The Internal Internship: Enabling Novel Opportunities for Undergraduate Data Science Experiential Education.

Jessica Liebowitz*
Timothy Hickey*
jkl@brandeis.edu
tjhickey@brandeis.edu
Brandeis University
Waltham, Massachusetts, USA

ABSTRACT

In this paper we reflect on what we have learned in the past three years creating a novel experiential education program in data science for undergraduates. In this Data Science Internal Internship (DSII), students are paid as employees of the university office by which they are tasked with helping to overcome administrative challenges through the introduction of data scientific solutions. Interns meet weekly with the DSII's program directors in the computer science department, receiving feedback on their design and implementation strategies, technical support as necessary, and guidance on organizational dynamics they encounter along the way. Similar to traditional internships, this offers students the opportunity to acquire data scientific expertise by working with a mentor in a professional setting. A key difference is that the mentor is the actual supervising client, typically with minimal data science expertise. This drives the uniqueness of the DSII's learning environment, which includes the guarantee to students that they will have access to the data they need to address the administrative challenge in question. It also grants students regular and direct contact with the supervising domain knowledge experts. This results in direct benefit to the university while providing students the unusual chance to serve as lead developers tackling high-value problems.

We explore the Internal Internship approach as a proof-of-concept, describing the infrastructure that will help make possible its replicability. The Internal Internship offers undergraduates the chance for demonstrable achievement through its expectations of high motivation and creativity in self-teaching, and of effective application of new technologies. These expectations emerge from real administrative challenges requiring innovative solutions, and through direct and dedicated relationships with respective administrators. Because of the distinctive learning environment it offers, we find that the Internal Internship concept is worth pursuing as a complement to the many new data science programs spawned over the past decade.

*Both authors contributed equally to this research.

SIGCSE Virtual 2024, December 5–8, 2024, Virtual Event, NC, USA

ACM ISBN 979-8-4007-0598-4/24/12
https://doi.org/10.1145/3649165.3690114

CCS CONCEPTS

• **Applied computing** → **Education**; • **Social and professional topics** → **Computer science education**.

KEYWORDS

Data Science, Experiential Learning, Internships, Higher Education Administration

ACM Reference Format:
Jessica Liebowitz and Timothy Hickey. 2024. The Internal Internship: Enabling Novel Opportunities for Undergraduate Data Science Experiential Education.. In *Proceedings of the 2024 ACM Virtual Global Computing Education Conference V. 1 (SIGCSE Virtual 2024), December 5–8, 2024, Virtual Event, NC, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3649165.3690114

1 INTRODUCTION

In our data analytic research into the higher education workforce, we discovered an untapped opportunity to create a new type of experiential learning program for data science students. This opportunity emerged from the finding, based on exploring job postings in the higher education sector, that research universities were among the fastest growing drivers of supply and demand for data scientific talent in the decade leading up to the COVID pandemic [10]. At the same time that more and more faculty were being called upon to bring data scientific skills into the classroom, fast-growing numbers of administrative job postings were calling for them too. This finding identified the potential value of bridging together the academic and the administrative functions of the university through a shared focus on data science.

To explore this potential, we launched a special internship program in 2021, in which top performing undergraduates are paired with administrators to apply data scientific skills to university administration and operations. This paper describes three years of experience with this Data Science Internal Internship (DSII), identifying distinctive learning experiences for participating undergraduates as well as specifying the program design characteristics that are responsible for that distinctiveness. We believe that the pilot program we designed, at an R1 AAU school, can serve a proof-of-concept role for other universities focused on building out multiple components of best-in-class data science education.

Deepening mastery and acquiring practical expertise through experiential learning is a key value for undergraduates that the Internal Internship program shares with other more established data science internship programs [3]. Experiential learning opportunities in technology-application fields have been growing in recent

years, thanks in part to the demand from employers for graduates who can "hit the ground running" [8]. They've been growing, also, thanks to the nature of data science itself, which is "necessarily highly experiential" [6]. Data science is both "a practiced art and a developed skill." As such, students need to "encounter frequent project-based, real-world applications with real data to complement the foundational algorithms and models" [6]. These insights about the pedagogical demands of technological learning are reflected in many educational subfields, including challenge-based learning and service learning [7]. The Internal Internship is consistent with this broadening response to the significance of experiential learning opportunities, which make it possible for students to "solidify their data skills in a way that a typical course cannot" [9].

The unorthodox nature of the pilot we designed lies in its Internal Internship structure, which enables undergraduates to take on mission-critical responsibilities within administrative offices, where dedicated supervisors expect students to help deploy effective data scientific solutions to work flow challenges. The Internal Internship is not a course, or a curriculum, and it is not a student organization. It is a system built on relationships among computer science faculty and university leaders for selecting trustworthy students, identifying data science projects with well-defined goals relevant to improving university operations, matching the two, and then overseeing the technical, security, and organizational support necessary to help each project arrive at a positive contribution to university administration.

To our knowledge, the infrastructure necessary to enable experiential learning for data science students in university operations has not yet been openly discussed or explored. In the first phase of our pilot program, which we report on here, we focused on building the infrastructure of relationships and expectations to ensure proper preparation, oversight, security, trust, and effective outcomes. With the essential qualities of this infrastructure now identified and in place, the DSII program will be able to move into its next phase, developing research findings that can be shared publicly regarding student learning, administrator learning, and metrics to evaluate operational impact. We elaborate now on the evidence of feasibility from our experience building the internal internship infrastructure for data science applications in higher education.

2 KEY COMPONENTS OF THE PROGRAM

The design of the Internal Internship is driven by its insistence on generating practical value for administrators participating in the program. This includes the way in which the Interns are selected, the projects defined and evaluated, and the oversight of student work progress conducted. It also helps explain why participating students are paid as employees rather than granted course credit: administrative mentors explicitly requested a paid, non-credit-bearing set-up for this program. Their experience with multiple forms of student engagement generated the strong conviction that students are much more likely to succeed in dedicating their intelligence, creativity, and resourcefulness to accomplishing office goals when they are not bound by course requirements.

2.1 Project Definition, Intern Selection, and Student Employment

DSII projects are defined by office managers identifying work-flow challenges in their divisions that would be well-served by data scientific applications. Many of the projects have focused on operational data that has great value to decision-makers but remains untapped, others reflect time budgets burdened by manual data analysis, some aim to leverage data scientific techniques to extract key pieces of information from voluminous types and numbers of documents, or to address increasing pressure from multiple stakeholders for trustworthy and interactive access to work-flow indicators. We work with administrators to help craft clear job descriptions for each project, but the scope, and the measure of success for each project depends on the operational office in question.

The supervising administrators were originally motivated by the university's senior leadership, as the University Provost and the Executive Vice President for Administration and Finance encouraged their direct reports and respective divisions to give it a try. Once the DSII began developing a track record of success, administrators became much more willing to consider the possibility of bringing undergraduates into their offices to help solve operational challenges through data scientific innovations.

The student selection process begins with a committee made up of computer science department faculty who review candidates' academic performance in foundational courses of the computer science major, in order to select finalist candidates for each open position. Faculty seek evidence of mastery of data structures and Python, and also, increasingly, of the basics of natural language processing, data visualization, and machine learning. Faculty also review, where applicable, evidence of relevant technical expertise developed through previous jobs and internships. But it's the supervising administrators themselves who interview finalist candidates and make their own hiring decisions based on who they believe will be best suited to accomplishing the project outcome in question.

Our Interns are equally paid student employees, working at least 10 hours per week, and reporting directly to the supervisor responsible for implementation of work-flow improvements. Interns usually work through two sessions – Fall, Spring, or Summer – and often work through a third session to help on-board new Interns, or to help automate their innovations. The Internal Internship does not offer students the opportunity to fulfill curricular requirements. In exchange, it offers well-mentored opportunities to design and implement solutions, avoiding the common disconnect in experiential learning programs between the classroom on the one hand and the business outcome on the other [1, 4].

2.2 Access to Data, Domain Knowledge Mentoring, and Weekly Cohort Meeting

Participating administrators commit to granting their Interns full access to all the data they will need in order to accomplish the stated goals as defined by the administrators themselves. This is unusual. Data for experiential learning purposes is often constrained by data security restrictions, legal barriers to cross-institutional collaboration, and pre-defined limits on the size and scope of data access [1, 4, 8]. Indeed, some data science faculty have made access to comprehensive data sets of "real" data a focus of their work,

since "it is difficult to find comprehensive case studies to support data science courses, especially cases that span from raw data to actual business outcomes" [5].

Our administrative partners are willing to take the unusual step of granting Interns data access because of the trust the program engenders. This trust stems from three sources: (a) explicit support and encouragement for the DSII from senior university leaders, including the Provost and the EVP for Finance and Administration, (b) oversight of progress for each student that we exercise through weekly meetings with the cohort, and (c) partnership that we've built with the university's data governance board as well as with the network security leaders within the university's central IT function[1]. This partnership includes continual training in the university's data security standards, and has led to the development of additional protocols and Non-Disclosure-Agreements specifically relevant to student workers and participating faculty.

Interns are expected to acquire deep domain knowledge quickly about their specific projects, and the DSII guarantees that students will have regular access to domain knowledge experts to help them do so. The working relationship between student and administrator is not inter-mediated by faculty or other educational experts providing guidance. This is unusual from the point of view of many data-focused internship or capstone experiential learning opportunities, which will often rely on the faculty supervisor to ensure effective solutions for those clients who "expect professional caliber" expertise [11]. We know from experience that our high expectations of students are justified: selecting for students who demonstrate mastery of data structures, coding, and related computational topics, has so far continued to generate a reliable, albeit small, pool of candidates capable of the self-teaching necessary to accomplish the operational goals defined by participating administrators.

The required weekly meeting of the Intern cohort oversees the individual progress of each student in achieving office-specific goals. We manage this meeting from an agile/scrum perspective: Interns learn to share crisp presentations each week about what they accomplished, what they aim to accomplish next week, what they faced as their greatest obstacle, and what they found particularly noteworthy, counter-intuitive, or intellectually compelling. Each Intern is required to present the weekly progress of their work through visualizations of key findings and essential domain knowledge. In this way, we ensure that we are hearing early of potential derailments – be they technical snags, or challenges in matters of communication or interpersonal interactions – which gives us the chance to nip them in the bud before they get serious enough to threaten the possibility of effective outcomes.

We bring faculty and administrators with relevant expertise or experience to give feedback, and to provide extra technical guidance on subfield-specific challenges. We encourage Interns to learn from each other as they discover they are often teaching themselves similar tools and techniques. We talk through situations where students are surprised to encounter organizational obstacles to technological effectiveness. We provide technological support. We also guide students in transforming their weekly pitches into refined end-of-session presentations about the impact of their work to an internal, cross-divisional audience of university leaders.

[1]As attested by the university's Deputy CIO, program interns are "helping to advance the Data Governance program, adhering to its protocols to ensure all data shared is accurate, accessible, complete, and secured [13].

3 PEDAGOGICAL STRUCTURE

3.1 Pairing of Student-Administrator

In the Internal Internship, students are in the unusual situation of being accountable for the relevance and impact of their applications. This accountability, which comes from the Intern's direct connection with the administrator, leads students to develop self-awareness about their lack of knowledge in the face of particular operational challenges. This self-awareness motivates students to figure out how to teach themselves the skills of technology application necessary to address the problem. It also motivates them to take advantage of the opportunity to work directly with their supervisors, acquiring the domain knowledge without which it would be impossible for them to evaluate the appropriateness of any proposed innovations. Students are required to learn, over time, how to translate details about their technological solutions into clear explanations of the value of their innovations. Students are expected not only to communicate what they are doing, but also to explain, teach, and provide compelling evidence to their supervisors of relevant impact. Otherwise, the administrator would not be approving the student's proposed innovations in work-flow.

In our experience, it is this direct relationship between student and administrator that drives the distinctiveness of the DSII learning environment. It is the source of the guarantee to participating students that they will receive full access to the office data necessary to create effective work-flow innovations. It is also the source of the intense motivation we see in DSII Interns to climb whichever learning curves they might encounter in order to create working solutions to their supervisors' administrative challenges.

3.2 Uncertainty

In the classroom, professors who create the assignments know the answers to every problem. Data sets in classroom projects must be clean enough and familiar enough to the professor so that students manipulating them can demonstrate their skills. This is often true for experiential learning programs or graded externships or capstone projects for undergraduates, where data sets need to be "pre-curated" and "clean" enough so that students can be guaranteed to show some degree of impact [14]. The DSII program places students in high-uncertainty situations where no one knows whether and how a particular data set might be made to yield value, nor whether a new set of data scientific techniques or tools might prove use-able or useful. This uncertainty – married to an infrastructure of oversight by computer science department DSII program leaders shoring up protections against the consequences of potential failure, and tied to participating administrators expecting effective outcomes – is the essence of the pedagogical structure of the program.

3.3 Complexity of technical knowledge

We have had colleagues ask why we don't target this program toward Masters students who would have more experience with data scientific tools and techniques, rather than undergraduates who end up teaching themselves these techniques as part of the

DSII experience. The main reason is that, so far, we've learned that university offices eager for data scientific innovations in their workflow do not require the most sophisticated of data science tools. The most accomplished of our bright undergraduates are fully capable of meeting the challenges that arise in university operations. They are also at a skill level where they can benefit greatly from the experience: one of high-intensity and steep learning curves on many fronts, but also one that is deeply gratifying because it is achievable and relevant.

We have found that the first few months of a DSII project are often devoted to developing clean, machine-readable data sets. At the same time, Interns are quickly learning the essential domain knowledge to be able to effectively collaborate with their administrative mentors. For example, one Intern-Admin team began with data cleaning, moved to visualizations of trends and disaggregation within the data relevant to office-specific metrics, launched a secure and user-friendly dashboard for interactive access to insights from the data, created a customized encoder to ensure anonymity of sensitive data components, was able to utilize those sensitive data components to generate new statistical analyses, and is now building on the results of those analyses to enable appropriate exploration of predictive analytics. These types of applications do not require graduate-level training. But they do require technical mastery of basics, confidence in self-teaching and in translating insights across domains, and creativity to help transform new technology applications into effective outcomes.

4 PRELIMINARY RESULTS

4.1 Data scientific innovations at work

The DSII program has just completed its 10th cohort, with 3 cohorts per year (Spring, Summer, Fall) since 2021, averaging 7 students per cohort, 6 projects per cohort, with 58 percent of the Intern spots held by male and 42 percent held by female undergraduates, totaling 31 individual students completing at least one session. The DSII Interns are leaving a legacy of success, generating reports from participating administrators of actual improvements in terms of data-informed decision-making, time-savings, more efficient use of existing analytical resources, and insights from operational data that would not otherwise have been analyzed.

For example, the Associate Provost for Research Administration, overseeing the building of an interactive dashboard capable of tracking and analyzing trends in all grant applications for sponsored research, described her office's experience with the DSII as follows: "Our office has been able to leap-frog in our understanding and use of our data, and has given us the ability to continually ask new questions of the data to help drive institutional decision-making" [13].

Similarly, the Executive Director of Prospect Development and Information Strategy for the fundraising division, overseeing the design of predictive models customized to the university's alumni donors, explained, "As we teach the students about fundraising data, they learn why the data is valuable and relevant, and then they implement their new analytical techniques, which in turn leads us to new insights too, not just about what our data is telling us, but also about how worthwhile it is to acquire the tools necessary for new ways of looking at it" [13].

We have been learning important lessons in the pilot phase. For example, early on in the program, two students chose to leave before the end of their respective sessions. One of those students decided to opt for an academically-focused research assistant position in a faculty member's lab instead. The other, similarly, expressed lack of interest in pursuing solutions to administrative challenges. We were surprised by these two students since the selection process includes interviews, run by the hiring managers themselves, which involve detailed conversations about project objectives. We have since redoubled our efforts to make clear to all candidates that they must be willing and eager to effectively follow the direction of their administrative supervisors if they are to take on the responsibilities of a DSII Intern. We understand that this aspect of the program is not a common experience for undergraduates, especially for those who have not yet had extensive experience in working in an office environment. We have not experienced the same challenge since.

On the administrative mentor side, we've learned to be more explicit about the requirements of effective project definition. For example, one of our early projects was motivated by a senior administrator who was needing data analytic support for a specific policy change. After extensive data engineering and statistical analysis conducted by the DSII Intern, it turned out that the validated results did not support the policy change that the administrator had assumed necessary. This "dog didn't bark" outcome was useful, but not exactly the most efficient way for the DSII program to impact operational effectiveness. We have not experienced this shortcoming again, since we are now more precise: DSII projects are expected to deploy data scientific innovations to create better access to data, and to rely on that access to illuminate, automate, and help with prediction for data-informed decision-making, not to assume or aim to confirm a predefined policy answer.

Participating offices have included Human Resources, Institutional Research, Academic Affairs and Registrar, Dean of Arts and Sciences, Graduate School of Arts and Sciences, Division of Student Affairs, Accounting and Financial Services, Information and Technology Services, Office of Technology Licensing, Institutional Advancement, Office of Research Administration, Facilities Administration, Transportation, and the Office of Investment Management.

4.2 Sustainability and Replicability

Our partnership with the network security leaders of the university, combined with a track record of three years of experience, is now making it possible to build a formal DSII Repository within central IT. This will streamline the administrative requirements of on-boarding and off-boarding of students through the program, automate associated security protocols, help ensure sustainability of on-going projects, as well as replicability of student-designed solutions. The DSII Intern dedicated to this project is now helping to assemble a system for secure storage and dynamic access to all the work created by DSII Interns, on a project-by-project basis. It will also include a public-facing component, enabling the sharing of student work and administrative impact in a way that is consistent with university security protocols.

We can share details from the DSII project in student affairs for a brief example of impact. In that project, administrators were

finding it difficult to leverage the value of the important national survey on student wellness they had been conducting on campus through the National College Health Assessment (NCHA). The resulting NCHA data was being delivered through SPSS, a statistical software package introduced over 50 years ago which lacks the capability to support quick access to key insights or to engage in interactive analytics on questions with the greatest relevance to decision-making.

The DSII Intern converted the SPSS data into formats readable by Python, making it possible to create a user-friendly tool for sophisticated engagement with the data. This was the first time the Intern faced the responsibility of transforming such a large and sensitive data set into a machine-readable format. It also required the student to evaluate various possible dashboard platforms, to best align the choice with the needs of the office. This was another type of practical implementation responsibility that the Intern encountered for the first time. Thanks to the Intern's motivation, dedication, and resourcefulness, and to the iterative process of 1-1 meetings and trial-and-error explorations together with the administrative mentor, the result was an effective solution. After selecting and teaching themselves to deploy a Python flask application utilizing Plotly Dash as the dashboard tool, the Intern translated the administrator's ranking of survey question importance, demographic characteristics of greatest relevance, and chosen bucketing of responses into easy-to-analyze data interactivity.

The resulting analyses of the NCHA data would not otherwise have been possible, including the disaggregation of data across multiple groupings, such as race, gender, academic program, and year of graduation. One immediate result of these new insights was a change in interventions related to alcohol consumption. Student affairs professionals are naturally concerned about risky behaviors involving alcohol, especially among first-year students, since the inexperience of first-years suggests that they stand to benefit most from extra guidance and education. Yet, the data scientific analyses of the campus-specific NCHA data, which could now be disaggregated by class year, demonstrated that on this specific campus, it was juniors and seniors who are most engaging in dangerous behaviors involving alcohol. As a result, the DSII's administrative mentor began redesigning alcohol safety programming to incorporate the extra support needed by upper-class students.

This type of experience – of the Intern working through data resources that are difficult to access, or otherwise messy, unreliable, or unwieldy, which in turn results in new opportunities for policy upgrades, innovations, and interventions that increase office effectiveness – this has been perhaps the most common arc that describes the DSII projects at work.

With regard to the resources necessary for long-term sustainability, the pilot phase of the past three years marks a clear path forward. The DSII program has been conceptualized, implemented, and managed by two members of the Computer Science department, the co-authors of this paper, one a senior faculty member who is former chair, and the other a research scientist. The university has provided some part-time staff support, but neither of the founder-directors have received additional pay for launching this program. Other members of the Computer Science department who have provided additional technical support to DSII Interns, attended some of the weekly meetings, or participated in the DSII finalist selection committee have also not received additional pay. Though the time commitment for this participation has been low, and those engaged with the DSII Interns and projects have appreciated the value and opportunity to be connected to the program, we believe that institutionalizing the DSII program for the long-run will require new funding. Phase II will lay out the conceptual details of a fundraising plan and budget to more formally link the DSII program to academic and administrative resources at the university.

5 INTERNAL INTERNSHIP SPECIFICATION

In our experience, four design characteristics define the educational value for data science students and the operational value for data-rich administrative offices that the Internal Internship makes possible. We speculate that these characteristics – data access for Interns, commitment to teaching and learning between intern and administrator, and high level strategic value of individual projects – seem more easily implementable in a higher education environment than in other industry sectors. In summary, those characteristics are as follows:

Data Access: granting access to students of raw, undigested data from organizations needing data scientific solutions is challenging. It is challenging for many reasons, including legal, organizational, and competitive ones. Looking inside the university to administrative offices for experiential learning opportunities in data science makes it much easier to mitigate many of these constraints on access.

Domain Knowledge Mentoring: domain knowledge expertise is crucial for all data scientists. DSII Interns have the unusual benefit of regular weekly meetings directly with their administrative mentors. As a result, students are continually testing their awareness of the problem that needs to be solved, designing technological solutions, iterating on those designs to better fit the expectations of their mentors, and learning how best to accomplish real impact. The DSII experience has revealed that for many administrators in a university environment, fulfilling teaching and mentoring roles is a welcome opportunity.

Translating from technology to outcomes: Since the DSII program places Interns in direct and on-going relationship with Administrators, with the Interns serving the lead role in design and implementation of data scientific innovations, Interns have the responsibility of translating, explaining, justifying, and guiding their supervisors on the practical relevance and value of technologies they are helping to implement. In other words, DSII Interns take on teaching and mentoring roles themselves, helping their supervisors learn about data science tools, and, consequently, also helping them to acquire practical experience in deploying those tools to improve data-informed decision-making.

Making a significant contribution to the university community: DSII projects are defined by and linked to high-level university administrators who are making important decisions about university operations. Improving these university operations through data science gives participating students a chance to leave a legacy of efficiency and effectiveness that will benefit future generations of students, faculty, and staff. The chance to make a positive and lasting difference fuels intense student motivation, a motivation that incorporates all the self-teaching of technological application skills,

acquisition of domain knowledge, and effective translation from technology to outcomes that characterizes the DSII Intern student experience.

In short, the DSII echoes the experiential learning benefits of traditional data science internships, adds the direct accountability for outcomes experienced by students in co-curricular activities like hack-a-thons, data-thons, skunkworks, and robotics clubs, and then infuses both of those qualities with unusually high strategic relevance for undergraduate data science students who care about making a lasting contribution to their university communities [2, 12].

6 EVALUATION AND FURTHER RESEARCH

During the first phase of launching the Internal Internship, we have been limited in our ability to design and share metrics of evaluation and impact on students, or on administrators and their offices. This is due to the private and sensitive nature of many aspects of university administration, and from not having yet developed policies for public distribution of participant experiences. The DSII will now be prepared to move beyond the pilot phase, thanks in part to the development of the DSII Repository. This Repository of proprietary documents, scripts, commentary, data, analytic findings, visualizations, tutorials, and custom-designed software tools, is being designed and overseen by the university's network security division within central IT, a team that includes a DSII Intern dedicated to this project. From the point of view of computer science educators, the most immediate relevance of this Repository is that it will include protocols for research surveys of program participants. It will also include public access to demonstrations and metrics of impact on a project-by-project basis.

The pilot phase of the Internal Internship has yielded implementation details about the feasibility as well as the special educational and operational value of placing undergraduates in roles as data scientific innovators in university administration. It has also illuminated the research questions most worth addressing in the next phase: (a) what are the learning outcomes of the DSII experience from the point of view of participating undergraduates, or, to put another way, what does it look like when students who are already demonstrating outstanding accomplishment are given unique opportunities to push their skills and talents even farther? In much the same way that Senior Honors Thesis students are evaluated by a rubric that measures the quality and significance of their independent research, so should DSII Interns be evaluated by the usefulness and impact of their technological innovations on their supervisors, the workflow of their offices, and on strategic responsibilities of related offices and of senior university leaders; (b) how might the DSII experience inform undergraduate data science education more broadly – are there specific coursework components Interns might identify as essential for all undergraduates studying data science? Might it be possible to transform non-sensitive aspects of individual DSII projects into innovative data science course materials? and (c) what is the impact of the DSII experience on administrative leaders, from a technology leadership and professional development point of view, as they mentor students to deploy data scientific innovations in their offices?

With regard to DSII's impact on career preparedness, we can share, anecdotally, comments from DSII alumni:

> "Many higher-level COSI courses require students to tap into online resources to teach themselves new tools and techniques, but nothing gave me as much experience in doing this as did the DSII. My admin mentors did not have opinions on how I achieved the outcomes they needed. They just needed me to get it done in a way that was useful for them. I am teaching myself new tools and techniques all the time now. It's something that everyone here [in this large tech company] learns how to do. But I'm a step ahead because I got to start learning how to do it in college."

> "Those weekly DSII sessions were especially valuable. I got better and better at being precise about what I had accomplished each week, what challenges I faced, and how I was planning to make progress the following week. All the technology and analysis teams here have some version of this process. I am so much better prepared than other recent college grads."

> "The DSII gave me the experience and therefore the confidence to learn how to speak with subject matter experts and decision-makers. I know the importance of reaching out for guidance about defining the scope of a problem that needs to be solved by technological tools. Thanks to this confidence, I've been noticeably more successful than my peers here [at this start-up company]."

7 CONCLUSION

Our reflection on three years of implementation experience with the DSII underscores the unique pedagogical characteristics of its Internal Internship structure, even while acknowledging its exploratory status. The pilot phase of the DSII can serve as a proof of concept for other universities. We believe it would be worthwhile to replicate even on the scale of one test project, with one willing administrator, one carefully selected student, and an interested member or two of the computer science department. We believe this because we've experienced unique benefits that can come from the Internal Internship. These benefits include an increased spirit of openness, creativity, and motivation for collaborative learning, among participating students, administrators, and faculty. Even though intangible, these benefits, we sense, redound to support the university's commitment to excellent data science education in the long run.

Since we found that the support of senior university leaders was essential in jump-starting the Internal Internship, we would be willing to offer personal testimonials to senior leaders of other universities whose computer science faculty might consider launching a similar pilot program. We see great potential from more formal research as well as from further deployment of the Internal Internship.

ACKNOWLEDGMENTS

We would like to thank Zhiyu Wang, Wenxiao Wang, and Rowan Scassellati for their research assistance on this paper.

REFERENCES

[1] Genevera I Allen. 2021. Experiential learning in data science: Developing an interdisciplinary, client-sponsored capstone program. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. 516–522. https://doi.org/10.1145/3408877.3432536

[2] Craig Anslow, John Brosz, Frank Maurer, and Mike Boyes. 2016. Datathons: an experience report of data hackathons for data science education. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*. 615–620. https://doi.org/10.1145/2839509.2844590

[3] Arko Barman, Su Chen, Andersen Chang, and Genevera Allen. 2022. Experiential Learning in Data Science Through a Novel Client-Facing Consulting Course. In *2022 IEEE Frontiers in Education Conference (FIE)*. IEEE, 1–9. https://doi.org/10.1109/FIE56618.2022.9962532

[4] Austin Cory Bart, Dennis Kafura, Clifford A Shaffer, and Eli Tilevich. 2018. Reconciling the promise and pragmatics of enhancing computing pedagogy with data science. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. 1029–1034. https://doi.org/10.1145/3159450.3159465

[5] Maxime C Cohen, C Daniel Guetta, Kevin Jiao, and Foster Provost. 2018. Data-driven investment strategies for peer-to-peer lending: a case study for teaching data science. *Big Data* 6, 3 (2018), 191–213. https://doi.org/10.1089/big.2018.0092

[6] Richard D De Veaux, Mahesh Agarwal, Maia Averett, Benjamin S Baumer, Andrew Bray, Thomas C Bressoud, Lance Bryant, Lei Z Cheng, Amanda Francis, Robert Gould, et al. 2017. Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and Its Application* 4 (2017), 15–30. https://doi.org/10.1146/annurev-statistics-060116-053930

[7] Silvia Elena Gallagher and Timothy Savage. 2023. Challenge-based learning in higher education: an exploratory literature review. *Teaching in Higher Education* 28, 6 (2023), 1135–1157.

[8] Sandra Gorka, Jacob R Miller, and Brandon J Howe. 2007. Developing realistic capstone projects in conjunction with industry. In *Proceedings of the 8th ACM SIGITE conference on Information technology education*. 27–32. https://doi.org/10.1145/1324302.1324309

[9] Johanna Hardin, Roger Hoerl, Nicholas J Horton, Deborah Nolan, Ben Baumer, Olaf Hall-Holt, Paul Murrell, Roger Peng, Paul Roback, D Temple Lang, et al. 2015. Data science in statistics curricula: Preparing students to "think with data". *The American Statistician* 69, 4 (2015), 343–353. https://doi.org/10.1080/00031305.2015.1077729

[10] Jessica Liebowitz, Timothy J. Hickey, and Matthew Ekins. 2021. Exploring the Impact of Computer Science on the Future of Higher Ed. (2021), 31 pages. https://doi.org/10.48617/pre.129 Paper presented at Data Science Initiative Series, Brandeis University.

[11] Susan E Martonosi and Talithia D Williams. 2016. A survey of statistical capstone projects. *Journal of Statistics Education* 24, 3 (2016), 127–135. https://doi.org/10.1080/10691898.2016.1257927

[12] Arnab Nandi and Meris Mandernach. 2016. Hackathons as an Informal Learning Platform *(SIGCSE '16)*. 346–351. https://doi.org/10.1145/2839509.2844590

[13] Office of Marketing and Brandeis University Communications. 2023. Brandeis Stories: Fostering Data-Driven Change. https://www.brandeis.edu/stories/2023/october/data-science-internship.html Accessed on March 28, 2024.

[14] Uzma Haque Syeda, Prasanth Murali, Lisa Roe, Becca Berkey, and Michelle A Borkin. 2020. Design study" lite" methodology: Expediting design studies and enabling the synergy of visualization pedagogy and social good. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13. https://doi.org/10.1145/3313831.3376829