ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE
Paper ID #38209

Multidisciplinary Engagement of Diverse Students in Computer Science Education through Research Focused on Social Media COVID-19 Misinformation

David Brown

David Brown is a Ph.D. student at the University of North Carolina at Charlotte. A former GAANN Fellow, David has received mentorship in STEM education practices while teaching introductory programming courses to undergraduate students. As a 3-Minute Thesis finalist at UNC Charlotte, David has demonstrated commitment for STEM communication to broader audiences. His research is focused on the application of statistical correlation, machine learning, dendrograms, and networks to analyze and visualize biological data.

Erfan Al-hossami

Erfan Al-Hossami is a Ph.D. student at the University of North Carolina at Charlotte. Erfan has been mentored in teaching CS1 since 2016 and then in CS education research. His work mainly focuses on predictive learning analytics and code generation. His research interests include Machine Learning, NLP with a particular focus on task-oriented dialogue systems like Siri and generating code from natural language.

Zhuo Cheng

Zhuo Cheng studies at UNC Charlotte as a Ph.D. student. Zhuo's research interest lies in Natural Language Processing and machine learning. He has been working on various projects including computational propaganda, hate speech detection, legal documents analysis and conversational AI. Zhuo is also interested in the application of NLP in the industry.

Samira Shaikh

Samira Shaikh is an Associate Professor of Computer Science at the University of North Carolina at Charlotte. Her expertise is in Computational Sociolinguistics, Dialogue Systems and Natural Language Generation.

Daniel Janies

In 2012, Daniel Janies joined the University of North Carolina at Charlotte as The Carol Grotnes Belk Distinguished Professor of Bioinformatics and Genomics. Dr. Janies received a Bachelor of Sciences in Biology from the University of Michigan in 1988 and a Ph.D. in Zoology from the University of Florida in 1995. Dr. Janies was a tenured faculty member in the College of Medicine at the Ohio State University where he served as a national principal investigator in the Tree of Life program of the NSF. Dr. Janies recent awards include DoD sponsored work to understand the spread of pathogens. Dr. Janies has advised the White House, the Pentagon, and testified to both Houses of Congress.

Mesbah Uddin

Dr. Mesbah Uddin is the Director of North Carolina Motorsports and Automotive Research Center and a Professor of Mechanical Engineering at the University of North Carolina at Charlotte. He is currently serving as the Chair of the Society of Automotive Engineers (SAE) Road Vehicle Aerodynamics Forum Committee, a committee responsible for

developing and maintaining SAE standards, technical papers, and special publications related to road vehicle aerodynamics and wind noise performance and test techniques. He is a member of UNC Charlotte Military Affairs Committee. In the past, he served as a member of North Carolina Governor's Motorsports Advisory Council during 2012-2017. He is also a member of the AIAA Turbulence Model Benchmarking Working Group. In the past, he served as senior CFD analyst at the then Chrysler's in its NASCAR, NHRA, street and passenger car computational aerothermal development programs. At present, Dr.Uddin's group focuses on the improvement of the aerothermal predictive capabilities of virtual and physical systems using machine learning and reduced order methods. As of today, he has graduated 6 PhD and 24 masters students, and published over 90 peer reviewed journal articles and conference papers.

Multidisciplinary Engagement of Diverse Students in Computer Science Education through Research Focused on Social Media COVID-19 Misinformation

1 Abstract

The ongoing COVID-19 pandemic has disrupted vital elements of personal and public health, society, and education. Increasingly with the viral pandemic, misinformation on health and science issues has been disseminated online. We developed an undergraduate training program focused on producing and presenting research to combat the rampant spread of this misinformation. Online misinformation represents a complex, multidisciplinary problem. Consequently, recruitment of students to the program was not exclusive to those from Computer Science or Science, Technology, Engineering, and Math (STEM) educational backgrounds. Participants were actively recruited from fields such as Linguistics, Social and Political sciences. This data analytics outreach program aimed to train educationally and demographically diverse undergraduate students in computational techniques and presentation skills through guided research regarding the current burst of misinformation. Over ten weeks, participants were instructed in an online curriculum covering five milestones: Python programming, data processing, machine learning with natural language processing, visualization, and presentation. Subsequently, participants were engaged in Computer Science research analyzing a real-world data set gathered from Twitter™ [1] between January and June 2020. Participants were organized into teams to investigate subtopics within the broader subject of misinformation: 1) detecting social media bot accounts, 2) identifying propaganda with computational methods, and 3) studying the discourse surrounding science preprints (i.e., papers that have been posted to the Internet but have not been peer reviewed). The program culminated in an exposition where each team presented research results to program officers, senior faculty, deans, government officials, and industry experts. Here we present the program curriculum, metrics of educational effectiveness, and feedback collected from participants.

[1] TWITTER, TWEET, RETWEET and the Twitter Bird logo are trademarks of Twitter Inc. or its affiliates. This talk uses TWITTER data but is not affiliated with or by TWITTER.

2 Introduction

The COVID-19 pandemic recorded 5.4 million deaths globally by December 2021 [1]. Within the United States, there have been 54 million documented cases and 820,000 deaths in that period [2]. The effects of the pandemic have been disruptive for public health, society, the economy, and education [3, 4, 5]. Misinformation regarding the COVID-19 pandemic is rampant [6, 7] and undermines our ability to successfully recover [8, 9]. Coinciding with the rise of misinformation, there has been a deluge of scientific papers that have not yet been peer reviewed (preprints) and associated Twitter posts on COVID-19 and its causative agent the SARS-CoV-2 virus [10].

Misinformation about the pandemic emerged online and spread rapidly [11]. Recent analyses indicate that unfounded claims in health-related misinformation aimed to sow distrust and commit fraud provides challenging competition with information from evidence-based sources such as the World Health Organization (WHO) and the US Centers for Disease Control (CDC) [12]. Appropriate and effective health guidance became obstructed through messages driven by propaganda and fear on social media platforms [13].

The global misinformation pandemic accompanying the spread of COVID-19 represents an opportunity for research. Abundant textual data and accompanying metadata are available from social media platforms and could prove useful for training computer science students [14]. Computer science education and research are necessary to abate this ongoing global misinformation crisis [14]. Therefore, a real-world data set was obtained from Twitter between January and June 2020 to investigate the online discourse surrounding the coronavirus pandemic [15]. A ten-week research program was developed to address the dual challenges of identifying effective computer science researchers and educating remote students. Through team science efforts, diverse, multidisciplinary research teams were recruited to investigate potential online misinformation.

Other research programs described in the literature for undergraduates include interdisciplinary research for medicine, biology, and computer science [16, 17], teaching research skills during a summer program [18], online research programs [19], and alternative roadmaps for STEM doctoral candidates [20]. Our program set up follows in the footsteps of interdisciplinary research programs, however it is novel from prior works, focusing on project-based learning, applying AI to the topic of health misinformation with a diverse set of undergraduates students across many majors. The program was also entirely delivered online.

3 Program Roadmap

Here we report on a fully remote program to educate a diverse cadre of college undergraduates in computational research techniques. In this program, learning outcomes focused on increasing participant familiarity with data science research methods (see Table 1). Participants formulated individual research questions within the broader research goals of their assigned team.

In the initial weeks of the program, we reviewed quantitative skills in analyzing large-scale language data, introduced design principles for data science analyses, and applied these new skills. Through guided collaboration sessions with graduate mentors, the participants gained

Table 1: Program Learning Outcomes

Objective	Description
Knowledge	Gain knowledge and skills to analyze large-scale language data.
Experimentation	Understand principles of design for social science experiments.
Hypothesis	Design theoretical experiments for making causal inferences.
Independence	Develop an independent research program.
Coding	Apply coding skills in execution of research.
Mentorship	Build mentor/mentee relationships with graduate collaborators.

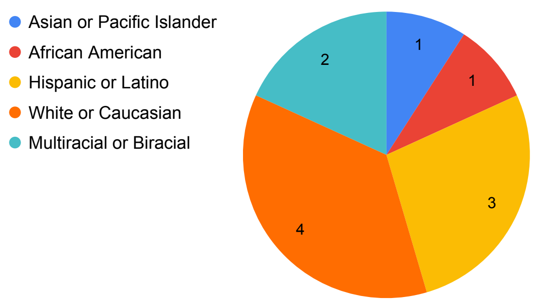
hands-on experience developing an independent research program in preparation for research fellowships, graduate school, and STEM research careers.

4 Program Recruitment

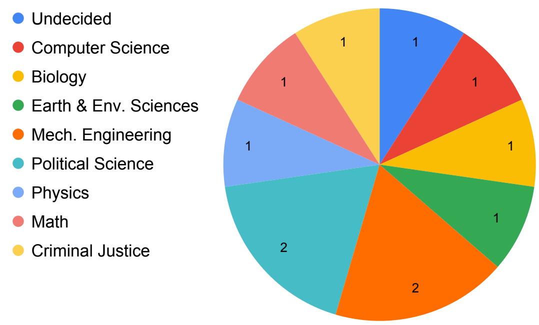
Misinformation is an emergent electronic threat to national security, personal, and public health. Online misinformation regarding COVID-19, and its causative agent the SARS-CoV-2 virus, has damaged our ability to recover from the pandemic. Some analyses suggest that false health-related misinformation provides challenging online social media competition for information from authorities in government, medicine, and research [12]. Program recruitment was targeted towards participants who identified a need in their academic or professional careers that corresponded with a need in society. This program provided an opportunity for undergraduates without a traditional background in computer science to engage in substantive research projects and improve their marketability.

Funding was obtained from the Office of Naval Research (ONR) to develop a training program for undergraduate Reserve Officer Training Corps (ROTC) cadets and US nationals in hands-on research activities in the general area of computational propaganda, in which our primary focus had been on the detection of online misinformation. The use of a paid summer research internship incentivized undergraduate students to apply for the available positions. Participants were compensated at a rate of $18 per hour for 15 hours of work per week during the ten week internship. Recruitment efforts were intentionally diverse. First, ROTC leaders promoted the program to undergraduate cadets at the University of North Carolina at Charlotte (UNC Charlotte). Next, the program was advertised to students in the College of Liberal Arts and Sciences, the College of Computing and Informatics, and The William States Lee College of Engineering through a website and distributed flyers. These efforts resulted in the reception of twenty-one applications.

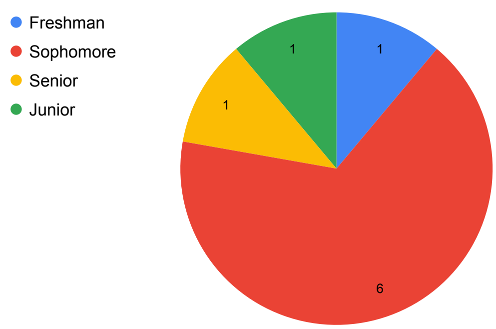
Recruitment focused on students whose immediate career goals aligned with national interests regarding complex electronic threats like information subversion or online misinformation. Secondarily, program criteria emphasized selecting high-performing, motivated individuals from a diverse range of academic backgrounds. Of utmost importance was the consistent



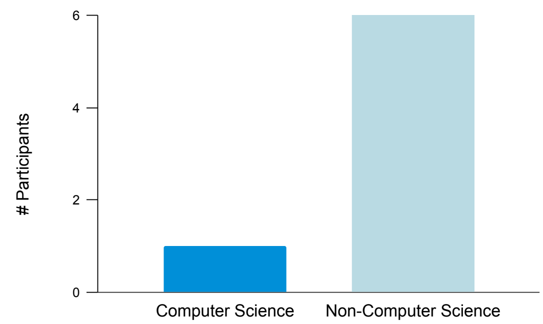
(a) Reported Race



(b) Reported Area of Study



(c) Reported Class Level



(d) Reported Computer Science Participation

Figure 1: Participant reported demographics by race, area of study, class level, and computer science participation

demonstration of academic achievement, regardless of discipline. Finally, individuals who appeared to prioritize research rather than training were chosen. From the initial group of twenty-one applicants, nine participants were ultimately accepted. The accepted participants' demographics are displayed in Figure 1.

An interest survey was sent to participants to inform the apportionment of individuals between the research teams and team roles. Participants who indicated a life sciences background were assigned to the "Science Misinformation" team, while those with a linguistics or political science background were assigned to the "Computational Propaganda" track. Similarly, individuals with some exposure to computer science would be given the "Data Integrity Officer" role, while a participant expressing a high degree of organization would serve as the "Operations Officer." Further descriptions of research teams and team roles can be found in Table 2 and Table 3 respectively.

5 Program Activities: Modules and Curriculum

The program was conducted during the summer of 2021 in accordance with state and local safety guidelines during the COVID-19 pandemic. Meetings were held using the Zoom™ application provided by the institution. All Zoom instructional sessions were recorded and made available for participants' asynchronous review in a shared Google™ Drive. At the close of each week, verbal

Table 2: Research Teams and Goals

Team	**Research Goal**
Bot Detection	Identify computer programs posing as human users.
Computational Propaganda	Study propaganda used by governments on social media.
Science Misinformation	Examine claims citing scientific research on coronavirus.

Table 3: Assigned Roles within Research Teams

Role	**Description**
Data Integrity Officer	Monitors data quality and analysis outputs.
Operations Officer	Documents the work done during the project.
Squad Lead	Organizes communication within team and between teams.

feedback was provided to each team by the associated Principal Investigator (PI) using the Zoom breakout rooms (one PI per team). The ten weeks were split into the Common, Research, and Presentation Modules as shown in Figure 2.

5.1 Common Module (CM)

The CM was designed to provide participants with a working knowledge of computer science and programming skills, irrespective of their educational background. At the close of the CM, expected learning outcomes included:

1. Participants will be able to use relevant programming software (IDEs),
2. Participants will be able to program in the Python programming language,
3. Participants will be able to manipulate and process tabular data using Python,
4. Participants will be able to apply NLP and AI techniques to the data and visualize outcomes,
5. Applying Learned Techniques to Investigate Misinformation in various Project Tracks.

In week 1, we focused on introducing general programming concepts using the Python 3 language. In week 2, we focused on the further development of Python programming skills by implementing the open-source machine learning (ML) library scikit-learn [21]. In week 3, we practiced these new skills with the introduction of natural language processing (NLP).

Each day of the CM was split between lectures, guided demonstrations, and practice time. Participants remained in the main Zoom “room", with as needed use of “Breakout Room" functionality to allow for smaller groups or individual questions. Jupyter Notebooks® and Google Colab™ were used to present lesson materials. The original starter codes were made available to the participants to function as workbooks and practice materials. Lectures and exercises were provided via starter code to increase participant familiarity with Python programming through experiential learning. For additional practice, participants were provided with a list of relevant courses and modules available online from DataCamp™ [22], for which an academic organization subscription had been obtained. Much emphasis was placed on developing fundamental skills

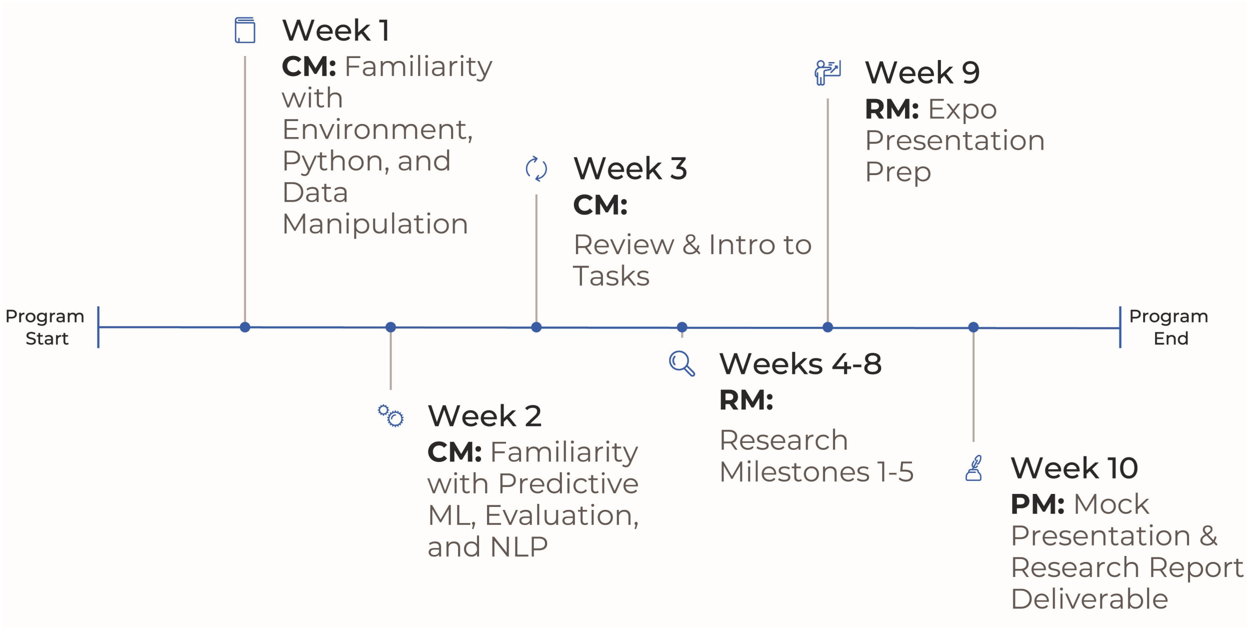


Figure 2: Program timeline including Common Module (CM), Research Module (RM), and Presentation Module (PM).

during the CM to ground the investigative process during the Research Modules.

5.2 Research Modules (RM)

Many common training data sets include the King County real estate set [23], the Iris data set [24], the Post Office data set [25], and movie reviews [26]. These data sets might appear unfamiliar for students outside of computer science, leading to difficulties learning and retaining new material. Internet users are often aware of Twitter and the basic format of a Twitter post or Tweet. Therefore, a Twitter data set [15] was supplied to the participants in the belief that relevant, guided research on familiar data is important for developing novel skills and practical abilities.

5.2.1 Data Set

In order to investigate online misinformation about the pandemic, Tweet identifiers from January to June of 2020 had been collected and filtered for relevance to COVID-19 [15]. The collection of Tweet identifiers was programmatically matched with the all associated data and metadata using the Twitter API in a process called “rehydration". This rehydrated data set was used by all three research tracks, using investigation-specific filters to limit the data size. For example, to arrive at the data subset for Computational Propaganda, the data was filtered for a short list of Twitter accounts associated with international governments [27]. These accounts represented government interactions with the general public on the topic of COVID-19. The Computational Propaganda subset was explored for differences in the online engagement of various governments regarding the pandemic. A different filtered subset of the data was used for the Bot Detection and Science Misinformation teams. These teams worked with Tweets linking to bioRxiv and medRxiv preprint servers. Preprint servers host early-release, complete, but non-peer reviewed drafts of scientific manuscripts. For Bot Detection, these preprint-associated Tweets were used to compare the activity of non-human users with human users. Science Misinformation investigated these same data for the potential misinformation about COVID-19 that was proliferated on Twitter using scientific literature.

5.2.2 Research Teams

Participants were split between the available research projects of Bot Detection, Computational Propaganda, and Science Misinformation (Table 2). Each team was subdivided into three roles: Squad Leader, Operations Officer, and Data Integrity Officer (Table 3). Both team and role assignments had been made by program leadership based on participant responses to the earlier interest survey following recruitment.

Table 4: Team Milestones during the Research Module.

Milestone	**Bot Detection**	**Computational Propaganda**	**Science Misinformation**
Data Processing	Extract features from preprint data set.	Pre-process Tweet content for textual data.	Clean and merge Twitter and preprint article data.
Application	Use starter code to train an ML model to classify accounts as bots or not.	Use starter code (BERT) to detect propaganda.	Use starter code to estimate sentiment and find user connections.
Data Analysis	Compare Tweet content and identify relative "velocity" of information (bot vs user).	Classify Tweets based on propaganda content.	Build networks for information spread by user relationships.
Data Visualization	Illustrate areas of dissimilarity between bots and users.	Compare differences in language and propaganda techniques by source.	Develop network graphs for user and preprint connections.
Presentation	Practice techniques for public speaking.	Practice techniques for public speaking.	Practice techniques for public speaking.

Standardized milestones, with one objective per week, were employed for the RM (see Table 4). The five milestones focused on data processing, applications of AI and NLP, analysis, visualization, and presentation. These milestone checkpoints served participants by subdividing the research process into digestible portions [28]. With the use of milestones, participants familiarized themselves with the data through exploratory data and statistical analysis in week 4 to build hypotheses for questions of interest. In week 5, participants applied the ML and NLP techniques they learned during the CM to their research questions. In week 6, we encouraged participants to further explore results and develop focused analyses. In week 7, we began condensing results into explanatory, useful, and relevant images, charts, and graphs. In week 8, we included an introduction to basic presentation design elements for a webpage, a poster, or an online presentation at a meeting. Each RM week ended with a review of each team and participant's progress to date, with feedback from the PIs to the entire cohort of participants and PI-to-team conferences.

5.3 Presentation Module (PM)

The remaining two weeks of the program were designated for preparation and presentation at the exposition. During week 9, participants were coached in visualization and verbal presentation in Zoom with revision during multiple practice sessions led by the PIs. Explicit instruction was devoted to traditionally accepted aspects of a STEM research program: the development of research questions, the application of methods, the observed results, and the informed conclusion. These sessions were used to gauge the final progress of the participants while performing the ultimate step of a research investigation – the exhibition of a clear set of objectives, results, and storyline.

During the PM, participants wrote individual research papers and verbally presented research results to a professional review committee exposition with a wide variety of expertise. Exposition committee members include U.S. military personnel concerned with emergent threats, university administrators and faculty, industry partners, government officials, and other subject matter experts. This diverse, expert audience gave the participants practice in communicating research and receiving feedback from decision makers, which are important skills for future careers in STEM and beyond.

6 Evaluation Approach

Two approaches were used to evaluate the program. First, participants were surveyed at the beginning and end of the program. These pre- and post-surveys were identical, collecting both scaled scores and open-ended responses. Second, anonymous qualitative reflections were collected weekly from the participants to gauge specific areas of strength, challenge, and improvement.

6.1 Pre- and Post-surveys

A questionnaire was developed to measure participant success in aggregate across the program. The pre-survey established a cohort baseline to be compared against the post-survey results. We used a Likert scale to measure participant agreement with provided statements. A comment section gathered additional feedback. The survey statements were to be scored from 1 (strongly disagree) to 5 (strongly agree). University Internal Review Board (IRB) approval was received before distributing the surveys. Provided statements focused on participants' perceptions of themselves and their abilities pertaining to higher education, STEM careers, research, and general computer science techniques (see Table 5). Examples of the statements included:

- *"I am able to design and conduct experiments using code."*
- *"I am able to analyze data using code and provide exploratory data analysis of big data sets."*
- *"I am able to deliver an oral presentation of the algorithm in technical terms."*
- *"I am able to demonstrate proficient use of modern computer tools for writing, presentation, project management, and group communication."*

Table 5: Descriptive statistics of mean, mode, median, and range of participants' responses to the computing skills and soft skills questionnaires. The scores for each questionnaire were computed by averaging each participant's Likert responses where 1 is strongly disagree and 5 is strongly agree.

Survey	**Descriptive Statistic**	**Computer Science Skills**	**Soft Skills Proficiency**
Pre-Survey	Mean	1.89	2.96
	Mode	1.14	2.67
	Median	1.57	3.00
	Range	3.43	3.00
Post-Survey	Mean	4.16	4.48
	Mode	-	4.67
	Median	4.14	4.67
	Range	1.57	1.33

6.2 Weekly Reflections

Weekly reflections were designed to evaluate the program curriculum and leadership team through participant morale. These reflections aided the leadership team in adjusting the pace of the curriculum modules on a weekly basis. Participants responded to the following questions:

- *"What went well this week?"*
- *"What did not go well this week?"*
- *"How can we improve your [program] experience?"*

Time was set aside in the final meeting of each week for participants to complete their reflections. The program leadership team collected and subsequently discussed anonymous responses during the end-of-week organizational meetings.

7 Results

7.1 Comparison of Participant Pre- and Post-Survey Scores

At the conclusion of the program, participants' responses to the pre- and post-surveys were compared. Some participants did not respond to the post-survey after the conclusion of the program. As participant responses were anonymized and compared in aggregate, that unresponsiveness is not considered to have greatly impacted the general result. Participants' self-rated computing skills from the pre- and post-surveys are reported in Figure 3. The initial pre-survey revealed only a single instance of agreement with the survey statements describing computer science skills. This coincided with the self-reported participant demographics (seen in Figure 1), where a single respondent indicated a computer science background. In contrast, no participants indicated a score less than "4" (agree) when responding to questions measuring computing skills in the post-survey. A two-tailed Mann-Whitney U test with a significance threshold of $p < 0.05$ was used to investigate the null hypothesis that the medians of the responses from the pre- and post-surveys were identical. The calculated U was 4, which was lower than the critical value of U at $p < 0.05$ (U = 12). The alternative hypothesis, that the two sets of participant responses had different medians, was accepted with high confidence. At the close of the program, as the responding participants indicated great improvement in self-reported perceptions of computer science ability and the alternative hypothesis had been accepted with high confidence, the program was interpreted to be successful, with respect to the small sample size.

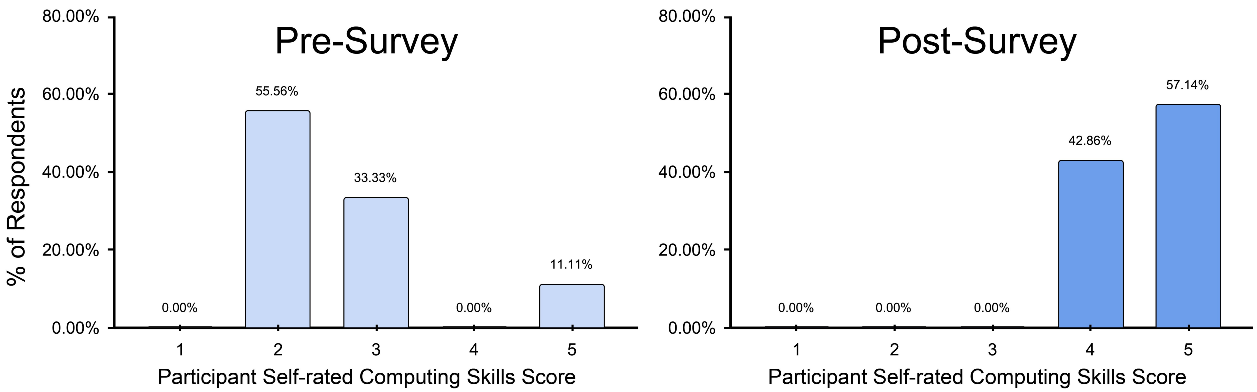


Figure 3: Responding participants' self-rated computing skills scores compared in the aggregate from pre- (left) and post-surveys (right).

7.2 Participant Responses Guided Curriculum Adaptations

Leadership had anticipated a wide range of familiarity and comfort with computer programming techniques, research, and STEM fields based on the program recruitment process. The pre-survey baseline established by the program revealed no major surprises about the participants. Participant pre-survey responses were used to inform adaptations of the curriculum. Specifically, CM week 3 was only designated to expose the participants to more detailed NLP methods. Responses to the pre-survey were used to preemptively modify the program curriculum to allow the possibility of using that week as either a teaching or a review week. This precaution was shown to be warranted upon receiving participants' reflection comments during the initial weeks of the CM. In a similar way, further responses indicated a desire for more practice, so an academic organization subscription was obtained to the DataCamp website. Weekly reflections also served to checkpoint participant morale and progress during the program, in both the CM and RM. Participant comments were examined in the aggregate from week to week.

8 Discussion

In light of the ongoing COVID-19 pandemic and the parallel spread of misinformation in online discourse, more students are personally aware of the difficulties facing them medically, academically, professionally, and societally. The use of a novel, contemporary, and highly relevant data set to apply computational research served to ground computer science concepts from ML and NLP into approachable research tasks. The above results, combined with the positive feedback from committee members, and a currently in-progress research paper, point to the efficacy of this approach for training small, diverse cadres of researchers.

A major factor in the program's success was the implementation of clear pathways of accountability supported by channels of communication. Assigned team roles allowed participants to build a sense of agency and ownership for their stake in the project through clearly defined responsibilities. This approach was mirrored in the feedback systems built into the program to ensure the participant's success. As a caveat for future implementations, participants' workload and daily engagement was demanding. Feedback channels were employed to measure this learning tension on a week by week basis and subsequently adjust the program. A similar approach might not be viable at a larger scale, which is an obvious limitation of this program.

Participant recruitment is essential to the program's success, and there is room for improvement. In the future, the pre- and post-surveys will be increased in complexity to allow for a larger array of scoring metrics. These improvements will improve the baseline to judge participant progress over the program.

Another area of improvement for this program would be to include a more comparative research design to increase the ability to measure educational outcomes. To address this concern, the pre-survey could be sent to all applicants (instead of only recruited participants) in order to provide a background value, not only for those selected for participation but also individuals from the institution that are being reached by the advertising and promotional efforts. Such a metric would give a year-to-year comparison for changes in the overall selection pool.

Additionally, the weekly reflection surveys could also be improved by adding objective or quantitative questions to specifically measure morale. For example, we could add: *"From 1 to 5, how successful did you feel this week in your goals?"* or *"From 1 to 5, how well did you understand your goals?"*

Availability of PI's and Graduate Research Assistants (GRA) are factors for scaling the program. The organizational ratio maintained within the program structure was 1 : 1 : 3 for PI : GRA : participants per team. To address this concern of scalability, the number of participants per GRA could be increased. However, the upper limit of participants that can be mentored by each GRA is likely small. Such a consideration is dependent on a large number of factors in the skills, comfort level, goals, and engagement of participants and GRAs. Those factors can be anticipated to a certain extent by advertisement, recruitment, and application processes.

The dual mentor-mentee relationships present in the program: between PIs and GRAs or GRAs and participants added further benefit to this program. Many graduate assistants in academia view their work as an inherently solo pursuit. Opportunities for team leadership or management are often overlooked. This academic reality is in stark contrast to the team-focused nature of professional working groups in the military, government, and industry, where the management and cooperation of individuals within a whole is paramount. During this program, the GRAs received training from the PIs in mentorship, organization, research program development, and team management. Here the program provided GRAs an opportunity to develop leadership experience, organizational skills, and research methods, in contrast to traditional teaching assistantships focused more exclusively on education.

9 Conclusion

This program demonstrates methods for teaching educationally diverse undergraduate students in computer science research through the implementation of an accountability-focused pedagogical method using novel real-world data sets. At the close of the program, participants from a wide range of educational backgrounds demonstrated their success by presenting individual portions of a greater research pursuit to a large group of experts in academia, government, and industry. The program demonstrates the importance of adaptability to effectively implement engaged online learning. Other computer science research mentorship programs could benefit from the research focus of the program, the deliberate requests for feedback, and the recruitment process described here.

We have demonstrated that students outside computer science disciplines can perform quality research and present their findings to a multidisciplinary audience of experts. Academic institutions and research laboratories should implement organizational models that emulate the multidisciplinary approaches in common use among industry professionals. To aid in this goal, open feedback channels, practice in analysis and presentation, are necessary for effective teamwork in solving real-world problems.

To address continuing and future emergent electronic threats like online misinformation, prospective analysts must be prepared. Analysts are high-performing individuals that demonstrate strong abstract reasoning skills, possess deep background knowledge, and develop strong self-motivational skills. These skills must be fostered through mentorship, communication, and practice. This program serves as a means to recruit and train future analysts, by extending opportunities for participation in computer science education to undergraduate students outside of computer science. Such multidisciplinary individuals can comprise teams that effectively address relevant subjects through the production and presentation of research to expert audiences.

10 Acknowledgments

The work presented in this paper was supported by the Office of Naval Research (ONR) Navy ROTC Cybersecurity Training Program grant number N00014-20-1-2804; authors appreciate Dr. Michael Simpson and Dr. R. Marielle Hjelm for their support and guidance during the execution of the project. We also thank Khyati Mahajan for her efforts in data collection, specifically, hydrating the tweets used in this research, and answering our questions on the dataset. The authors appreciate all the support from the Office of OneIT, University Research Computing staff at the University of North Carolina at Charlotte who have assisted us in utilizing high-performance computing resources for this project.

References

[1] World Health Organization (WHO), "COVID-19 weekly epidemiological update, edition 72, 28 December 2021," WHO, Tech. Rep. 72, Dec. 2021.

[2] E. Dong, H. Du, and L. Gardner, "An interactive web-based dashboard to track COVID-19 in real time," *Lancet Infect. Dis.*, vol. 20, no. 5, pp. 533–534, May 2020.

[3] S. X. Zhang, Y. Wang, A. Rauch, and F. Wei, "Unprecedented disruption of lives and work: Health, distress and life satisfaction of working adults in China one month into the COVID-19 outbreak," *Psychiatry Res.*, vol. 288, p. 112958, Jun. 2020.

[4] A. El Masri and E. Sabzalieva, "Dealing with disruption, rethinking recovery: Policy responses to the COVID-19 pandemic in higher education," *Policy Design and Practice*, vol. 3, no. 3, pp. 312–333, Jul. 2020.

[5] A. F. Tasso, N. Hisli Sahin, and G. J. San Roman, "COVID-19 disruption on college students: Academic and socioemotional implications," *Psychol. Trauma*, vol. 13, no. 1, pp. 9–15, Jan. 2021.

[6] J. Roozenbeek, C. R. Schneider, S. Dryhurst, J. Kerr, A. L. J. Freeman, G. Recchia, A. M. van der Bles, and S. van der Linden, "Susceptibility to misinformation about COVID-19 around the world," *R Soc Open Sci*, vol. 7, no. 10, p. 201199, Oct. 2020.

[7] S. Evanega, M. Lynas, J. Adams, and K. Smolenyak, "Coronavirus misinformation: Quantifying sources and themes in the COVID-19 'infodemic'," https://www.uncommonthought.com/mtblog/wp-content/uploads/2020/12/Evanega-et-al-Coronavirus-misinformation-submitted_07_23_20-1.pdf, accessed: 2022-2-9.

[8] B. S. Fakhruddin, K. Blanchard, and D. Ragupathy, "Are we there yet? The transition from response to recovery for the COVID-19 pandemic," *Prog Disaster Sci*, vol. 7, p. 100102, Oct. 2020.

[9] World Health Organization (WHO), "Novel Coronavirus(2019-nCoV) situation report-13," WHO, Tech. Rep. 13, Feb. 2020.

[10] N. Fraser, L. Brierley, G. Dey, J. K. Polka, M. Pálfy, F. Nanni, and J. A. Coates, "The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape," *PLoS Biol.*, vol. 19, no. 4, p. e3000959, Apr. 2021.

[11] J. J. Van Bavel, K. Baicker, P. S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M. J. Crockett, A. J. Crum, K. M. Douglas, J. N. Druckman, J. Drury, O. Dube, N. Ellemers, E. J. Finkel, J. H. Fowler, M. Gelfand, S. Han, S. A. Haslam, J. Jetten, S. Kitayama, D. Mobbs, L. E. Napper, D. J. Packer, G. Pennycook, E. Peters, R. E. Petty, D. G. Rand, S. D. Reicher, S. Schnall, A. Shariff, L. J. Skitka, S. S. Smith, C. R. Sunstein, N. Tabri, J. A. Tucker, S. v. d. Linden, P. van Lange, K. A. Weeden, M. J. A. Wohl, J. Zaki, S. R. Zion, and R. Willer, "Using social and behavioural science to support COVID-19 pandemic response," *Nat Hum Behav*, vol. 4, no. 5, pp. 460–471, May 2020.

[12] S. Chen, Q. Xu, J. Buchenberger, A. Bagavathi, G. Fair, S. Shaikh, and S. Krishnan, "Dynamics of health agency response and public engagement in public health emergency: A case study of CDC tweeting patterns during the 2016 Zika epidemic," *JMIR Public Health Surveill*, vol. 4, no. 4, p. e10827, Nov. 2018.

[13] V. H. Murthy, "Confronting health misinformation: The U.S. surgeon general's advisory on building a healthy information environment [internet]," US Department of Health and Human Services, Tech. Rep., 2021.

[14] O. Lee and T. Campbell, "What science and STEM teachers can learn from COVID-19: Harnessing data science and computer science through the convergence of multiple STEM subjects," *J. Sci. Teacher Educ.*, vol. 31, no. 8, pp. 932–944, Nov. 2020.

[15] E. Chen, K. Lerman, and E. Ferrara, "Tracking social media discourse about the COVID-19 pandemic: Development of a public coronavirus twitter data set," *JMIR Public Health Surveill*, vol. 6, no. 2, p. e19273, May 2020.

[16] D. S. Raicu and J. D. Furst, "Enhancing undergraduate education: A reu model for interdisciplinary research," in *Proceedings of the 40th ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 468–472. [Online]. Available: https://doi.org/10.1145/1508865.1509027

[17] P. O. Franco, J. S. Ramírez-Lugo, H. O. Zuazaga, M.-E. P. Hernández, L. R. Pericchi, and J. E. García, "Enhancing undergraduate education and curriculum through an interdisciplinary and quantitative initiative to broaden participation in big data," in *Proceedings of the 18th Latin American and Caribbean Consortium of Engineering Institutions (LACCEI) International Multi-Conference for Engineering, Education and Technology*, Virtual Conference, July 2021.

[18] J. Fairley, J. Auerbach, A. Prysock, L. Conrad, and G. May, "Teaching research skills in summer undergraduate research programs," in *2008 Annual Conference & Exposition*, no. 10.18260/1-2–4091. Pittsburgh, Pennsylvania: ASEE Conferences, June 2008, https://peer.asee.org/4091.

[19] M. Ross, E. Litzler, and J. Lopez, "Meeting students where they are: A virtual computer science education research (CSER) experience for undergraduates (REU)," in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, ser. SIGCSE '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 309–314. [Online]. Available: https://doi.org/10.1145/3408877.3432547

[20] A. Rorrer, D. K. Pugalee, P. Ramaprabhu, M. Uddin, H. P. Cherukuri, T. Xu, and D. Prajapati, "Pathways to entrepreneurship (PAtENT) program: Reimagining STEM doctoral programs," in *2021 ASEE Virtual Annual Conference Content Access*. ASEE Conferences, 2021.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011.

[22] DataCamp, "DataCamp: Learn R, Python & data science online," https://www.datacamp.com/, 2022, accessed: 2022-2-9.

[23] harlfoxem, "House sales in King County, USA," 2016.

[24] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, Sep. 1936.

[25] J. J. Hull, "A database for handwritten text recognition research," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 550–554, May 1994.

[26] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon, USA: Association for Computational Linguistics, Jun. 2011, pp. 142–150.

[27] N. Nigro, "Hamilton monitored accounts on Twitter," https://securingdemocracy.gmfus.org/hamilton-monitored-accounts-on-twitter, accessed: 2022-2-9.

[28] M. M. Dorodchi, N. Dehbozorgi, A. Benedict, E. Al-Hossami, and A. Benedict, "Scaffolding a team-based active learning course to engage students: A multidimensional approach," in *2020 ASEE Virtual Annual Conference Content Access*. peer.asee.org, 2020.

Appendix A Program Survey

Cyber-ONR Program Survey

Welcome to the Cyber-ONR Summer Program! Thanks for participating in our survey! The Google Form is split into two steps: The first step is to consent and the second step is the survey.

You will need to be logged in with your UNCC Gmail account to complete this step.

* Required

Consent to Participant in a Research Study

Title of the Project: Engaging Future Generation War Fighters and Civilians in Research to understand Cyber and Electronic Warfare, Cyber Defense, and Data Driven Modeling Tools
Principal Investigator: Mesbah Uddin, Ph.D., UNC Charlotte
Co-investigators: Samira Shaikh, Ph.D., UNC Charlotte; Dan Janies, Ph.D., UNC Charlotte
Graduate Research Assistants: Erfan Al-Hossami, UNC Charlotte; David Brown, UNC Charlotte
Study Sponsor: This research is being funded by the Office of Naval Research.
You are invited to participate in a survey. Participation in this survey is voluntary. The information provided here is to help you decide whether or not to participate. If you have any questions, please ask.

Important Information You Need to Know

The purpose of this survey is to understand how effective the summer program would be/was in your motivation for continue education in a STEM field, or participate in Navy program and also in Cyber-Defense research.

Description of Participation

1. You will be asked to read and agree to this consent form. This step will take about 1 minute.
2. You will be asked to complete an online survey that contains 14 questions. Your task is to give your thoughts on these 14 questions. This step will take about 10-15 minutes.

Eligibility

You may participate in this study if you are above 18 years old or older and if you can comfortably communicate in spoken English and are currently participating in the Summer program funded by this research.

Length of Participation

Participation should take approximately 20 minutes.

Risks of Participation

The risks associated with this survey are minimal and no greater than what you would experience in daily life.

Benefits of Participation

You will not be paid for your participation in this survey.

Volunteer Statement

You are a volunteer. The decision to participate is completely up to you. If you decide to do the survey, and then change your mind, you may stop at any time without any consequences. Participation in this survey, or withdrawing at any time, will not affect your status in any way.

Confidentiality:

Any information about your participation, including your identity, is completely confidential. The following efforts will be taken to protect confidentiality and privacy:

All data will be stored on a password protected server. All data from the survey will be pooled and published in aggregate form only. No identifiers, including the study ID number, will ever be connected to any quotation. After this study is complete, identifiers will be removed from the data and the data could be used for future research studies or distributed to another investigator for future research studies without additional informed consent.

Fair Treatment and Respect

UNC Charlotte wants to make sure that you are treated in a fair and respectful manner. Contact the UNCC Office of Research Protections and Integrity (704.687.1871) if you have any questions about how you are treated as a study participant. If you have any questions about the project, please contact Dr. Samira Shaikh (sshaikh2@uncc.edu)

Approval Date

This form was approved by the UNCC Institutional Review Board for use on XX,XXXX for one year.

Participant Consent

I have read the information in this consent form. I can ask questions about participating in this study at any time by emailing Dr. Samira Shaikh (sshaikh2@uncc.edu. I am at least 18 years of age and I agree to participate in this research project. To receive a copy of this consent form I take screen shots of the consent text and/or email Dr. Samira Shaikh (sshaikh2@uncc.edu for an electronic copy of the consent.

1. Please select whether you would like to participate or not using the two options below. If you indicate that you do not want to participate, the survey will end. If you indicate that you agree to participate, please select that option below and then click the button at the bottom right of the screen to continue. *

   Mark only one oval.

   - I agree to participate   *Skip to question 2*
   - I do not wish to participate

Survey Questions

2. Your Name *

   ____

3. Gender *

   ____

4. Are you a first generation student? *

   Mark only one oval.

   - Yes
   - No
   - Maybe

5. Are you part of military family? *

Mark only one oval.

- ( ) Yes
- ( ) No
- ( ) Maybe

6. I am able to design and conduct experiments using code *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	( )	( )	( )	( )	( )	Strongly Agree

7. I am able to explain the functions of an algorithm for a complete solution *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	( )	( )	( )	( )	( )	Strongly Agree

8. I am able to use data acquisition code and understand their proper use and interpretation *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	( )	( )	( )	( )	( )	Strongly Agree

9. I am able to analyze data using code and provide exploratory data analysis of big data sets *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	Strongly Agree

10. I am able to recognize the need for performance measures (accuracy, F-score, AUC) and include estimates of combined uncertainties with reported values, based on reasonable baselines *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	Strongly Agree

11. I am able to design an algorithm based on requirement laid out from a research question *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	⬭	⬭	⬭	⬭	⬭	Strongly Agree

12. I am able to implement basic algorithms that use machine learning as part of an overall system *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

13. I am able to outline and explain a procedure for algorithm solution that begins with user requirements, develops functional specifications, analyzes a variety of alternatives, and choose a final design to best meet user requirements *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

14. I am able to deliver an oral presentation of the algorithm in technical terms *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

15. I am able to demonstrate proficient use of modern computer tools for writing, presentation, project management, and group communication *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

16. I am motivated to pursue a STEM career in the field of Cyber Defense and Electronic Warfare *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

17. I am motivated to pursue a career in the Navy *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

18. I am motivated to pursue higher education (graduate degree in data science, cyber defense, or electronic warfare) *

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

19. Your comments about your experience in the program: *

____
____
____
____
____