WIP: *Python for Everyone* as a Mathematics GE Course: Broaden Participation and Enhance Data Science Career Pipeline

Wendy Lee
Department of Computer Science
San José State University
San José, CA, USA
wendy.lee@sjsu.edu

Melody Moh
Department of Computer Science
San José State University
San José, CA, USA
melody.moh@sjsu.edu

Rula Khayrallah
Department of Computer Science
San José State University
San José, CA, USA
rula.khayrallah@sjsu.edu

Nada Attar
Department of Computer Science
San José State University
San José, CA, USA
nada.attar@sjsu.edu

Kathy Lam
Department of Computer Science
San José State University
San José, CA, USA
kathy.lam@aqtual.com

***Abstract*—This work-in-progress (research-to-practice) paper describes our approach to designing a General Education (GE) Math course, aimed at enabling students not majoring in Computer Science (CS) to learn and apply computer programming in their own fields of study. Research has shown that underrepresented groups face barriers such as economic and educational disparities, and cultural biases, which limit their access to Data Science (DS) and Artificial Intelligence (AI) education. Given the rapid growth of these technologies, the critical need for a diverse and inclusive workforce is increasingly evident. To address this, educational opportunities tailored for these groups are essential. Our course, *Python for Everyone*, is a Mathematics/Quantitative Reasoning GE offering and a core requirement for our BS-DS program, designed to meet these needs. It welcomes students from all disciplines, integrating programming skills with mathematical concepts, thereby democratizing access to DS and AI skills, and fostering interdisciplinary collaboration. The course design ensures students from various majors can recognize the relevance and application of computer programming in their fields. This paper delves into the curriculum's theoretical foundations, its structure, and the outcomes. Notable achievements include enhanced gender and diversity representation: 39% of the participants were female, surpassing the 25% in traditional CS majors, and the enrollment of underrepresented minority (URM) students reached 19%, compared to 7% in typical CS courses. These outcomes demonstrate the course's effectiveness in fostering diversity, inclusion, and accessibility in DS education. *Python for Everyone* serves as a pioneering model, guiding students, particularly from underrepresented backgrounds, and providing a pathway to successful careers in DS and AI. Our ongoing efforts involve conducting more surveys and assessments to refine our course, focusing on incorporating more real-world applications to highlight the relevance of DS across disciplines, aiming for a more inclusive workforce, and assessing the impact of this course in enhancing educational and career opportunities for URM students. We believe this paper is a valuable resource for educators dedicated to enhancing GE quantitative reasoning courses and broadening career opportunities in CS, DS, and AI for students from underrepresented groups.**

***Index Terms*—Diversity concerns, Computing skills, Interdisciplinary**

I. INTRODUCTION

In recent years, Data Science has grown significantly, becoming influential in fields like finance, health care, and scientific research [1], [20], [30]. This growth has created a high demand for skilled data scientists, while also highlighting the need for a diverse and inclusive workforce. Currently, underrepresented groups such as women, racial and ethnic minorities, and those from lower socioeconomic backgrounds face significant challenges, with limited opportunities in Computer Science and Data Science [2], [6], [33], [34]. Prevailing stereotypes reinforce the idea that women do not belong in these fields. [13], [22], [25]. Although Data Science offers unique opportunities to bridge these gaps, women represent only 15% to 22% of professionals in the field, with just 18% holding leadership positions at leading tech companies [5].

To address these disparities, we introduced *Python for Everyone*, a Mathematics/Quantitative Reasoning General Education (GE) course. General Education in the United States refers to a broad curriculum that provides students with a well-rounded foundation of knowledge and skills across various disciplines, regardless of their major. Unlike previous initiatives focused on specific majors, our course meets the Mathematics/Quantitative Reasoning GE requirement while introducing Python programming for practical mathematical applications. This inclusive approach makes data science and computing accessible to all students from any major. This course is also a core requirement of the BS Data Science program, providing a solid foundation in both mathematics and programming [28].

This paper provides a comprehensive overview of the curriculum development process for *Python for Everyone*. We explore the theoretical frameworks and concepts used in designing the course and highlight their relevance to Data Science applications. By sharing our experiences, we aim to guide similar initiatives promoting diversity and inclusion in Data Science.

II. Related Work

A. Entry-Level Computer Programming Courses for Non-Computing Majors

In recent years, data-centric introductory computing courses have encouraged non-computing majors to explore programming. Erkan and Lee bridged the gap between Computer Science courses for majors and non-majors by enhancing a non-majors course focused on tabular data [14]. Liu et al. developed an entry-level data science course tailored for non-computing majors with positive results at Rochester Institute of Technology [24]. Burridge and Fekete designed a first-year data science course teaching Python through practical patterns, providing foundational skills for both majors and non-majors [8]. UC Berkeley's "Foundations of Data Science" (Data 8) course emphasizes real-world applications and societal implications while nurturing computational skills, diversity, and ethical awareness [36]. However, these courses focus on using libraries like Scikit-Learn, Pandas, Matplotlib, and NumPy without covering underlying mathematical concepts.

B. Integrating Computer Programming and Mathematics into the Data Science Curriculum

Early career experiences in Data Science underscore the need for students to understand how computer science, mathematics, statistics, and applied fields intersect [11], [19]. Integrating computer science into the mathematics curriculum has been advocated in recent studies [12], [15], [19]. Kao calls for integrating computer science courses into traditional mathematics to bridge the gap between theory and application, catering to applied math majors and equipping them with relevant knowledge for data science, machine learning, and AI [19]. Vírseda introduces theoretical computer science concepts early in the undergraduate mathematics curriculum, showing positive impacts on student performance [12]. Such integration demonstrates how programming can deepen understanding of mathematical theories and their applications, yet these integrated courses are not designed for non-computing majors.

III. Curriculum development

A. Python for Everyone: Filling the Gap in Data Science Education for Non-Computing Majors

Python for Everyone addresses a gap in introductory programming courses by integrating mathematics and programming techniques for non-computing majors to solve real-world problems. Current data-centric courses focus on programming libraries for data analysis without covering underlying mathematical concepts, and integrated CS-math curricula often cater primarily to computing and math majors. To address this, we developed *Python for Everyone*, a General Education (GE) course that merges introductory statistics with Python programming. It provides early exposure to the interdisciplinary nature of Data Science for students across disciplines. Drawing from well-established literature, the course aims to enhance student engagement and participation by expanding the range of GE courses in CS.

Key factors driving this initiative include:

- **Interdisciplinary Engagement**: Computing courses should incorporate interdisciplinary perspectives, enriching students' understanding [9], [10], [21], [27], [35].
- **Increased Participation**: Diverse perspectives in computing can boost participation in Data Science [4], [23].
- **Inclusivity for Other Majors**: Integrating programming into a math GE course makes programming more accessible to students from various majors [17], [18].

B. Theoretical Frameworks

Our curriculum emphasizes inclusive pedagogy, active learning strategies, and authentic projects. Collaborative learning fosters a supportive community, teamwork, and communication skills. The course is designed based on best practices in Computer Science Education, aiming to:

- Provide relevant and practical instruction and assignments [23], [29].
- Foster a supportive and welcoming environment [4], [32].
- Use diverse assessments for feedback to students and instructors [3], [37].
- Implement Universal Design for Learning (UDL) and promote active learning via pre-class videos, hands-on exercises, group projects, and problem-solving activities [7], [16], [31].

IV. Curriculum Overview

A. Python for Everyone Concepts

Python for Everyone introduces Python programming and essential mathematical concepts, accessible to students from various backgrounds with no prior programming experience. The course aims to equip students with skills for careers in Data Science and covers:

- Python Programming Fundamentals for Data Science:
  - Basic syntax, data types, control structures
  - User-defined functions
  - Mathematical libraries: Pandas, NumPy, SciPy
  - Data visualization libraries: Matplotlib, Seaborn
- Mathematical Concepts in Data Analysis:
  - Central tendency, variability, normal distribution
  - Standardized scores, sampling distribution, standard error
  - Hypothesis testing, T-tests, One-way ANOVA, Chi-squared test
  - Correlation and Regression
- Effective Communication of Insights from Data

B. Course Structure

Python for Everyone is an inclusive and interactive in-person course that blends pre-class videos, concise lectures, hands-on activities, and collaborative group projects. All materials are organized in Canvas, our Learning Management System (LMS). An interactive digital textbook [26] allows students to learn at their own pace, including homework exercises with adaptive learning and automatic grading for prompt feedback. Each class session features an instructional assistant providing immediate, personalized support during hands-on activities.

1) Pre-class Learning and Formative Assessment: Pre-class videos cover upcoming topics, and short quizzes assess comprehension before class (Fig. 1). This approach enhances student preparation and fosters productive in-class discussions.

2) Classroom Activities and Summative Assessment: The class comprises four components: Lecture, Hands-on Activities, Discussion, Group Projects, and Assessment.

- **Lecture**: Live coding using Google Colab, a cloud-based platform for writing, executing, and sharing Python code, exemplifies programming concepts (Fig. 2).
- **Hands-on Activities**: Students work independently or collaboratively on assignments using Google Colab.
- **Discussion**: Students share solutions, challenges, and successes with the class.
- **Group Term Projects**: Students collaborate in groups to complete a month-long data analysis project using datasets relevant to their interests and applying their skills to solve real-world problems. At the end, each group presents their project results in class.
- **Assessment**: Individual tests, group projects, and exams.

This interactive approach empowers students with practical knowledge, enabling them to confidently apply Python and mathematical principles in Data Science.

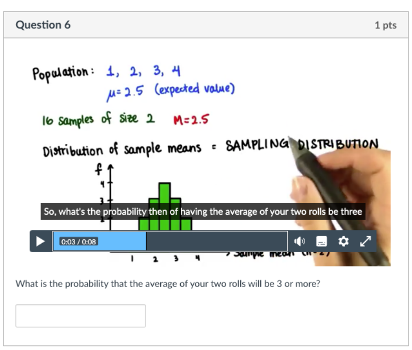

Fig. 1. Example of a pre-class video in Canvas and the corresponding quiz question, with closed captioning provided as an option.

C. Assess Pre- and Post-Course Student Expectations and Feedback

Survey analysis of pre- and post-course expectations and feedback provides valuable insights into the effectiveness of our Python programming course and its impact on students. The goal is to identify strengths and areas for improvement while broadening access to Data Science education.

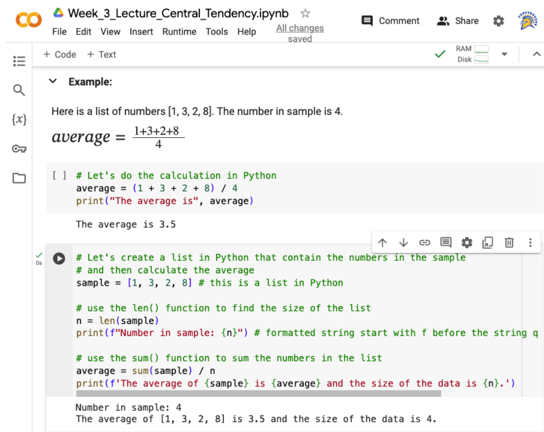

Fig. 2. Example of a Google Colab notebook with lecture materials

1) Pre-Course Student Expectations: A pre-course survey gauges students' expectations and motivations for enrolling in *Python for Everyone*. It collects data on their coding knowledge, Data Science interests, and career aspirations, helping us tailor the curriculum to meet their needs.

2) Post-Course Student Feedback: A post-course survey gathers feedback on the learning experience, curriculum, teaching methods, assignments, and resources. It assesses student satisfaction and identifies areas for improvement.

3) Survey Analysis: Comparing pre- and post-course surveys helps assess how well the course meets students' expectations and learning objectives.

TABLE I
UNDERGRADUATE ENROLLMENT STATISTICS ACROSS DIFFERENT TIERS AT UNIVERSITY1 IN AY 2022-2023

Average Fall 2022 to Fall 2023 *	Headcounts	Female	URM
All Colleges *	26,277	48%	34%
College of Science *	2,695	48%	22%
Computer Science major *	777	25%	7%
Data Science major (Fall 2023)	234	24%	9%
Python for Everyone *	135	39%	19%

V. RESULTS

A. Python for Everyone Bridges Gender and Minority Gaps

In 2022, we launched the *Python for Everyone* course with encouraging outcomes. The course, integral to the data science major, attracted a diverse student body. Notably, 51% of enrolled undergraduates were from non-Data Science and non-STEM disciplines (Fig. 3). The course improved gender and diversity representation, with 39% female participants (compared to 25% in Computer Science) and 19% underrepresented minority (URM) students (versus 7% in Computer Science, see Table 1). Our institution is located in California, where underrepresented minority (URM) students typically include those from Hispanic or Latino/a/x, Black or African American, Native American or Alaska Native, and Native Hawaiian or Other Pacific Islander backgrounds, as they are historically underrepresented in higher education, particularly in STEM fields.

In fall 2023, we introduced our BS Data Science (BSDS) program, which had demographics similar to BS Computer Science (BSCS). However, *Python for Everyone* featured a more diverse student body, highlighting its appeal to a broader range of students (Table I).

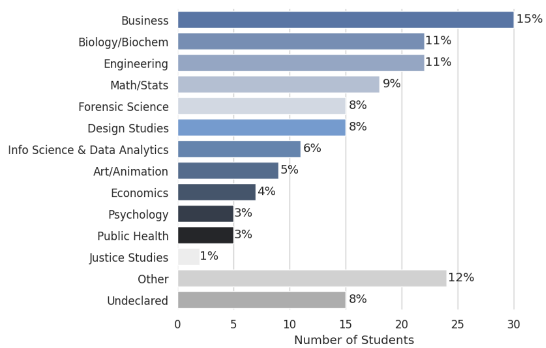

Fig. 3. Non-Data Science Students completed Python for Everyone in AY2022-2023 by Majors.

Most notably, *Python for Everyone* drew a notable interest from Latinx students, constituting 13% of the participants, along with 6% representation of Black/African-American students. This result holds considerable importance, given that prior to the course's introduction, Computer Science majors witnessed meager representation with only 5% Latinx students and 1% Black/African-American students.

B. Gender and Minority Representation in Course Outcomes

Our data reveals higher success rates among female and Non-URM students. Specifically, 83% of female students completed the course successfully, compared to 78% of male students. Non-URM students had an 82% success rate versus 71% for URM students. Conversely, URM students had a higher DFW (Drop, Fail, Withdrawal) rate of 29% compared to 18% for Non-URM students. These disparities underscore the need to address underlying factors to create an equitable academic environment.

C. Pre-course survey highlights students' motivations in learning programming

The pre-course survey reveals that students are optimistic about learning Python and have clear expectations for the course. Key highlights include:

- Motivations: Learning Python for personal, professional, and community growth, applying it to personal life and future careers.
- Interdisciplinary Interest: Exploring Data Science, Bioinformatics, UI/UX design, and expanding programming knowledge for job opportunities.
- Technical Skills: Building skills for a career as a financial analyst and recognizing programming's versatility.
- Future Goals: Preparing for future careers and master's degrees in Data Science or Computer Science while fulfilling programming requirements.

D. Students' Post-course Reflection and Feedback

To improve future student success, we analyze post-course surveys to identify strengths and challenges. This helps us implement strategies to enhance learning and course outcomes.

1) Key Strengths:

- Engaging Learning: An engaging, interactive environment with multimedia resources and hands-on activities makes learning enjoyable and memorable.
- Practical Projects: Authentic projects involving real-world data analysis help students apply Python skills and prepare for career opportunities.
- Supportive Environment: An inclusive and supportive atmosphere encourages participation, collaboration, and seeking help from instructors and peers.

2) Areas for Improvement:

- Clarifying Complex Topics: Some students need more clarity on complex topics. We will provide additional explanations, examples, and resources to improve comprehension.
- Balancing Course Pace: To accommodate varying learning speeds, we will adjust the course pace and offer supplementary materials for self-paced learning.
- Timely Feedback on Assessments: Timely feedback is crucial for student progress. We will establish realistic grading timelines, use automated grading tools, and encourage peer review and self-assessment.

VI. Conclusion

This paper outlines our experiences in designing the *Python for Everyone* course as a Mathematics/Quantitative Reasoning GE course, focusing on promoting broader participation and pathways to Data Science careers. We incorporate theoretical foundations such as inclusive pedagogy, and active learning strategies to create an engaging learning environment, fostering fundamental programming skills, critical thinking, and a deep understanding of mathematics in Data Science.

We identified course strengths through analysis, including well-structured content, engaging teaching methods, relevant assignments, and meaningful group projects, resulting in positive changes in students' perceptions of Python programming and Data Science. Areas for improvement were also identified via survey analysis, highlighting the need for clearer instruction, timely feedback, and balanced course pace. Using these insights, we refine the course to enhance future learning experiences and align with the evolving needs of Data Science students.

Our goal is to share the challenges, lessons learned, and ongoing efforts to guide similar initiatives and encourage innovative approaches to advancing diversity and accessibility in Data Science.

ACKNOWLEDGMENT

This project was funded by the Center of Inclusive Computing (CIC) at Northeastern University. We would like to express our gratitude to Felix Muzny, Jodi Tims, and Carla Brodley for their valuable suggestions and guidance in developing an inclusive computing program at San José State University.

REFERENCES

[1] Norita Ahmad, Areeba Hamid, and Vian Ahmed. Data science: Hype and reality. *Computer*, 55(2):95–101, 2022.

[2] Noura Albarakati. Analyzing factors that contribute to attrition of underrepresented undergraduate students in computer science, 2022.

[3] Noura Albarakati, Lisa DiPippo, and Victor Fay-Wolfe. Rethinking cs0 to improve performance and retention. In *Proceedings of the 23rd Australasian Computing Education Conference*, pages 131–137, 2021.

[4] Christine Alvarado, Zachary Dodds, and Ran Libeskind-Hadas. Increasing women's participation in computing at harvey mudd college. *acm Inroads*, 3(4):55–64, 2012.

[5] Knowledge at Wharton Staff. Women in data science: Moving from inclusion to influence. https://knowledge.wharton.upenn.edu/article/women-data-science-moving-inclusion-influence/, 2023. 19 January, 2024.

[6] Christopher S Awad, Youping Deng, John Kwagyan, Abiel Roche-Lima, Paul B Tchounwou, Qingguo Wang, and Muhammed Y Idris. Summary of year-one effort of the rcmi consortium to enhance research capacity and diversity with data science. *International Journal of Environmental Research and Public Health*, 20(1):279, 2022.

[7] Sheryl Burgstahler. Universal design: Implications for computing education. *ACM Transactions on Computing Education (TOCE)*, 11(3):1–17, 2011.

[8] Joshua Burridge and Alan Fekete. Teaching programming for first-year data science. In *Proceedings of the 27th ACM Conference on on Innovation and Technology in Computer Science Education Vol. 1*, pages 297–303, 2022.

[9] Valerie Carr, Morris Jones, and Belle Wei. Interdisciplinary computing: Applied computing for behavioral and social sciences. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pages 400–406, 2020.

[10] Xiuqing Chen, Wei Liu, et al. The value of python programming in general education and comprehensive quality improvement of medical students based on a retrospective cohort study. *Journal of Healthcare Engineering*, 2022, 2022.

[11] Andrea Danyluk and Paul Leidig. Computing competencies for undergraduate data science curricula: Acm data science task force. 2021.

[12] Rafael del Vado Vírseda. Learning from the impossible: Introducing theoretical computer science in cs mathematics courses. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pages 952–958, 2021.

[13] Sylvain Duranton, Jörg Erlebach, Camille Brégé, Jane Danziger, Andrea Gallego, and Marc Pauly. What's keeping women out of data science? https://www.bcg.com/publications/2020/what-keeps-women-out-data-science, 2020. Accessed: 19 January 2024.

[14] Ali Erkan and Adam Lee. Connecting majors/non-majors courses via tabular data. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*, pages 235–236, 2019.

[15] Kathi Fisler, Emmanuel Schanzer, Steve Weimar, Annie Fetter, K Ann Renninger, Shriram Krishnamurthi, Joe Gibbs Politz, Benjamin Lerner, Jennifer Poole, and Christine Koerner. Evolving a k-12 curriculum for integrating computer science into mathematics. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pages 59–65, 2021.

[16] Richard M Gargiulo and Debbie Metcalf. *Teaching in today's inclusive classrooms: A universal design for learning approach*. Cengage Learning, 2022.

[17] Pingzhang Gou, Yuyue Han, and WenXuan Zhang. Diversified teaching evaluation of python programming based on obe concept. In *2021 16th International Conference on Computer Science & Education (ICCSE)*, pages 402–406. IEEE, 2021.

[18] Jessen Havill. Embracing the liberal arts in an interdisciplinary data analytics program. In *Proceedings of the 50th ACM technical symposium on computer science education*, pages 9–14, 2019.

[19] Chia Hung Kao. Enriching undergraduate mathematics curriculum with computer science courses. *International Journal of Engineering Pedagogy*, 11(5), 2021.

[20] Z Faizal Khan and Sultan Refa Alotaibi. Applications of artificial intelligence and big data analytics in m-health: a healthcare system perspective. *Journal of healthcare engineering*, 2020:1–15, 2020.

[21] Natalia Khuri, Wendy Lee, K Virginia Lehmkuhl-Dakhwe, Miri Van-Hoven, and Sami Khuri. Interdisciplinary minor in bioinformatics: First results and outlook. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pages 407–412, 2020.

[22] Colleen M Lewis, Ruth E Anderson, and K Yasuhara. I don't code all day: Fitting in computer science when the stereotypes don't fit. In *Proceedings of the 2016 ACM Conference on International Computing Education Research*, ICER'16, pages 23–32, New York, NY, USA, 2016. Association for Computing Machinery.

[23] Albert Lionelle, Josette Grinslad, and J Ross Beveridge. Cs 0: Culture and coding. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pages 227–233, 2020.

[24] Xumin Liu, Erik Golen, Rajendra K Raj, and Kimberly Fluet. Offering data science coursework to non-computing majors. In *Proceedings of the 2nd International Workshop on Data Systems Education: Bridging education practice with education research*, pages 44–49, 2023.

[25] A. Master, A. N. Meltzoff, and S. Cheryan. Gender stereotypes about interests start early and cause gender disparities in computer science and engineering. *Proceedings of the National Academy of Sciences*, 118(e2100030118), 2021.

[26] B. Miller. *Programming in Python 3*. zyBooks, a Wiley brand, 2020. Accessed 2023.

[27] Melody Moh, Rafael Alvarez-Horine, Sagar Satish Chandawale, and Susmit Arun Mogarkar. On interdisciplinary student background: A successful course integrating teaching and research. In *2013 3rd Interdisciplinary Engineering Design Education Conference*, pages 56–62. IEEE, 2013.

[28] Melody Moh, Rula Khayrallah, Wendy Lee, Teng-Sheng Moh, David Taylor, and Ching-Seh Mike Wu. Diversifying the computing workforce – rapid development of interdisciplinary computing-based undergraduate programs. In *2024 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2024.

[29] Briana B Morrison, Beth A Quinn, Steven Bradley, Kevin Buffardi, Brian Harrington, Helen H Hu, Maria Kallia, Fiona McNeill, Oluwakemi Ola, Miranda Parker, et al. Evidence for teaching practices that broaden participation for women in computing. In *Proceedings of the 2021 Working Group Reports on Innovation and Technology in Computer Science Education*, pages 57–131. 2021.

[30] Mahyuddin KM Nasution. Understanding data toward going to data science. In *Computer Science On-line Conference*, pages 478–489. Springer, 2022.

[31] Christine F Reilly, Laura M Grabowski, and Gustavo Dietrich. Improving the success of non-traditional students in an introductory computing course. In *2021 IEEE Frontiers in Education Conference (FIE)*, pages 1–5. IEEE, 2021.

[32] Lauren Rich, Heather Perry, and Mark Guzdial. A cs1 course designed to address interests of women. *Acm sigcse bulletin*, 36(1):190–194, 2004.

[33] Yana Samuel, Jean George, and Jim Samuel. Beyond stem, how can women engage big data, analytics, robotics and artificial intelligence? an exploratory analysis of confidence and educational factors in the emerging technology waves influencing the role of, and impact upon, women. *arXiv preprint arXiv:2003.11746*, 2020.

[34] Linda J Sax, Kari George, Daniel Harris, and Fay Cobb Payton. Reframing the representation of black students in undergraduate computing. *Journal of Women and Minorities in Science and Engineering*, 26(4), 2020.

[35] Andrea Tartaro, Christopher Healy, and Kevin Treu. Computer science in general education: beyond quantitative reasoning. *Journal of Computing Sciences in Colleges*, 32(2):177–184, 2016.

[36] Eric Van Dusen, Anthony Suen, and Cathryn Carson. Innovation in undergraduate data science education. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pages 1397–1397, 2020.

[37] Hidayet Suha Yüksel and Nevin Gündüz. Formative and summative assessment in higher education: Opinions and practices of instructors. *European Journal of Education Studies*, 2017.