Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025
ASEE
Paper ID #45880

Integration of Data Science Modules Across Interdisciplinary Courses at Multiple Institutions: Analysis of Students' and Faculty Perspectives

Mr. Md. Yunus Naseri, Virginia Polytechnic Institute and State University

Yunus Naseri is a Ph.D. candidate in the Department of Civil and Environmental Engineering at Virginia Tech (VT). He joined VT as a master's degree student through a Fulbright Scholarship in 2018. His research focuses on data science literacy integration across STEM+C disciplines and data science application in water use across different economic sectors.

Dr. Vinod K. Lohani, Virginia Polytechnic Institute and State University

Dr. Vinod K. Lohani is a Professor of Engineering Education at Virginia Tech. He served as a Program Director in the Division of Graduate Education, NSF for 4 years (2020-24). In this capacity, he was deeply engaged with the NSF Research Traineeship (NRT), Innovations in Graduate Education (IGE) and CAREER programs and also participated in several NSF-wide working groups on semiconductors and quantum information science and engineering (QISE).

Dr. Manoj K Jha P.E., North Carolina A&T State University

Dr. Manoj K Jha is an associate professor in the Civil, Architectural, and Environmental Engineering department at the North Carolina A&T State University. His research interests include hydrology and water quality studies for water resources management u

Dr. Gautam Biswas, Vanderbilt University

Gautam Biswas is a Cornelius Vanderbilt Professor of Engineering, and Professor of Computer Science, Computer Engineering, and Engineering Management in the Department of Computer Science at Vanderbilt University.. He is also a Senior Research Scientist

Caitlin Snyder

Caitlin Snyder is a PhD student in the department of Computer Science at Vanderbilt University. Her research focuses on understanding how students work collaboratively in open-ended learning environments with the end goal of developing semi-automated anal

Dr. Steven X. Jiang, North Carolina A&T State University

Dr. Steven Jiang is an Associate Professor in the Department of Industrial and Systems Engineering at North Carolina A&T State University. His research interests include Human Systems Integration, Visual Analytics, and Engineering Education.

Caroline Benson Sear, Virginia Polytechnic Institute and State University

Integration of Data Science Modules Across Interdisciplinary Courses at Multiple Institutions: Analysis of Student and Faculty Perspectives

I. Introduction

Data science is emerging as a field that is revolutionizing science and industries alike, with work across nearly all domains becoming increasingly data-driven, affecting both the jobs that are available and the skills required by employers. As more data and analytical methods become available, more aspects of the economy, society, and daily life will become dependent on data-driven decision-making. Recognizing this shift, the National Academies of Sciences (2018) emphasizes that academic institutions must prioritize developing "a basic understanding of data science in all undergraduates" to prepare them for this new era [1]. This is particularly crucial for STEM graduates, who must develop varying levels of expertise in working with data – the ability to understand, interpret, and critically evaluate data, as well as to use data effectively to inform decisions. The recent emergence of large language models (LLMs) such as ChatGPT, which are becoming increasingly proficient at basic to advanced data science skills, has not made acquiring these competencies in undergraduate programs obsolete but rather more relevant, as critical thinking abilities developed through data science literacy education are essential for analyzing LLM outputs [2]. Moreover, when properly integrated into pedagogical practices, these LLMs can facilitate the teaching of data science literacy skills through enhanced personalized learning approaches [3].

Data science literacy education typically follows two main approaches: standalone courses (including general and core disciplinary courses, immersive degrees, minors, certificates, and MOOCs (massive open online course)) or integration within existing disciplinary courses. While standalone approaches are common, students often struggle to apply these skills within their disciplinary contexts [4], [5]. The integrated approach offers a more sustainable and evidence-based method for introducing data science literacy into established curricula [6], [7], helping bridge the instructional gap while aligning with learning theory principles that emphasize building upon students' previous knowledge and experience [8], [9]. While this approach can effectively increase student data science competencies, instructors face significant challenges, including curriculum constraints and supporting students with varying levels of data science familiarity [10], [11]. Although research studies exist on data science integration in both single disciplines (e.g., [12], [13]) and multiple disciplines (e.g., [7], [14]), many prior efforts were standalone approaches that isolated data literacy from disciplinary contexts. Furthermore, principles of integration across STEM disciplines based on input from diverse student and instructor populations are missing, leaving a significant gap in developing common principles for cross-disciplinary data science integration.

Our work adopted an integrated approach to infusing data science into various undergraduate science and engineering courses through discipline-specific modules developed via a multi-university research-practice partnership (RPP), an effective strategy for enhancing the impact of education research on educational practice [15]). Over the four years of our project period, we collected comprehensive data from both instructors and 1000+ students (based on student online surveys) from three participating universities. Instructor data, gathered through semi-structured online interviews conducted twice during the project, captured their experiences during module development and implementation. Student data, collected through pre- and post-implementation online surveys each semester, included demographic information and both Likert-scale and open-ended questions about their perceptions of data science and self-perceived learning of specific module topics (for detailed methodology, see [10], [11], [16], [17]). Our previous work explored instructor perspectives on this integration approach (see [10]) and its efficacy based on students' self-perceived learning and changes in their data science perceptions after taking one or more data science modules (see [17]). This paper synthesizes our project implementation findings into concise lessons learned to inform instructors and departmental members interested in similar data science integration approaches within their curricula. Specifically, we aim to elaborate on key considerations and strategies for effective integration based on instructor and student perspectives. We also discuss the future endeavors of the instructors and the project as a whole and describe the experiences of interdisciplinary graduate students, serving as graduate research assistants (GRAs) in this project, who were deeply engaged with the instructors in implementing various project activities.

The insights presented in this paper offer novel guidance for decision-making in data science integration across diverse STEM disciplines, addressing the current gap in cross-disciplinary data science integration strategies. The remainder of the paper is organized as follows: Section II provides detailed project background; Section III presents key lessons learned from implementation; Sections IV and V explore graduate research assistants' experiences in collaborating with instructors along with future steps for both instructors and the project as a whole, respectively; and Section VI offers our concluding remarks.

II. Project background

Project history: The 2nd author on this paper served as the Director for Education and Global Initiatives at an interdisciplinary research institute during 2016-19 at the lead university, Virginia Tech (VT). During 2017 and 2018, this person served as a diversity champion of this interdisciplinary institute and facilitated the participation of the institute in the HBCU/MSI (Historically Black Colleges and Universities/Minority Serving Institutions) Research Summits at VT. The goal was to invite faculty and students from HBCU/MSI institutions to VT to form partnerships and collaborations in research as well as introduce students from these institutions to the faculty, labs, and facilities at VT in their areas of interest. Typically, these research summits were held in the month of October. A few faculty and students from North Carolina Agricultural and Technical State University (NC A&T) – an HBCU – participated in the 2017 HBCU/MSI research summit at VT. The visitors were taken to various labs and departments to explore collaboration possibilities. It was during this visit in 2017 that the idea to develop a research proposal for the NSF/IUSE (National Science Foundation/Improving Undergraduate STEM Education) program was conceived. The team also decided to invite a data science expert from Vanderbilt University (VU) to join the team. The goal was to incorporate data science modules into a variety of interdisciplinary undergraduate courses at these institutions and develop best practices. The team submitted the proposal in late 2017 which was not successful. The faculty and student team from NC A&T was again invited to the 2018 research summit at VT and the IUSE proposal was revised and resubmitted and the proposal became successful in 2019 and formally the project began in the fall of 2019 with a workshop of all team members that was organized at VT. The 2nd author took an assignment at the NSF in January 2020 and one of his colleagues from the VT took over the PI responsibility and led the project activities in collaboration with team members from NC A&T and VU. In March 2024, the 2nd author took over the project responsibilities again upon completion of his NSF assignment. The project is in No Cost Extension period currently and will expire in September 2025.

Project goals and organization: The project's overarching goal is to develop and implement an interdisciplinary collaborative approach to foster data science expertise among undergraduate students across various STEM disciplines, including engineering, computer science, environmental science, and biology. This goal was pursued through three primary objectives by: (1) integrating real-world data from high-frequency monitoring systems, specifically water monitoring at VT and traffic monitoring at VU, (2) conducting evidence-based research on student learning across diverse demographics, disciplines, institutions, and academic settings, and (3) developing and implementing transferable learning modules to extend the project's impact beyond the partnering universities through existing partnerships.

The project was structured as an RPP, comprising instructors and researchers from the three participating universities, an educational research and assessment firm, and an industry advisory panel with experts from both private and public sectors. This RPP structure enabled flexible responses to partner needs throughout the project (for detailed RPP information, see [10]). Six STEM instructors from participating universities, representing varying levels of data science teaching experience and comfort, collaborated to integrate data science into their courses through discipline-specific modules. These instructors, responsible for teaching courses that differed in academic level, student background, and instructional modality, independently designed modules aligned with their course content, syllabus, and student learning objectives. Each instructor developed one to three data science modules and implemented them multiple times during the project, refining them through implementation experience and discussions with other partners in the project. This resulted in twelve modules developed and implemented across six different STEM courses (Table 1).

Table 1. Participating courses and their discipline-specific modules and implementation semesters.

Course (Course Abbreviation)	University	Department (Discipline)	Module(s)	Implementation Semester(s)
Monitoring and Analysis of the Environment (MAE)	Virginia Tech	School of Plant and Environmental Sciences (Sciences)	Errors in Measured Data	Spring 2020, Spring 2021, Spring 2022
Engineering Hydrology (EH)	North Carolina Agricultural and Technical State University	Department of Civil, Architectural, and environmental engineering (Engineering)	Time-series Analysis of Precipitation Data	Spring 2020, Spring 2021, Spring 2022
Hydrology (HYDRO)	Virginia Tech	Department of Civil and Environmental Engineering (Engineering)	Frequency Analysis in Hydrology	Fall 2020, Fall 2021, Fall 2022
Smart Cities (SC)	Vanderbilt University	University Course (Engineering)	Confidence Interval, Clustering, Supervised Learning	Spring 2020
Ecology (ECO)	Virginia Tech	Department of Biological Sciences (Sciences)	Introduction to data science: Visualization and Interpretation; Ecology is Data!; Effect of acid rain on aquatic and terrestrial ecosystems	Spring 2020, Spring 2021, Spring 2022
Engineering Statistics (ES)	North Carolina Agricultural and Technical State University	Department of Industrial & Systems Engineering (Engineering)	Basic statistics; Hypothesis testing	Spring 2020, Fall 2020, Fall 2021, Spring 2022, Fall 2022

Project achievements: Over the four-year study period, six instructors across engineering and science disciplines from three universities developed and implemented twelve data science modules, embedding data science concepts into disciplinary topics using real-world data (Table 1). These modules have been made publicly available, with eleven modules shared on the project website (ds4stem.org) and one module hosted on HydroLearn (hydrolearn.org), a public educational platform for earth sciences. The project directly engaged more than 1,200 students (based on the participating course rosters), with a significant proportion representing traditionally underrepresented groups in STEM disciplines.

The project's impact extends beyond module development and implementation and sharing. Students demonstrated significant improvement in their self-assessed understanding of data science topics covered after completing one or more modules, which aligned with instructors' assessments of student performance. Furthermore, analysis of student surveys revealed positive changes in their perceptions of data science across constructs of motivation, skills, interest, and confidence (Figure 1; see [17]). The project's findings have been disseminated through multiple channels, including two peer-reviewed conference papers ([11],

[16]), two full journal articles ([10], [17]), several conference presentations, and a workshop, contributing to the broader discourse on data science integration in STEM education.

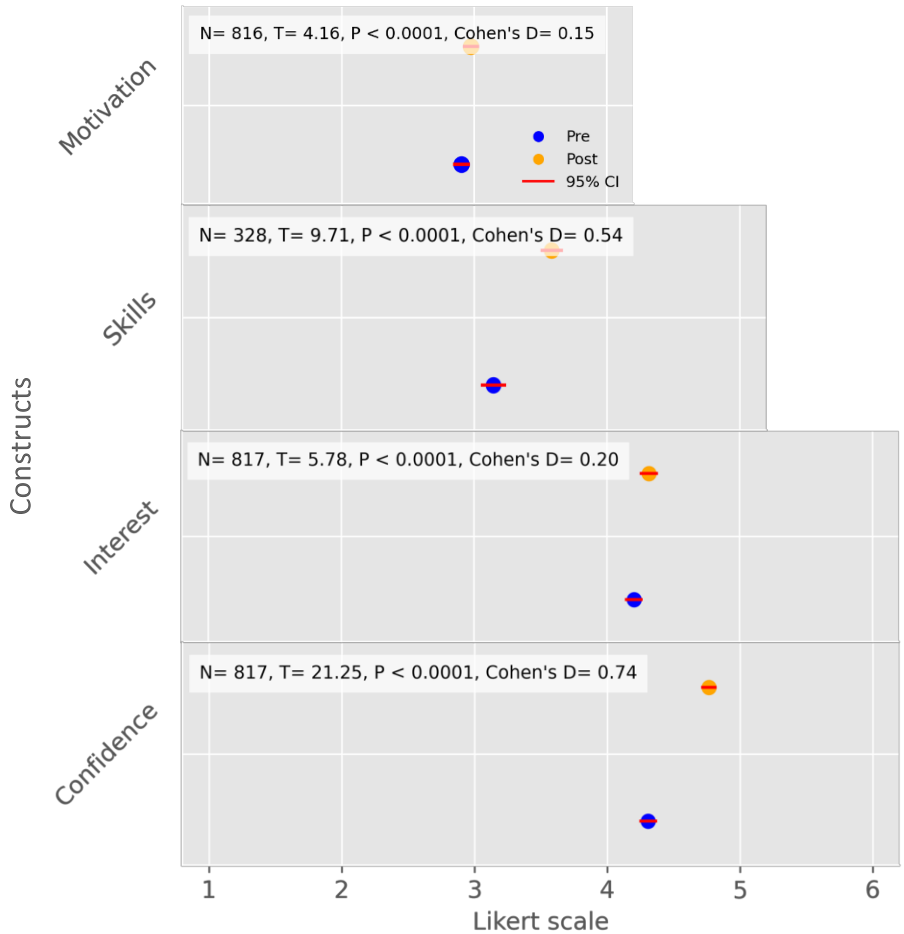

Figure 1. Pre-post survey results of student perception of data science. This figure presents a comparison of pre- and post-survey responses across four constructs: motivation, skills, interest, and confidence in data science. Paired t-tests revealed statistically significant improvements ($p < 0.0001$) in all constructs, with confidence showing the largest positive change. Blue and orange dots represent pre- and post-survey means respectively, with red horizontal lines indicating 95% confidence intervals. For detailed statistical analysis and methodology, refer to [17].

III. Key considerations for integrating data science across disciplines

Instructor autonomy: While common practice in data science education initiatives often involves separate teams developing learning modules apart from instructors (e.g., [18]), our findings reveal that instructors highly value having autonomy over content development when embedding data science into disciplinary contexts. This autonomy enabled them to tailor modules to specific course needs and constraints, particularly when the goal is infusing data science content into disciplinary topics using real-world data, rather than adding separate content alongside existing material. Instructors leveraged their deep understanding of course practicalities to optimally adjust the depth and breadth of integrated data science content, helping minimize difficulties with fitting new material into existing curricula. This autonomy facilitated integration that felt natural rather than forced, as instructors could seamlessly weave data science concepts into disciplinary topics such as flood frequency analysis in upper-level hydrology courses.

Instructor autonomy in content development may offer several additional benefits. Students engage more readily with the data science content when the content aligns directly with their disciplinary learning and career goals, rather than appearing as an alien addition to their coursework [6]. Instructors appreciated the flexibility they had to continuously adapt their content based on student feedback gathered through in-class surveys, enabling easier and iterative improvements to the modules. Furthermore, this autonomy potentially enhances the sustainability of integration efforts beyond the project's funding period – a crucial concern noted in the literature [1], [19]. When instructors directly embed data science into disciplinary content rather than treating it as an add-on to an already full curriculum, the integration is more likely to persist as a permanent course component.

While emphasizing instructor autonomy, we found that certain structural elements remained essential for successful integration. These included establishing common definitions for learning objectives and outcomes through consensus agreement, developing accurate assessments, and maintaining a shared framework for module development ([10], [11]). We also acknowledge that embedding data science content into disciplinary contexts requires varying levels of effort across disciplines, as some fields are traditionally less amenable to data-driven methods (e.g., STEM vs. social sciences disciplines). In such cases, instructors may initially benefit from additional support through data science integration initiatives. However, with the increasing availability of disciplinary-relevant data and analysis methods resulting from the growing acceptance of data-driven approaches across disciplines, such seamless integration of data science content will become increasingly easier.

Data science and disciplinary topic selection: Instructors strategically focused on teaching discipline-agnostic data science fundamentals, particularly data visualization and basic statistical analyses, that could be readily applied within disciplinary contexts (Figure 2). When selecting specific disciplinary topics for integration, instructors considered multiple factors: alignment with course syllabi, appropriateness for students' academic level, significance within the discipline, and availability of relevant data. For instance, in the upper-level civil engineering hydrology course from VT, the instructor integrated data science into flood and drought frequency analysis because it was a core learning outcome of the course, suited to multiple phases of the data science lifecycle, and had readily available public data from the United States Geological Survey website.

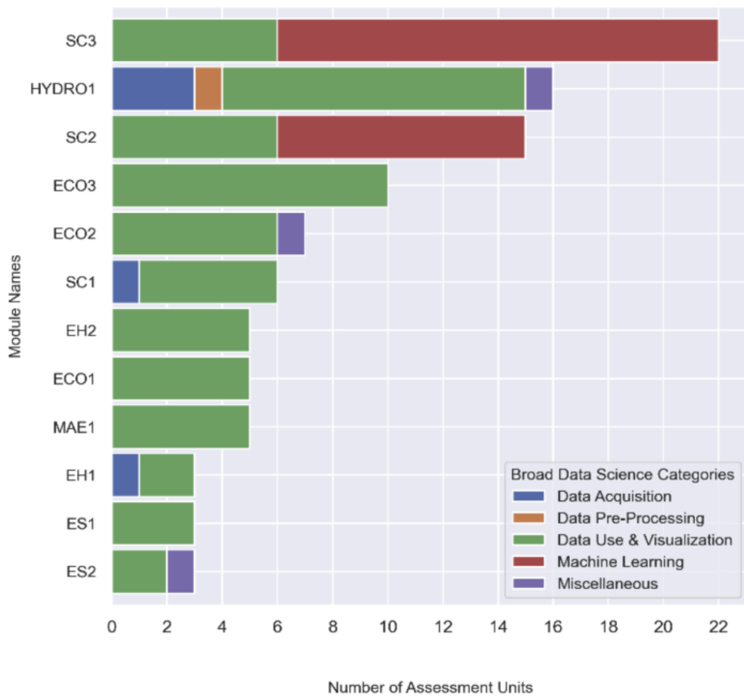

Figure 2. Distribution of data science topics across course modules. The figure displays the distribution of broad data science categories across 12 modules developed in six disciplinary courses (Table 1), showing the number of assessment units allocated to each category. Data Use & Visualization was the predominant category, accounting for 66% of all assessment units and appearing in every module. Machine Learning represented 25% of assessments, concentrated in the SC2 and SC3 modules. The remaining categories - Data Acquisition, Data Pre-Processing, and Miscellaneous - each comprised 5% or fewer of the total assessment units. This distribution suggests instructors favored discipline-agnostic topics focused on data use and visualization, which encompassed creating and interpreting visualizations and basic statistical analyses. For detailed information about the data science categories, their formulation method, and assessment unit calculations, refer to [10], [16]. Note: The module codes on the y-axis represent specific modules from different courses: SC1, SC2, and SC3 are modules from Smart Cities (SC); HYDRO1 is from Hydrology (HYDRO); ECO1, ECO2, and ECO3 are from Ecology (ECO); EH1 and EH2 are from Engineering Hydrology (EH); MAE1 is from Monitoring and Analysis of the Environment (MAE); and ES1 and ES2 are from Engineering Statistics (ES).

The use of real-world, discipline-relevant datasets proved particularly effective in increasing student engagement and interest, as evidenced by student responses to open-ended survey questions as well as instructors' perceptions of student engagement and interest. This finding aligns with established research showing that authentic data and real-world applications enhance student motivation and learning outcomes in STEM education [20], [21]. When students work with actual disciplinary data rather than constructed examples, they better understand the relevance of data science to their future careers and develop more realistic expectations about data analysis challenges they might encounter professionally.

However, instructors faced a common pedagogical challenge in balancing breadth versus depth of topic coverage. This tension emerged particularly when deciding between teaching more data science skills thoroughly (depth) versus covering more disciplinary content (breadth). This challenge reflects a well-documented issue in STEM education where instructors must carefully weigh the trade-offs between comprehensive coverage and deep understanding [22]. The challenge becomes particularly acute in data science integration because students need sufficient time to develop practical competency with new analytical tools while still mastering core disciplinary concepts. Instructors addressed this by carefully selecting specific data science skills that most directly supported their disciplinary learning objectives, rather than attempting to cover all aspects of data science as evidenced by the popular choice of Data Use & Visualization topic category (Figure 2).

Early focus on student motivation: Our analysis of students' self-reported motivation, interest, and confidence revealed that their initial perceptions were the strongest predictors of their final perceptions after completing a data science module (Figure 3). This suggests that inspiring students' early motivation, interest, and confidence in data science can significantly influence their engagement throughout the course. Student responses to open-ended questions emphasized the value of seeing direct connections between data science skills and their future careers, indicating that real-world applications resonated strongly with them. These findings have important practical implications for instructors integrating data science into their courses. Based on these results, we recommend that instructors begin their data science integration with explicit discussions about the relevance of data science to students' disciplines and future careers. Such discussions can help establish the foundational motivation that appears crucial for sustained engagement. This approach aligns with previous research demonstrating that pretreatment motivation significantly impacts learning across different contexts [23], [24]. The timing of these motivational discussions appears particularly important, as our data shows that early perceptions strongly influence how students engage with and perceive data science throughout the rest of the course.

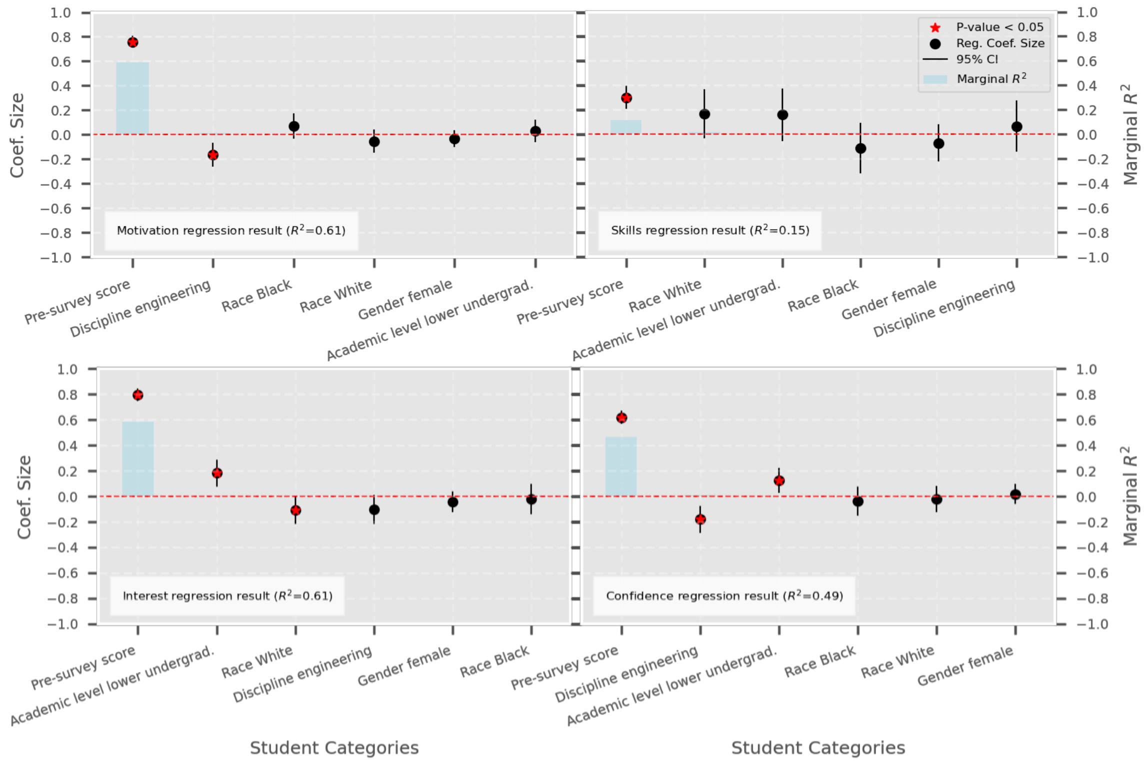

Figure 3. Summary result of regression models for each construct. The left y-axes for each plot represent the coefficient size and the right y-axes represent the marginal coefficient of determination (R^2) (ranging from 0 to 1) which indicates the relative contribution of each independent variable to the unadjusted R^2 for the entire model. The x-axes indicate student groupings (i.e., independent variables) included in each model (which are the same across all four models) and are ordered based on their respective coefficient absolute values which differ across models based on regression results. The reference categories are academic level upper undergraduate (for the included category of academic level lower undergraduate), discipline science (for the included category of discipline engineering), race other (for the included racial identities of White, and Black), and gender male (for the included gender identity of female). The red stars on the coefficients indicate statistical significance at 5% significance level. For detailed statistical analysis and methodology, refer to [17].

Scaffolding needs: Instructors acknowledged the challenge of addressing varying student backgrounds in data science, particularly in courses drawing students from different disciplines and academic levels. This concern was echoed in students' responses to open-ended survey questions (Figure 4), where they expressed apprehension about differing prior experiences with data science, pace of instruction, and learning curves associated with various tools. To address these varying backgrounds, instructors implemented multiple support strategies. For example, they developed comprehensive tutorials and encouraged both in-class and online peer learning through discussion boards where more experienced students could help others. Group work proved particularly effective, with instructors designing problem-based learning activities that allowed students to collaborate and learn from peers with different expertise levels. Instructors also provided flexibility in tool selection, allowing students to choose between familiar tools like Excel and more advanced options, while ensuring adequate tutorial support for those ready to tackle more sophisticated approaches. Additional support included providing reference materials and centralized websites (such as the one in this project, ds4stem.org) containing data background information, data science glossaries, and tutorial videos. Student feedback indicated strong appreciation for these scaffolding approaches, particularly the tutorials and group work strategies (Figure 4).

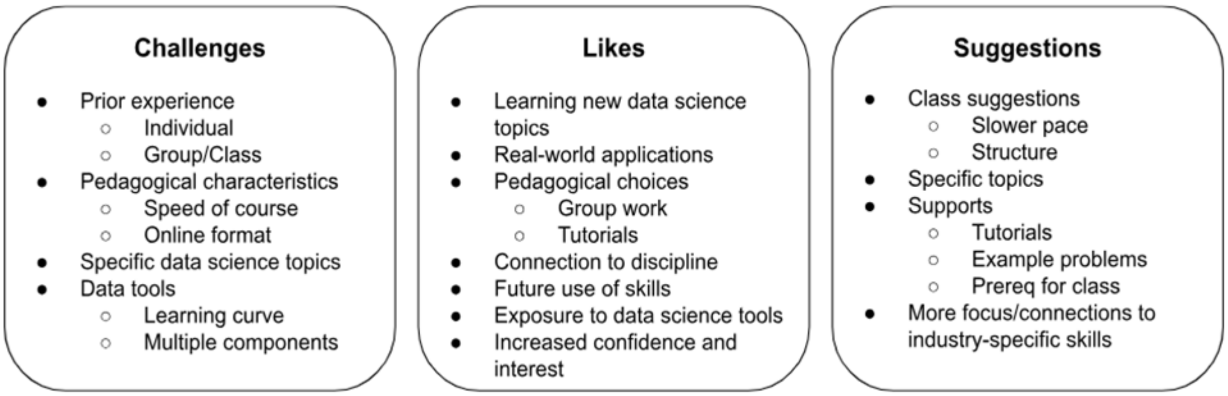

Figure 4. Summary of student open-ended survey responses regarding data science integration. The figure summarizes key themes from student open-ended survey responses regarding their experiences with data science integration in their courses, organized into three categories: challenges, likes, and suggestions. Challenges predominantly centered around prior experience levels, pedagogical aspects, specific data science topics, and use of data tools. Students' likes focused on learning new concepts, real-world applications, pedagogical approaches, and connections to their disciplines. Suggestions primarily addressed class structure improvements, topic-specific adjustments, and requests for additional learning support. For a detailed analysis of student responses and comprehensive findings, refer to [17].

Instructors identified timely and constructive feedback as another crucial scaffolding element. They implemented various feedback mechanisms, including just-in-time feedback during in-class group work sessions, where students could receive immediate guidance from both peers and instructors. For larger classes where individual attention was challenging, instructors utilized self-paced modular tests with automated feedback and in-class polling platforms (such as Microsoft Forms and Poll Everywhere) to gauge understanding and provide immediate clarification. They also emphasized the importance of following up after graded assignments and sharing detailed rubrics to help students understand performance expectations and areas for improvement.

Beyond student-focused scaffolding, instructors emphasized the need for support in developing clear learning objectives, assessments, and rubrics during module development. They advocated for a gradual approach to implementing new data science tools and concepts, starting with familiar topics while progressively incorporating more advanced elements. While teaching assistants could help develop tutorials for new tools, instructors noted that time constraints often influenced their tool selection, particularly when considering whether they could effectively teach complex programming tools to students with limited prior exposure. This highlighted the importance of balancing pedagogical ambition with practical time limitations when integrating new data science elements into existing courses.

Integration into disciplinary context: Instructors reported achieving a relatively smooth integration of data science content into their disciplinary courses, attributing this success primarily to the quantitative nature of their STEM disciplines and their existing pedagogical approaches that emphasized active learning through interactive, problem-based exercises. A key advantage of their approach was the ability to maintain most existing disciplinary content, as data science concepts were embedded within rather than added alongside disciplinary topics. This stands in contrast to similar initiatives where instructors struggled with removing disciplinary content to accommodate additional data science materials [25].

The instructors emphasized that their goal was not to sacrifice disciplinary content but rather to enhance it through data-driven skill development. They focused on integrating data science principles into the teaching of fundamental disciplinary concepts, carefully selecting topics that naturally lend themselves to data analysis and visualization. While acknowledging that not all disciplinary topics are equally amenable to data science integration, instructors strategically identified opportunities to weave data science principles into sections that already involved data analysis or quantitative reasoning.

Even in courses with constraints around covering broad disciplinary content and less emphasis on skills development, instructors suggested creative ways to incorporate data science concepts. Their suggested solutions included integrating data science perspectives into content delivery, such as using data visualizations to explain concepts or incorporating simulation-based examples into lectures. This approach would allow exposing students to data science thinking without significantly altering such courses' primary focus on disciplinary content.

IV. Graduate research assistant perspectives

GRAs, representing diverse disciplines such as civil and environmental engineering, computer science, environmental science, and biology, appreciated the opportunity to collaborate with faculty across diverse STEM disciplines. They believe their collaboration enhanced their ability to communicate effectively across disciplinary boundaries. They think this experience proves especially valuable for their professional development, as the contemporary workforce requirements demand increased emphasis on interdisciplinary collaboration across industries.

The GRAs noted that these cross-disciplinary communication skills would likely benefit their future careers, whether in academia or industry.

The GRAs emphasized the importance of project management approaches in facilitating successful outcomes. They identified several key elements that contributed to effective project execution: the establishment of clear multi-scale goals (monthly, semestral, and annual) aligned with broader project objectives, regular meetings among GRAs and with instructors and project PIs to assess progress, and clear communication of expectations and deliverables. The assignment of specific GRAs to individual courses proved particularly effective, creating clear lines of responsibility and enabling detailed oversight of course-specific implementations. These efforts facilitated efficient project management while ensuring consistent support for instructors throughout the module development and implementation process.

V. Future work for instructors and the project

The project team recognizes the transformative potential of LLMs in enhancing data science education and plans several expansions incorporating these tools. Both students and instructors have identified a significant need for personalized learning experiences due to varying levels of data science expertise and different learning pace requirements among students. Instructors believe LLMs can help address these challenges by providing customized support for concept understanding and a smooth introduction to data analysis tools such as coding, particularly for students with limited prior exposure to data science. However, instructors emphasize the importance of treating LLMs as assistive tools rather than authoritative sources, encouraging students to maintain critical thinking and responsibility for their learning outcomes. Beyond data science integration, instructors are exploring innovative applications of LLMs in their disciplinary teaching. For example, one instructor at VT has developed an interactive assignment using custom GPTs (a feature accessible through the ChatGPT Plus subscription) to simulate stakeholder engagement in water resource management challenges, enabling students to gain practical experience in understanding and integrating diverse perspectives in environmental decision-making. Additionally, the research group of another investigator is developing a specialized LLM-based chatbot that will enable students to interact dynamically with module content while enriching existing materials through curated web-based resources, supporting both learning and content enhancement objectives.

The project team continues their commitment to disseminating best practices and fostering broader adoption of data science integration approaches. Following the success of the project's workshop titled "Integrating Data Science Modules in Engineering and Science Courses" at the 2023 ASEE Annual Conference (June, Baltimore), which aimed to share effective strategies for embedding data science content within disciplinary contexts, a second dissemination workshop is planned for 2025. This workshop, to be hosted at NC A&T, an

HBCU, will specifically target instructors from civil & environmental engineering, computer science, environmental science, industrial and systems engineering, and biological systems engineering faculty, furthering the project's goal of expanding data science education opportunities across diverse institutional contexts. The team also plans to develop a follow-up NSF proposal for integrating AI-related concepts into the courses.

VI. Summary

With the rise of data deluge across all sectors of society, there is an increasing need for data-literate graduates across disciplines and sectors of the economy, particularly in STEM fields [1]. While educational institutions have begun efforts to meet this demand [26], effective pedagogical approaches for developing data literacy remain an area of active investigation. While there are many different approaches to incorporating data-literacy learning objectives within existing disciplines, a modular approach offers a flexible and effective method [27]. This project set out to develop and implement an interdisciplinary approach for integrating data science expertise across undergraduate STEM curricula through a multi-university RPP. The implementation involved six instructors from three different universities, one public (VT), another private (VU), and the third an HBCU (NC A&T), developing and deploying twelve discipline-specific data science modules, directly engaging over 1,200 students, with significant representation from traditionally underrepresented groups in STEM. The project demonstrated considerable success, evidenced by significant improvements in students' self-reported understanding of data science concepts and positive shifts in their perceptions across motivation, skills, interest, and confidence. The project's impact extends beyond direct implementation through multiple peer-reviewed publications, conference presentations, publicly available educational modules, and workshops for instructors. We have distilled the key implementation findings into five critical considerations for successful data science integration: instructor autonomy in content development and adaptation, strategic selection of data science and disciplinary topics, early emphasis on student motivation, comprehensive scaffolding approaches for diverse student backgrounds, and seamless integration within disciplinary contexts. These considerations are intended to facilitate effective data-literacy module development and implementation for instructors and departmental leaders seeking to infuse data-literacy learning objectives within their disciplines. Looking forward, the project continues to evolve through exploring innovative applications of LLMs for personalized learning experiences and expanding its reach through targeted workshops for interdisciplinary instructors.

Acknowledgments: This research is supported by NSF-IUSE grants #1915538, #1915487, and #1915268. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] E. National Academies of Sciences and Medicine, *Data Science for Undergraduates: Opportunities and Options*. Washington, DC: The National Academies Press, 2018. doi: 10.17226/25104.

[2] G. Karaali, "Artificial Intelligence, Basic Skills, and Quantitative Literacy," *Numeracy*, vol. 16, no. 1, Jan. 2023, doi: https://doi.org/10.5038/1936-4660.16.1.1438.

[3] H.-W. Choi and S. Abdirayimov, "J Multimed Inf Syst: Revolutionizing Education: The Era of Large Language Models," 2024, doi: 10.33851/JMIS.2024.11.1.97.

[4] S. Hester, S. Buxner, L. Elfring, and L. Nagy, "Integrating quantitative thinking into an introductory biology course improves students' mathematical reasoning in biological contexts," *CBE—Life Sci. Educ.*, vol. 13, no. 1, pp. 54–64, 2014.

[5] J. L. Klug, C. C. Carey, D. C. Richardson, and R. Darner Gougis, "Analysis of high-frequency and long-term data in undergraduate ecology classes improves quantitative literacy," *Ecosphere*, vol. 8, no. 3, p. e01733, 2017.

[6] A. Bundy, "Australian and New Zealand Information Literacy Framework," Australian and New Zealand Institute for Information Literacy, Adelaide, 2004.

[7] D. Deb, M. Fuad, and K. Irwin, "A Module-based Approach to Teaching Big data and Cloud Computing Topics at CS Undergraduate Level," in *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*, ACM, 2019. doi: 10.1145/3287324.3287494.

[8] K. Hunt, "The challenges of integrating data literacy into the curriculum in an undergraduate institution," *IASSIST Q.*, vol. 28, no. 2–3, p. 12, 2005.

[9] A. Janiak and R. Rudek, "Experience-based approach to scheduling problems with the learning effect," *IEEE Trans. Syst. Man Cybern.-Part Syst. Hum.*, vol. 39, no. 2, pp. 344–357, 2009.

[10] Md. Y. Naseri et al., "Integrating Data Science Into Undergraduate Science and Engineering Courses: Lessons Learned by Instructors in a Multiuniversity Research–Practice Partnership," *IEEE Trans. Educ.*, pp. 1–12, 2024, doi: 10.1109/TE.2024.3436041.

[11] C. Snyder et al., "Understanding Data Science Instruction in Multiple STEM Disciplines," in *2021 ASEE Virtual Annual Conference Content Access*, 2021.

[12] C. Dichev and D. Dicheva, "Towards Data Science Literacy," *Procedia Comput. Sci.*, vol. 108, pp. 2151–2160, 2017, doi: 10.1016/j.procs.2017.05.240.

[13] T. Donoghue, B. Voytek, and S. E. Ellis, "Teaching Creative and Practical Data Science at Scale," *J. Stat. Data Sci. Educ.*, vol. 29, no. sup1, pp. S27–S39, 2021, doi: 10.1080/10691898.2020.1860725.

[14] D. Schuff, "Data science for all: a university-wide course in data literacy," in *Analytics and Data Science*, Springer, 2018, pp. 281–297.

[15] C. E. Coburn and W. R. Penuel, "Research–Practice Partnerships in Education: Outcomes, Dynamics, and Open Questions," *Educ. Res.*, vol. 45, no. 1, pp. 48–54, Jan. 2016, doi: 10.3102/0013189X16631750.

[16] M. Y. Naseri et al., "A modular approach for integrating data science concepts into multiple undergraduate STEM+ C courses," presented at the 2022 American Society Engineering Education Annual Meeting, 2022.

[17] M. Y. Naseri et al., “Integrating Data Science into Undergraduate Science and Engineering Courses through Discipline-specific Modules: The Students’ Perspective,” 2025, *Edarxiv*. [Online]. Available: https://osf.io/preprints/edarxiv

[18] K. A. Dill-Mcfarland et al., “An integrated, modular approach to data science education in microbiology,” *PLOS Comput. Biol.*, vol. 17, no. 2, p. e1008661, 2021, doi: 10.1371/journal.pcbi.1008661.

[19] J. Bonnell, M. Ogihara, Y. Yesha, and I. Bojanova, “Challenges and Issues in Data Science Education,” *Computer*, vol. 55, no. 2, pp. 63–66, 2022, doi: 10.1109/mc.2021.3128734.

[20] S. Kross and P. J. Guo, “Practitioners teaching data science in industry and academia: Expectations, workflows, and challenges,” in *Proceedings of the 2019 CHI conference on human factors in computing systems*, 2019, pp. 1–14.

[21] A. A. McNamara, *Bridging the gap between tools for learning and for doing statistics*. University of California, Los Angeles, 2015.

[22] D. Coil, M. P. Wenderoth, M. Cunningham, and C. Dirks, “Teaching the Process of Science: Faculty Perceptions and an Effective Methodology,” *CBE—Life Sci. Educ.*, vol. 9, no. 4, pp. 524–535, Dec. 2010, doi: 10.1187/cbe.10-01-0005.

[23] J. M. Hamm, R. P. Perry, J. G. Chipperfield, P. C. Parker, and J. Heckhausen, “A motivation treatment to enhance goal engagement in online learning environments: Assisting failure-prone college students with low optimism,” *Motiv. Sci.*, vol. 5, no. 2, pp. 116–134, 2019, doi: 10.1037/mot0000107.

[24] M. R. Hasannejad, M. Zoghi, and H. D. Asl, “Motivation, tasks, attitudes: The influence of motivational pre-task strategies on tasks performance and tasks engagement,” *J. Lang. Learn. Teach.*, vol. 7, no. 2, pp. 113–128, 2017.

[25] M. Haynes, J. Groen, E. Sturzinger, D. Zhu, J. Shafer, and T. Mcgee, “Integrating Data Science into a General Education Information Technology Course,” in *Proceedings of the 20th Annual SIG Conference on Information Technology Education*, ACM, 2019. doi: 10.1145/3349266.3351417.

[26] V. Merwade and B. L. Ruddell, “Moving university hydrology education forward with community-based geoinformatics, data and modeling resources,” 2012, Accessed: Mar. 10, 2025. [Online]. Available: https://hess.copernicus.org/articles/16/2393/2012/

[27] A. Bundy, “Australian and New Zealand Information Literacy Framework,” Australian and New Zealand Institute for Information Literacy, Adelaide, 2004.