Matthew Brett, Ross Barnowski, and Jean-Baptiste Poline. 2018. Teaching computational reproducibility for neuroimaging. *Frontiers in Neuroscience* 12 (2018), 727.
- [12] Joel Ostblom and Tiffany Timbers. 2022. Opinionated Practices for Teaching Reproducibility: Motivation, Guided Instruction and Practice. *Journal of Statistics and Data Science Education* 30, 3 (2022), 241–250. <https://doi.org/10.1080/26939169.2022.2074922>
- [13] Dipankar Raychaudhuri, Ivan Seskar, Max Ott, Sachin Ganu, Kishore Ramachandran, Haris Kremono, Robert Siracusa, Hang Liu, and Manpreet Singh. 2005. Overview of the ORBIT radio grid testbed for evaluation of next-generation wireless network protocols. In *IEEE Wireless Communications and Networking Conference, 2005*, Vol. 3. IEEE, 1664–1669.
- [14] Dipankar Raychaudhuri, Ivan Seskar, Gil Zussman, Thanasis Korakis, Dan Kilper, Tingjun Chen, Jakub Kolodziejewski, Michael Sherman, Zoran Kostic, Xiaoxiong Gu, et al. 2020. Challenge: COSMOS: A city-scale programmable testbed for experimentation with advanced wireless. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. Association for Computing Machinery, New York, NY, USA, 1–13.
- [15] Lars Vilhuber, 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 (2022), 274–281. <https://doi.org/10.1080/26939169.2022.2074582>
- [16] Lisa Yan and Nick McKeown. 2017. Learning networking by reproducing research results. *ACM SIGCOMM Computer Communication Review* 47, 2 (2017), 19–26.
- [17] Burak Yildiz, Hayley Hung, Jesse H Krijthe, Cynthia CS Liem, Marco Loog, Gosia Migut, Frans A Oliehoek, Annibale Panichella, Przemyslaw Pawelczak, Stjepan Picek, et al. 2021. ReproducedPapers.org: Openly teaching and structuring machine learning reproducibility. In *Reproducible Research in Pattern Recognition: Third International Workshop, RRPR 2021, Virtual Event, January 11, 2021, Revised Selected Papers*. Springer International Publishing, 3–11.