XSEDE EMPOWER: Engaging Undergraduates in the Work of Advanced Digital Services and Resources

Aaron Weeden
Shodor Education Foundation
Durham, NC
aweeden@shodor.org

ABSTRACT

To address the need for a diverse and capable workforce in advanced digital services and resources, the Shodor Education Foundation has been coordinating an undergraduate student program for the Extreme Science and Engineering Discovery Environment (XSEDE). The name of the program is EMPOWER (Expert Mentoring Producing Opportunities for Work, Education, and Research). The goal of the program is to engage a diverse group of undergraduate students in the work of XSEDE, matching them with faculty and staff mentors who have projects that make use of XSEDE services and resources or that otherwise prepare students to use these types of services and resources. Mentors have coordinated projects in computational science and engineering research in many fields of study as well as systems and user support. Students work for a semester, quarter, or summer at a time and can participate for up to a year supported by stipends from the program, at different levels depending on experience. The program has run for 11 iterations from summer 2017 through fall 2020. The 111 total student participants have been 28% female and 31% underrepresented minority, and they have been selected from a pool of 272 total student applicants who have been 31% female and 30% underrepresented minority. We are pleased that the selection process does not favor against women and minorities but would also like to see these proportions increase. At least one fourth of the students have presented their work in articles or at conferences, and multiple credit the program with moving them towards graduate study or otherwise advancing them in their careers.

Keywords

Undergraduate student programs, Advanced digital services and resources, HPC, Computational science, Data science, Internships

1. INTRODUCTION

To address the need for a diverse and capable workforce in advanced digital services and resources, the Shodor Education Foundation [8] has been coordinating an undergraduate student program since summer 2017 for the Extreme Science and Engineering Discovery Environment (XSEDE) [10]. The name of the program is EMPOWER (Expert Mentoring Producing Opportunities for Work, Education, and Research). The goal of the program is to engage a diverse group of undergraduate students in

the work of XSEDE, matching them with faculty and staff mentors who have projects that make use of XSEDE services and resources or that otherwise prepare students to use these types of services and resources. Mentors have coordinated projects in computational science and engineering research in many fields of study as well as systems and user support. Students work for a semester, quarter, or summer at a time and can participate for up to a year supported by stipends from the program. Students participate at one of three different levels depending on their existing experience: Learners are trained on new skills and knowledge, Apprentices apply their skills and knowledge to supervised tasks, and Interns work more independently.

2. METHODS
2.1 Recruiting Mentors and Students

We recruit mentors and students by promoting the program in XSEDE newsletters [12]; engaging with the Campus Champions community [3], who participate themselves and/or spread the word on their own campuses; and with help from the XSEDE Broadening Participation team [1], who engage with historically underrepresented faculty and students through conference and campus visits.

2.2 Receiving and Reviewing Applications

Shodor maintains an in-house application website [9] and database, originally developed to coordinate workshop registrations for the National Computational Science Institute [7]. Student and mentor application forms for EMPOWER have been developed based on those of the Blue Waters Student Internship Program [4]. The mentor form requests the mentor's affiliation/role with XSEDE, the project title and summary, student job description, use of XSEDE resources, contribution to the community, start and end dates, location, participation level, training plan, number of students the mentor can support, student names (if already identified), and additional student prerequisites and qualifications. The student application form requests GPA; subject areas studied; subject interests; relevant courses and grades; relevant work and internship experiences; career goals; interests in contributing to XSEDE; and experiences with mathematics, computing, application software, programming languages, Unix/Linux, parallel computing, visualization, data science, and machine learning.

When students apply, they can optionally indicate a gender (Male or Female) and/or an ethnicity (African-American, American Indian or Alaskan, Asian or Pacific Islander, Hispanic, Caucasian, or Middle Eastern). For our analysis, we consider the following ethnicities to be underrepresented minorities: African-American, American Indian or Alaskan, and Hispanic.

As applications come in, the program coordinator reviews the submissions, looking for issues that may need to be addressed, such as start and end dates that do not line up with the dates of the

program, or mentors indicating preferred students who have not themselves submitted applications (or vice versa for students and their preferred mentors). The coordinator corresponds with mentors to identify which student applications should be matched with which mentor applications during the review process.

With the goal of avoiding biases during the review process, an anonymous review is conducted. The coordinator prepares a Portable Document Format (PDF) file for each mentor/student application pair in which names, pronouns, genders, ethnicities, institutions, locations, and URLs are removed. The process of preparing these PDF files is partially automated and partially manual. A script written in the PHP programming language automatically searches for gender pronouns and replaces them with an indicator that a gender pronoun has been removed. The script also allows for other search strings to be specified by the coordinator manually (i.e. those indicating names, pronouns, genders, ethnicities, institutions, locations, and URLs). These search strings are automatically replaced with indicators that the string has been removed. The coordinator reads each application looking for search strings to include and types them into a spreadsheet. The spreadsheet is set up to construct a URL that is used to run the PHP script. A separate Bash script requests each URL and downloads the resulting output into PDF files. The result is a collection of anonymized PDF files that are given to reviewers.

The coordinator recruits three reviewers for each PDF file from the community: primarily from the Campus Champions, the XSEDE Community Engagement and Enrichment staff, and former mentor participants of EMPOWER. Reviewers score each PDF using a Google Form that presents ten prompts on a 1–5 scale: 1) the project contributes to the work of XSEDE, 2) the proposed level of participation (Learner, Apprentice, or Intern) is appropriate, 3) the training plan is appropriate, 4) the project is suitably scoped given the start and end dates, 5) the mentor is likely to do a good job supporting the student, 6) the student is new to the XSEDE community, 7) the project is a good fit for the student's expressed interests and skill level, 8) the student is likely to do a good job in the project, 9) the student's participation will advance them in their career path as a user or facilitator of advanced digital services and resources, and 10) the match of mentor, student, and position should be selected for the program.

For a given PDF and reviewer, a weighted score is calculated based on the reviewer's scores for each prompt using the formula below.

$$4*(P10) + 3*(P1+P5+P8+P9) + 2*(P2+P3+P4+P7) + P6$$

A final score for each PDF is calculated using an iterative model that takes into account the weighted scores and reviewer leniencies, based on the Differential Model described in [5]. PDF files are ranked by this final score, reviewer comments and other considerations (such as previous participation) are taken into account, and final selections are made by the coordinator.

2.3 Training Students
The student participants of the summer 2017 and summer 2018 programs were invited to participate in the Petascale Institute [11], a two-week training event conducted by Shodor for the undergraduate students of the Blue Waters Student Internship Program. Students from EMPOWER who were interested and available attended this training, where they learned how to apply parallel and distributed computing concepts to computational simulation and modeling using the Blue Waters supercomputer [2] as the example architecture. Funding for the Petascale Institute ended in 2018.

Outside of the Petascale Institute, Shodor has so far not conducted training for the EMPOWER students. Instead, the mentor participants in the program provide training to the students whom they are mentoring. The modes of training have varied and include formal courses, informal lectures and one-on-one tutoring, and self-learning using online resources.

2.4 Reviewing Student Work
EMPOWER students complete monthly progress reports in which they describe accomplishments, issues, and what they plan to accomplish before the next report. In reading these reports, the EMPOWER coordinator looks for mentions of publications or presentations that have been produced or prepared by the students, as well as other highlights of significant accomplishments. These highlights are reported to XSEDE and the National Science Foundation as evidence of impact of the program. Students receive their stipend payments once they have completed all of their required monthly reports.

The EMPOWER coordinator also occasionally conducts optional small surveys of the students and mentors to obtain specific information, such as the number of hours per week that mentors are putting into mentoring their students or comments about the impact of the program on students' career paths.

3. RESULTS
There have been 11 iterations of the EMPOWER program from summer 2017 through fall 2020 (one each per fall, spring, and summer). The 111 total student participants have been 28% female and 31% underrepresented minority, and they have been selected from a pool of 272 total student applicants who have been 31% female and 30% underrepresented minority.

At least one fourth of the students have published articles or presented at conferences about their EMPOWER work. We have also heard anecdotally from students about the impact of the program on their career paths, including highlights such as being selected as a college's valedictorian, securing cooperative education, receiving offers from Research Experience for Undergraduate programs, deciding what to study in graduate school, impressing job recruiters, improving public speaking skills, leading and tutoring new student researchers, and networking and meeting new collaborators.

In an optional survey, mentors reported putting in an average of 2–10 hours of mentoring per week per student.

4. DISCUSSION
The program has had essentially equal proportions of female student applicants and female student participants, as well as essentially equal proportions of underrepresented minority student applicants and underrepresented minority student participants, which suggests the selection process does not favor against female or minority students. We would like to increase the proportions of female and minority applicants and participants. We will change the application form to use more standard and inclusive categories. For example, we will update the optional "gender" prompt in our account creation form to provide a "Non-binary" option as well as a "Prefer to Self-describe" option. We will also use the race and ethnicity categories from [6] as a starting point for updating our form's "ethnicity" prompt, as well as taking into account inclusivity research about the categories. We would also like to do a more thorough study of the impact of the program on student career paths and do other analyses of our data. The EMPOWER program will run for five more iterations through summer 2022.

6. REFERENCES

[1] Linda Akli. 2018. XSEDE: Tackling Diversity and Inclusion in Advanced Computing. *Computing in Science & Engineering* 20, 3 (May–Jun. 2018), 71–72. DOI: https://doi.org/10.1109/MCSE.2018.03202635.

[2] Brett Bode, Michelle Butler, Thom Dunning, Torsten Hoeer, William Kramer, William Gropp, and Hwu Wen-Mei. 2013. The Blue Waters Super-System for Super-Science. *Contemporary High Performance Computing: From Petascale toward Exascale*, 339–366. CRC Press.

[3] Marisa Brazil, Dana Brunson, Aaron Culich, Lizanne DeStefano, Douglas M. Jennewein, Tiffany Jolley, Timothy Middelkoop, Henry J. Neeman, Lorna I. Rivera, Jack A. Smith, and Julie A. Wernert. 2019. Campus Champions: Building and sustaining a thriving community of practice around research computing and data. In *PEARC '19: Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning)* 78 (July 2019), 1–7. DOI: https://doi.org/10.1145/3332186.3332200

[4] Patricia Jacobs, Phillip List, Mobeen Ludin, Aaron Weeden, and Robert M. Panoff. 2014. The Blue Waters Student Internship Program: Promoting Competence and Confidence for Next Generation Researchers in High-Performance Computing. *2014 Workshop on Education for High Performance Computing* (Nov. 2014), 49–55. DOI: https://doi.org/10.1109/EduHPC.2014.6

[5] Hady W. Lauw, Ee-Peng Lim, and Ke Wang. 2007. Summarizing Review Scores of "Unequal" Reviewers. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, April 26–28, 2007, Minneapolis, MN, 539–544. DOI: https://doi.org/10.1137/1.9781611972771.58

[6] United States Office of Management and Budget (OMB). 1997. *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. Notice 62 FR 58782. 9 pages. Document number 97-28653. Retrieved from https://www.federalregister.gov/documents/1997/10/30/97-28653/revisions-to-the-standards-for-the-classification-of-federal-data-on-race-and-ethnicity/.

[7] Shodor. 2020. The National Computational Science Institute. Retrieved from http://computationalscience.org/.

[8] Shodor. 2020. The Shodor Education Foundation, Inc. Retrieved from http://www.shodor.org/.

[9] Shodor. 2020. XSEDE EMPOWER Program. Retrieved from http://computationalscience.org/xsede-empower/.

[10] John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, and Nancy Wilkins-Diehr. 2014. XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* 16, 5 (Sept.–Oct. 2014), 62–74. DOI: https://doi.org/10.1109/MCSE.2014.80

[11] Aaron Weeden, Colleen Heinemann, Skylar Thompson, Cameron Foss, Mobeen Ludin, and Jennifer K. Houchins. 2019. The Blue Waters Petascale Institute: Longitudinal Impact and Assessment-Driven Development of an Intensive, Hands-on Curriculum for Teaching Applications in HPC. In *PEARC '19: Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning)* 85 (July 2019), 1–8. DOI: https://doi.org/10.1145/3332186.3337771

[12] XSEDE. 2020. Impact. Retrieved from https://www.xsede.org/impact/.