The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
ASEE
Paper ID #44284

Assessing Chemical Engineering Students' Perspectives on Data Science and Its Integration in the Academic Curriculum: Implications for Industry Readiness

Dr. Betul Bilgin, The University of Illinois at Chicago

Betul Bilgin is a Clinical Associate Professor of Chemical Engineering (CHE) at the University of Illinois Chicago (UIC).

Naomi Groza, The University of Illinois at Chicago

Assessing Chemical Engineering Students' Perspectives on Data Science and Its Integration in the Academic Curriculum: Implications for Industry Readiness

In the rapidly evolving landscape of chemical engineering (CHE), the incorporation of data science has gained increasing importance. To equip students with the skills required for a data-driven industry, it is crucial to understand their perceptions of data science and their willingness to embrace it in their academic and professional journey. This study engages a diverse group of chemical engineering students across different academic levels to explore their viewpoints on data science and its potential integration into the academic curriculum.

The instrument assesses four crucial constructs: interest, career aspirations, perceived value, and self-efficacy regarding data science. The study delves into students' prior exposure to data science concepts, their recognition of data science's relevance in chemical engineering, and their readiness to engage with data science tools and techniques. Students, both seniors, juniors and sophomores in CHE, participated in the interviews. The study's findings reveal that, despite limited exposure within their coursework, most students acknowledge the growing significance of data science in the chemical engineering field. The majority express a keen interest in expanding their knowledge of data science and are receptive to its integration into their academic and future career paths.

Moreover, this research identifies barriers to the incorporation of data science in the CHE curriculum, such as the need for additional resources and training for both students and faculty. The insights presented in this study offer valuable guidance for educators and industry professionals seeking to seamlessly embed data science into the chemical engineering curriculum and better prepare students for a data-centric industry.

This paper provides a comprehensive overview of interview development, data distribution, and key findings. It underscores the urgency of further research to enhance the integration of data science in the CHE curriculum and the essential role of preparing students for an industry that increasingly relies on data analytics and computational techniques.

Introduction

The integration of data science in chemical engineering is a rapidly evolving field, with a focus on data management, statistical and machine learning, and data visualization [1]. This integration is crucial for handling the increasing complexity and size of data sets in chemical engineering research and practice [2]. Data science has particularly impacted molecular science in chemical engineering, with applications in molecular discovery and property optimization [3]. The development of a cyberinfrastructure for data-driven design and exploration of chemical space further underscores the potential of data science in transforming chemical research [4]. The alignment of data analytics and strategy is transforming the chemical industry, with data playing a crucial role in production, research, marketing, and customer service strategies [5]. The use of big data and analytics in chemical engineering is driving innovation, increasing manufacturing efficiency, and fostering collaboration with other industries [6]. It is also critical to form a team proficient in skills, a comprehensive understanding of data science, and a

concentrated emphasis on the genuine requirements pivotal to the digital transformation within the chemical industry [7], [8].

The integration of data science into chemical engineering curricula is crucial for preparing students to meet modern industry challenges [1]. This integration can be achieved through the modification of existing courses, introduction of electives, or the creation of specialized programs [9]. Data science principles, including data management, statistical and machine learning, and visualization, are particularly important in this context [2]. Furthermore, the use of computer-aided modeling and simulation (CAMS) can enhance the problem-solving capabilities of chemical engineering graduates [10]. Such studies are instrumental in guiding the formulation of an academic curriculum designed to furnish chemical engineering students with the essential competencies and knowledge requisite for navigating a future dominated by data.

Instructors endeavoring to introduce new pedagogical approaches encounter several well-documented challenges, including constraints on time, diversity in student academic backgrounds, their own limited training in new domains, insufficient institutional support, the necessity to adhere to extensive curricular content, and a scarcity of resources [11]-[14]. The integration of data science into curricula introduces further complexities, given its status as a burgeoning interdisciplinary field that synthesizes elements of mathematics, statistics, computer science, and engineering, along with critical aspects of design, communication, and decision-making theory [15].Additionally, students in disciplines just beginning to adopt data science often lack both the curriculum support and foundational knowledge required for engaging with data science-intensive courses. Faculty are thus challenged with not only devising courses or modules that cater to students of varied educational backgrounds but also with the need for personal upskilling and a reevaluation of their traditional disciplinary paradigms. Identifying a starting point necessitates considerable effort at both the individual and organizational levels.

Building on the challenges faced by instructors in integrating data science into their curricula, it's crucial to consider the student perspective, which plays a pivotal role in the successful adoption of these new pedagogical approaches. Students' perceptions toward data science, their openness to embracing this interdisciplinary field, and their readiness to engage with its complexities significantly influence the effectiveness of educational innovations. Understanding these perceptions is essential for educators to tailor their teaching strategies, ensuring that they not only overcome the logistical and institutional barriers but also address the educational and motivational needs of their students. Research on the topic of student perceptions towards data science has predominantly focused on the K-12 education sector. The significance of comprehending these perceptions is underlined by Kross, who advocates for educators, especially those with practical industry experience, to cultivate empathy towards the diverse backgrounds and expectations of students [16]. This aspect gains particular importance in the context of secondary education, for the integration of data science as a key interdisciplinary skill [17]. The selection of educational tools in teaching data science is another critical consideration, with Israel-Fishelson offering an insightful review of tools tailored for K-12 learners [18]. Furthermore, a comprehensive Data Science Education Framework for iSchools was introduced, emphasizing methodologies that are user-centric, tool-oriented, and application-focused [19].

This study focuses on students' attitudes and perceptions towards data science. It underscores the importance of the efforts to create accessible and engaging resources that support the introduction of data science concepts across a diverse range of undergraduate courses. By prioritizing the students' perspective, the study aims to foster a more inclusive and effective learning environment that prepares students for the demands of a data-driven world.

Motivation

To adequately equip students for the evolving demands and prospects of a data-centric industry, it is essential to explore their views on data science. This includes understanding its significance and practicality in relation to their field of study, alongside gauging their willingness to integrate data science principles into their academic and future professional pursuits. Understanding students' attitudes towards data science is essential for several reasons:

1. **Curriculum Development**: Insight into students' perceptions allows educators and curriculum developers to tailor educational programs that not only meet the technical demands of the industry but also align with students' interests and aspirations. This ensures the development of engaging and relevant content that enhances learning outcomes and fosters a positive attitude towards data science.
2. **Overcoming Barriers to Learning**: By identifying any misconceptions or apprehensions students may have about data science, educators can design interventions that address these issues directly. This might include demystifying data science, showcasing its integral role in solving real-world chemical engineering problems, and providing supportive learning environments that encourage experimentation and inquiry.
3. **Enhancing Motivation and Engagement**: Understanding students' willingness to engage with data science can help educators employ strategies that boost motivation. For example, integrating practical, hands-on projects that demonstrate the tangible benefits of data science in chemical engineering can make the subject matter more compelling and relevant to students' future careers.
4. **Bridging the Skills Gap**: Acknowledging and acting upon students' perceptions and willingness to learn data science can lead to the development of targeted programs that bridge the gap between traditional chemical engineering education and the emerging needs of the industry. This involves not only imparting technical data science skills but also fostering a mindset oriented towards innovation, continuous learning, and adaptability.
5. **Facilitating Industry-Academia Collaboration**: Understanding student perspectives can also facilitate stronger collaborations between academic institutions and the chemical industry. By aligning educational outcomes with industry needs, students can graduate with a skill set that is immediately valuable and relevant, thereby enhancing their employability and contributing to the advancement of the field.

Methodology

Research Rationale and Design

This study is driven by the imperative to understand chemical engineering students' perceptions and attitudes towards data science, recognizing its crucial role in their future professional landscapes. By exploring how students perceive the relevance of data science to their careers and

their attitudes towards acquiring data science skills, the research seeks to contribute valuable insights into curriculum development and align educational practices with industry expectations. Furthermore, identifying the challenges and barriers students encounter in learning data science will enable educators and policymakers to implement targeted interventions that enhance the educational framework, thereby better preparing students for the demands of a data-driven industry. Guided by the research rationale outlined, this study is designed to address the following research questions:

1. *How do chemical engineering students perceive the relevance of data science to their careers, and what are their attitudes toward learning data science skills?*

2. *What challenges and barriers, if any, do students face in acquiring data science skills, and to what extent do their perceptions and attitudes align with the expectations of the chemical engineering industry regarding data science?*

The study adopts a qualitative research design to gather comprehensive data on students' perceptions, attitudes, and experiences. This design facilitates a nuanced understanding of the complex factors influencing students' views on data science and their readiness to engage with it in their professional lives.

Context

This study is conducted at University of Illinois Chicago (UIC) Chemical Engineering Department. UIC is a Carnegie Classification Research I institution and Hispanic-Serving Institution since 2014, is at the forefront of addressing the challenges and leveraging the opportunities presented by its diverse and vibrant academic community. The College of Engineering (COE) alone enrolls 4,934 students, with notable diversity among its ranks: 22% first generation, 23.6% Hispanic, 22% female. Each year, approximately 30-35 students enter as first-year students in the UIC Chemical Engineering department, and 20-25 students transfer to department from community colleges. Among the first-year chemical engineering students, nearly 27% identify as Hispanic, and nearly 41% are female.

Participants

The participants in this study were undergraduate chemical engineering students from various academic years, specifically sophomores, juniors, and seniors. These students were recruited through an email invitation sent to all chemical engineering students across the first year, sophomore, junior, and senior levels, offering a $25 incentive for participating in an interview. The recruitment process aimed to gather a diverse group of volunteers from each class level to participate in the study.

Despite efforts to include students from all academic years, no volunteers were obtained from the first-year cohort. This absence of first-year participants is attributed to the lack of chemical engineering courses offered to them, which might have led to a feeling of disconnect or unpreparedness to contribute meaningfully to the study on data science integration in their curriculum.

From the pool of volunteers, five students from each class level (sophomore, junior, and senior) were randomly selected, with stratification by gender to ensure diversity. In total, 14 students participated in the interviews, comprising 5 seniors, 5 juniors, and 4 sophomores. The absence of first-year students notwithstanding, the selected participants offered a broad perspective across the different stages of the chemical engineering undergraduate program.

Data Collection

Data collection for this study was conducted through comprehensive interviews with all participating students. These hour-long interviews were carried out by fellow undergraduate students who had not only completed human subject research training but were also directly involved in this research project. The choice of peer interviewers was strategic, aiming to create a relaxed and relatable atmosphere that encouraged participants to express their thoughts and experiences freely and openly.

The structure of the interview was meticulously designed to evaluate four critical constructs that are pivotal in understanding students' engagement with data science within the chemical engineering curriculum [20]. These constructs are:

1. **Interest**: Participants' enthusiasm and curiosity about data science.
2. **Career Aspirations**: How participants see data science fitting into their future professional plans.
3. **Perceived Value**: The importance participants attribute to data science skills in the context of their education and future career in chemical engineering.
4. **Self-Efficacy**: Participants' confidence in their ability to learn and apply data science skills effectively.

To thoroughly assess these constructs, the interview questions were carefully crafted and organized (sample questions are listed in *Table 1*). Participants were asked a series of questions tailored to elicit detailed responses that would shed light on their perceptions and attitudes towards data science. The questions aimed to explore not only the participants' current engagement with data science but also their long-term view of its relevance and utility in their prospective careers.

Table 1: Sample Interview Questions

Construct	Sample Questions
Understanding of Data	What does data mean to you? How important is data in the chemical engineering profession?
Definition of Data Science	How would you define data science? What does data science mean to you?
Current and Desired Skills	What data science skills do you currently possess? What data science skills are you planning to acquire prior to graduation?
Importance of Skills in Career	In your opinion, how important is it to have these skills in chemical engineering careers?

Data Analysis

Qualitative data from interviews were transcribed and subjected to thematic analysis. This process involved individual and group coding the data, identifying themes and patterns, and interpreting the findings in the context of existing literature and the study's objectives. The study adhered to ethical guidelines, ensuring the confidentiality and anonymity of participants. Informed consent was obtained from all participants, and they were informed of their right to withdraw from the study at any time without penalty.

Results and Discussions

To analyze the qualitative data obtained from the interviews, a coding system was established to categorize responses according to the four constructs outlined in the study: Interest, Career Aspirations, Perceived Value, and Self-Efficacy regarding data science. Beyond exploring the four primary constructs, students were also queried about their understanding of definition of data science, the current state of its integration into the chemical engineering curriculum, and their suggestions for improvement. This thematic coding allowed for the systematic organization of data, facilitating the identification of prevalent themes, patterns, and frequencies of codes related to each construct. The coding process was iterative, involving initial coding to capture broad themes followed by focused coding to refine the understanding of each construct.

The frequencies of codes within each construct were visually represented in *Figure 1*, which illustrates the distribution and prominence of specific themes discussed by the participants. This visualization aids in identifying which aspects of data science are most engaging to students, their views on the relevance of these skills to their future careers, the value they place on data science education, and their confidence in their abilities to succeed in this area.

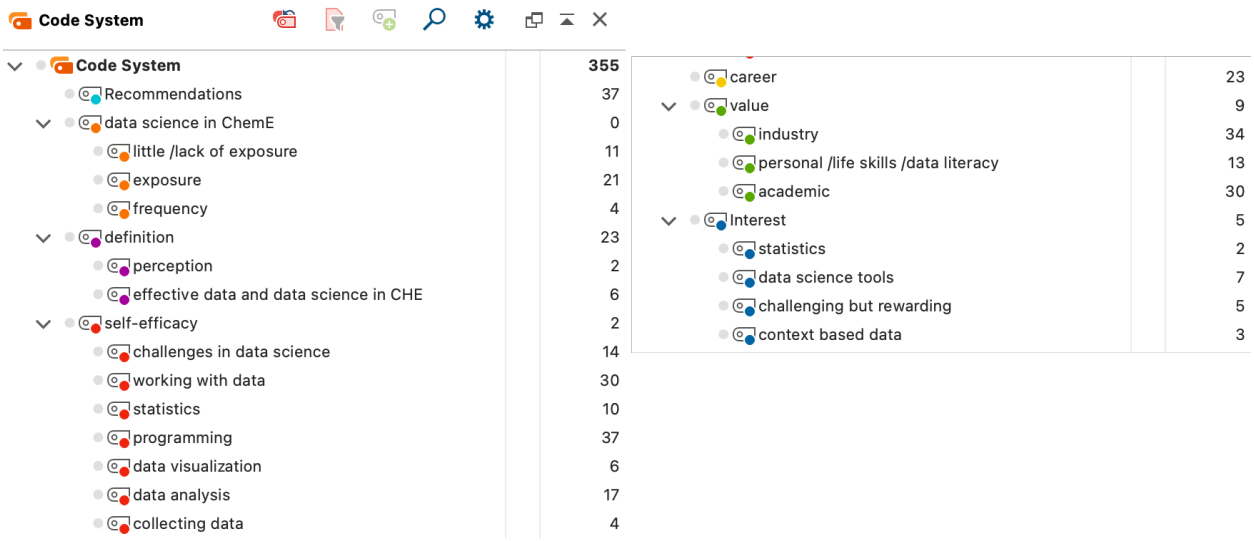

Figure 1: Code system and frequencies.

The analysis of code frequencies revealed several key findings:

Definition of Data and Data Science:

The participants' responses revealed a range of understandings regarding the definition of data and data science, shaped largely by their previous experiences and exposure. While some

students articulated their definitions with a degree of confidence, others expressed uncertainty, indicating that their conceptualizations of data and data science are evolving. The responses can be categorized into three main themes: "Numbers and Analysis," "Interpreting Data," and "Unsure what data/data science is."

Numbers and Analysis
Students within this category perceive data as a numerical representation of systems or phenomena and view data science as the discipline focused on analyzing these numerical datasets to derive meaning and insights.

- "Data is a numerical description of a system, and data science is a group of numerical data, analysis, and meaning of it." This response underscores the quantitative nature of data and the analytical processes involved in data science.
- Another student elaborated, "Data is the representation of information, and data science is the study of data." Here, the emphasis is on data science as a field dedicated to understanding and analyzing data representations.
- Furthermore, a participant described data science in terms of practical application: "Data is putting variables into equations; data science is getting meaning out of it." This highlights the process of transforming raw data into actionable insights through analytical methods.

Interpreting Data
In this theme, students focus on the importance of contextualizing and interpreting data, suggesting that a deep understanding of the subject matter is essential for accurate data analysis.

- One student noted the critical role of context: "To contextualize data, you need to know the subject to assess if data analysis is done correctly; your code might work, but that does not mean you are getting the right answers." This statement reflects the idea that data science involves more than just numerical analysis; it requires an integration of subject matter expertise to ensure meaningful interpretations.
- Another perspective offered was, "Data is in sensors and economics in chemical engineering; data science is interpreting these values and creating a story." This view emphasizes the narrative aspect of data science, where data from diverse sources is synthesized into coherent stories that inform decision-making processes.

Unsure What Data/Data Science Is
A segment of the participants expressed uncertainty about the precise definitions of data and data science, reflecting a perception of these concepts as integral, yet diffuse, components of the broader STEM fields.

- "I did not think about data science as its own field; it is part of everything in STEM." This comment suggests that, for some students, data science is not seen as a distinct discipline but rather as an inherent aspect of scientific inquiry and problem-solving across all areas of STEM.

Interest in Data Science:
The participants expressed a keen interest in data science skills, although they often did not explicitly identify these skills as "data science." Instead, these capabilities are viewed as essential tools that facilitate more effective research and analysis within their field. The students' interests

are categorized into four main themes: "Statistics," "Data Science Tools," "Context-Based Data," and "Challenging but Rewarding."

Statistics
Some students showed a preference for certain types of mathematical analysis while expressing gaps in their education, particularly in statistics, which is foundational to data science.

- "I like straightforward math, I really like calculus. I never took statistics." This statement reflects an appreciation for the mathematical underpinnings of data science, yet acknowledges a lack of exposure to statistical methods critical for data analysis.

Data Science Tools (mainly Excel, VBA)
Students expressed a specific interest in utilizing data science tools, particularly Excel and VBA, to deepen their analysis and enhance their research capabilities.

- "I definitely want to use Excel to figure out data get a more in-depth analysis of data." This indicates a desire to leverage widely available software tools to perform complex data analyses.
- Another student added, "It definitely is something that interests me, and I will be taking future courses at least to develop my programming skills." This shows a proactive approach towards acquiring the necessary data science skills, emphasizing the importance of programming in their toolkit.

Context-Based Data
Participants also highlighted the relevance of data science in their everyday lives and its specific applications within chemical engineering (CHE).

- The interest in "uses in everyday life but interested in CHE data" suggests that students value data science not only for its general utility but also for its potential to solve discipline-specific problems.

Challenging but Rewarding
Students acknowledge the challenges associated with working with data, especially in research contexts where outcomes are uncertain. Despite these challenges, they find the process rewarding.

- "I like working with data. Sometimes it's frustrating, because if you're working on research, you really don't know what the right answer is supposed to be." This quotation captures the inherent uncertainties in research and the problem-solving satisfaction that comes from navigating these challenges.

Self-Efficacy in Working with Data:
The participants' responses indicate a substantial level of exposure to data handling, which has contributed to a generally high degree of self-efficacy regarding their data-related skills. This exposure spans across their academic experiences in chemical engineering (CHE) and other disciplines, as well as in their personal and professional lives. The students' confidence in working with data is reflected in the quotations and can be organized into two categories: "Working with Data in CHE and Non-CHE Context" and "Confident/Comfortable in Working with Data."

Working with Data in CHE and Non-CHE Context

Students reported working with data in various contexts, both within chemical engineering courses and in other scientific labs, suggesting that their educational programs provide a broad foundation in data handling.

- "CHE 205: data visualization and how to use data, using data as opposed to collecting your data is easier." This response indicates that students find working with pre-existing data sets more straightforward than collecting data themselves, highlighting the importance of courses that focus on data analysis and visualization.
- Another participant mentioned "collecting data in labs, analyzing data in homework assignments," showing that data skills are being practiced not only in lab settings but also reinforced through coursework.
- Practical applications were also noted: "in chemistry lab and unit operations lab," suggesting that hands-on experience in labs is a critical component of their data education.
- One student detailed their professional experience: "at job inventory management and cost via Excel," which shows that data skills are directly applicable and valuable in workplace settings.

Confident/Comfortable in Working with Data

A significant number of students expressed confidence in their data skills, attributing this to the amount of exposure and practice they have received.

- "Very comfortable within engineering and money handling context" suggests that students feel at ease when working with data in familiar contexts, drawing parallels between engineering data analysis and financial management.
- "More exposure meant more confidence—learning about one data set helped to understand others in the same context" indicates that students recognize the value of experience in building their confidence, understanding that skills developed in one area can be transferable to other data sets and contexts.

Perceived Value of Data Science in Career

Participants unanimously recognized the importance of data science skills for their current and future careers, with some emphasizing the growing significance of programming abilities. Their acknowledgement of the value of data science is reinforced by the increasing accessibility and availability of data, which in turn amplifies the need for data science competencies. The quotations provided can be summarized below. Students clearly see data science as an integral part of their professional development and career advancement within the sciences, particularly in engineering.

- "It's super important. Science is a lot of data. For example, if we were trying to make a power plant as efficient as possible, we got to run a bunch of like experiments to figure out which set of equipment or materials best for the efficiency, and that's all like because of data. So it's definitely important for us to learn all the data before we graduate, all the data but like the techniques to use data." This response encapsulates the belief that data science is not only relevant but crucial for optimizing processes and equipment in real-world engineering scenarios.
- Another student highlighted the burgeoning demand for such skills: "I would say currently they're in very high demand, and I think, moving into the future, they're gonna

stratospherically in demand. Because if you have data science skills and one particular niche field wearing, you're considered a subject matter expert like chemical engineering." This perspective underscores the competitive edge that data science skills provide, positioning individuals as experts in their field and enhancing their employability.

Student Performance in Data Science Skills:
The student responses regarding their experience and performance in statistical and programming skills, which are essential components of data science, revealed a mix of proficiency levels. While some students had direct experience through coursework, others had not formally taken statistics courses or deeply engaged with complex programming. Their self-evaluation of performance varied, often depending on whether they had taken a specific course on the subject.

Statistics
A significant number of students reported a lack of formal training in statistics, a fundamental component of data science, which impacts their ability to perform statistical analysis. Only three took stat courses, only two mentioned using stats in CHE courses. This indicates that despite the use of statistics in various courses, formal education in this area is not widespread among the participants. Their ability to perform in statistics is often judged based on whether they have taken a specific statistics course, leaving many without a solid foundation in this area.

Programming
Programming appears to be a more familiar area for the students, with a range of experiences across different platforms and languages. All have some level of programming (VBA, Matlab, Python)." This suggests that while programming is part of their academic experience, its relevance to data science is not always made explicit. The level of confidence varies many are comfortable with basic programming but weak on complex programming. Students report comfort with fundamental programming but acknowledge their limitations with more advanced programming tasks, indicating an area for potential growth and development within the curriculum.

Data Visualization
Data visualization is an area where students feel confident, particularly when it comes to basic tasks. They are confident with simple Graphing and plotting. The ability to create basic graphical representations of data is well within the comfort zone of most students, signifying that at least some data presentation skills are being effectively imparted through their education.

Challenges in Data Science:
The responses from chemical engineering students regarding the challenges they face in data science primarily revolve around the issue of contextualizing data. Many students recognize that without understanding the context or meaning behind data, it merely represents a collection of numbers without actionable insights. Experienced students, in particular, report difficulties in managing and interpreting large datasets. The challenges are categorized into three main themes: "Coding and Using Software," "Not Able to Contextualize/Make Use of Data," and "Working with Large Datasets."

Coding and Using Software
Students expressed challenges related to the technical aspects of data science, such as coding and the use of software tools necessary for data analysis.

- "Coding, running scripts, VBA, making ODE functions work, need practice to code." This quote illustrates the technical barrier students face, indicating a need for more practice and training to become proficient in coding, which is integral to data science.

Not Able to Contextualize/Make Use of Data
A common challenge is the ability to understand and apply data to real-world situations, transforming it from abstract numbers to valuable information.

- "Useless if you do not know how to utilize it. Scary and overwhelming sometimes." This response reflects the apprehension and difficulty students experience when they cannot contextualize data.
- "Applying data into specific applications" further emphasizes the struggle to translate data analysis into practical applications, highlighting a gap between technical skills and real-world implementation.

Working with Large Datasets
Handling large volumes of data is another significant challenge reported by students, particularly in terms of analysis and deriving meaningful conclusions.

- "Working with large datasets and drawing conclusions with those, finding patterns, correlations are harder with large data sets." Students acknowledge the complexity that comes with large datasets, which requires advanced analytical skills to uncover patterns and correlations.
- "Navigating through data and finding the correct value (reading tables)" points to the difficulty in sifting through extensive data to extract relevant information, a skill that becomes increasingly important with the growth of big data.

Data Science in the Chemical Engineering Curriculum
The participants in this study conveyed that while they encounter data regularly in their chemical engineering (CHE) courses, they do not typically identify these experiences with the formal terminology of "data science." This lack of recognition persists regardless of class year or gender among the students. Their insights can be organized into reflections on "CHE 200 level classes" and "CHE 300 level/senior year classes," as well as a general perspective on "Classes overall."

CHE 200 Level Classes (CHE 201 and CHE 205)
Students receive their initial exposure to working with data in their sophomore courses, yet they do not equate this with learning data science.

- "205 and 201 HW problems given some data" and "examples in 201 and 205 HW, data sets and analyzing data. Data science skills make everything quicker and easier." These responses illustrate that while students are engaging with data and its analysis, they view these tasks as separate from the data science field.

CHE 300 Level/Senior Year Classes (CHE 301, 321, 341, 396-397)
As students progress to higher-level courses, the complexity of data work increases, but again, it is not explicitly associated with data science.

- "CHE 301, HW assignments, plot, find regression" and "project in CHE 321, Matlab." These examples show that students undertake more sophisticated data analysis tasks, such as regression and computational modeling, which are foundational to data science, but the connection to data science as a discipline is not made clear in the curriculum.

Classes Overall

Students recognize that they are employing data science techniques throughout their coursework; however, these techniques are not labeled as "data science" within the curriculum.

- "Data science techniques are employed but not called data science" and "assignments in ChemE, raw data to conclusion." Students are applying data science methods to draw conclusions from raw data, yet the curriculum does not frame these experiences within the data science paradigm.

Recommendations: Enhancing Data Science in the Chemical Engineering Curriculum

The chemical engineering students at UIC have provided a range of suggestions for incorporating data science more effectively into their curriculum. Their recommendations touch on the integration of programming languages, the timing and relevance of data science education, and the use of real-world examples to underscore the importance of data analysis in their field. Here is a synthesis of their recommendations along with direct quotations:

Programming and Software Tools

- Students suggest a stronger emphasis on programming languages known to be useful in chemical engineering contexts, such as VBA, MATLAB, Python, and C++.
    - "VBA like excel VBA on one hand… it's really effective in chemical engineering... Matlab and... Python and C++... those could be... included maybe as a component of trying to write some programming."

Early Exposure and Skill Development

- The timing of data science education is critical, with a call for introducing relevant skills earlier in the academic journey to better prepare students for the workforce.
    - "If we are not developing more skills of data science in our sophomore to junior years, it kind of hinders us for our future of preparation in the work life area where data will be everywhere anyway."

Dedicated Data Science Courses

- There is a strong demand for dedicated data science courses tailored to chemical engineering that focus on industrial applications of data analysis.
    - "Those classes should definitely be replaced with a chemical engineering data science class… data analysis... it should be more industrial related... a rigorous data science class for chemical engineering that analyzes data from the industry."

Practical Lab Experience

- Introducing lab classes earlier in the curriculum that emphasize data collection and analysis is recommended to provide practical, hands-on experience.
    - "I feel like maybe a lab class earlier in our curriculum would be a little bit beneficial than waiting until senior year to take a lab."

Real-World Applications

- Students recognize the value of learning from real-world data and suggest the curriculum should include more authentic industrial data sets for analysis.
    - "It's a realistic can be or that has been seen in industry. I feel like using real world examples would be pretty cool."

Enhanced Data Analysis Education

- There is a call for a greater focus on data analysis techniques, such as regression and pattern recognition, throughout the chemical engineering curriculum.
    - "I think if we had more education, just in terms of simple regression, polynomial, fitting things of that nature, both in excel as well as python, that would be pretty sweet."

Integration of Data Science Throughout the Curriculum

- Students propose that data science concepts should not be confined to a single course but integrated throughout the chemical engineering curriculum.
    - "Maybe adding in a lecture here and there, or implementing it across of the whole undergraduate classes... give a data science-centered problem."

Application and Relevance

- Emphasizing the relevance of data science problems to actual chemical engineering practices is seen as crucial for student engagement and understanding.
    - "The biggest component would be some form of actual application, it needs to be something that's seen by the student as relevant to their education in their career."

A Survey Analysis of Data Science Skills and Perceptions

We also conducted surveys to assess the current state of data science skills among undergraduate students at the University of Illinois Chicago (UIC), focusing on sophomores and seniors. The survey, which garnered responses from 49 sophomore and 62 senior students, explored both the self-assessed proficiency levels in various data science skills and the perceived importance of these skills for their future careers.

Participants were asked to rate their proficiency in skills such as Python, Data Visualization, Critical Thinking, Excel, Communication Skills, Statistics, and Machine Learning on a Likert scale of 5; “Not Know At All”, “Know Very Little”, “Know Some”, “Confident” and “Expert”. Additionally, they evaluated the importance of these skills in their careers on a 4-point Likert scale; ‘Very Unimportant”, “Unimportant”, “Important” and “Very Important. We compared the combined responses on “Confident” and “Expert” in analyzing students’ confidence level and combined "Important" and "Very Important" responses for analysis of students’ perception of importance of these skills. This dual-focused approach allowed for a comprehensive understanding of not only where students see themselves skill-wise but also where they believe they should be focusing their educational efforts. The Figure 2 compares the self-assessed proficiency levels of sophomore and senior students across various data science skills, while Figure 3 represents the perceived importance of those skills.

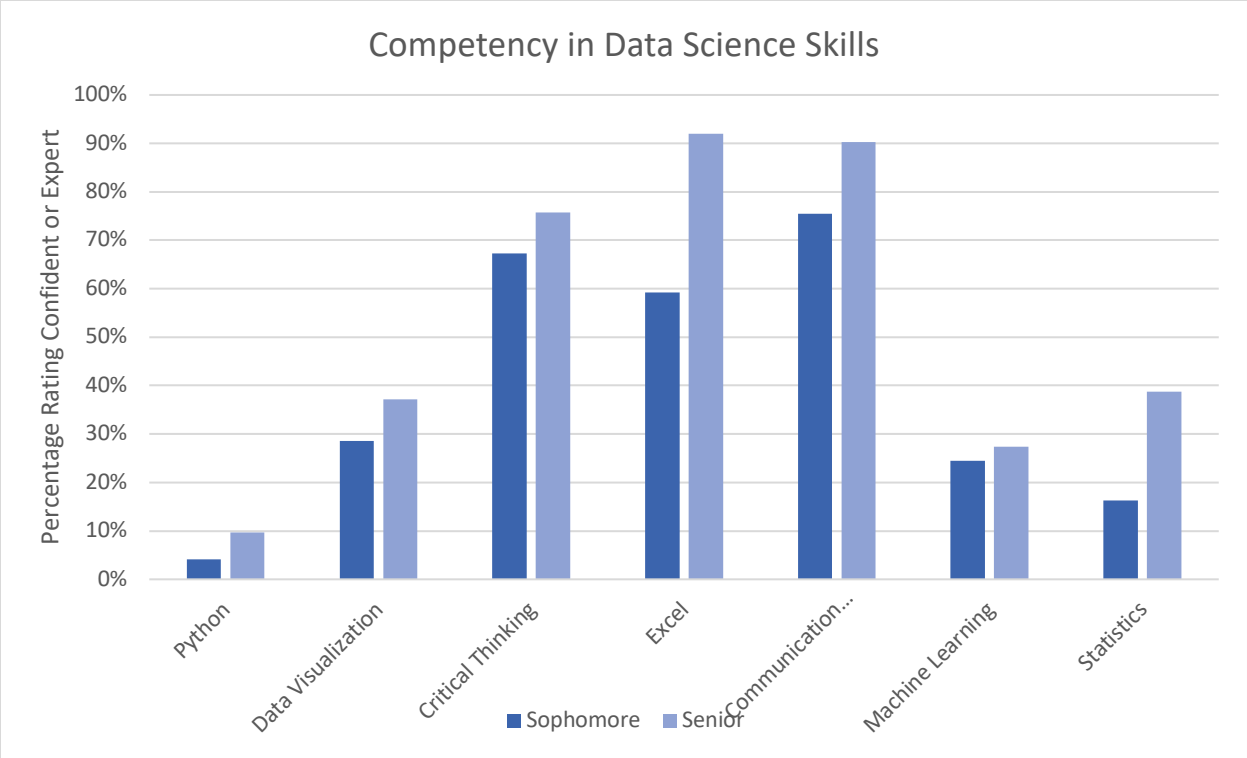

Figure 2: Self-assessed proficiency levels of sophomore and senior chemical engineering students at UIC across various data science skills, including Python, Data Visualization, Critical Thinking, Excel, and Communication Skills.

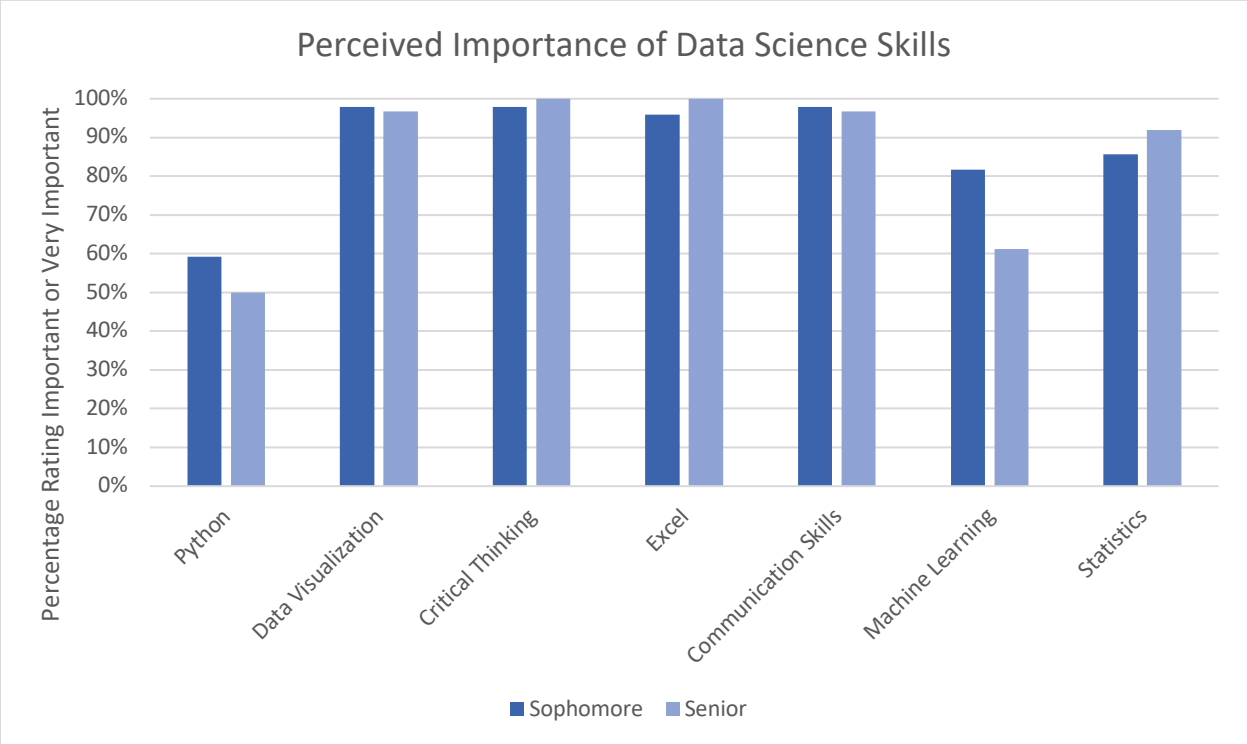

Figure 3: The perceived importance of data science skills for chemical engineering careers.

The analysis of survey results revealed key insights into the self-assessed proficiency of chemical engineering students in critical data science skills. Confidence levels in areas like Critical Thinking and Excel showed significant growth from sophomores to seniors, suggesting a strong alignment with the curriculum's focus. However, proficiency in Python and Machine Learning remained relatively low, despite the recognized importance of these skills.

Students rated Critical Thinking, Excel, and Data Visualization highly in terms of career importance, with seniors recognizing an almost universal importance for Critical Thinking and Excel. Interestingly, the perceived importance of Python decreased from sophomores to seniors, while Machine Learning saw a significant drop in importance among senior respondents, suggesting a possible shift in focus or a nuanced understanding of the chemical engineering profession's demands.

A notable gap between confidence and perceived importance was evident in Statistics and Machine Learning. Despite high importance ratings, students expressed relatively low confidence in these areas, highlighting potential opportunities for curriculum development. The decrease in the perceived importance of Machine Learning from sophomores to seniors also suggests a need for educational institutions to better integrate and clarify the role of emerging technologies in chemical engineering.

Conclusions

The exploration into chemical engineering students' perspectives on data science has provided illuminating insights into their understanding, interest, perceived value, and self-efficacy regarding data science, as well as the challenges they face and the current state of data science within their curriculum. It is evident that while students are exposed to elements of data science throughout their chemical engineering education, there is a distinct gap in the explicit recognition and definition of these skills as part of the data science domain.

Students have articulated a strong interest in data science skills, recognizing their utility in both academic and professional settings. However, there is a clear need for more structured educational pathways that align these interests with formal data science training, particularly in statistical methods and advanced programming. Confidence in data handling is varied, suggesting a demand for curricular enhancements that provide deeper, more consistent exposure to data science concepts.

The challenges students encounter, such as contextualizing data and navigating large datasets, underscore the necessity for a curriculum that goes beyond traditional engineering education to include robust data science components. Such a curriculum would not only improve technical proficiencies but also foster a more intuitive understanding of data analysis and its applications in real-world engineering problems.

Recommendations from the students for curricular improvements include the integration of programming languages relevant to chemical engineering, early and continued exposure to data science concepts, dedicated data science courses with industrial relevance, and the inclusion of practical lab experience that emphasizes data collection and analysis. Moreover, the application

of real-world examples and the continuous reinforcement of data analysis techniques are seen as critical for translating theoretical knowledge into practical skills.

In conclusion, this study highlights a critical moment in chemical engineering education. As the industry advances towards a more data-driven approach, there is a pressing need for academic programs to adapt and evolve. By integrating comprehensive data science education into the chemical engineering curriculum, we can prepare students not only to meet the current demands of their field but to excel as innovators and leaders in the data-rich landscape of the future.

References

[1] T. A. Duever, "Data science in the chemical engineering curriculum," *Processes,* vol. 7, *(11),* pp. 830, 2019.

[2] D. A. Beck *et al*, "Data science: Accelerating innovation and discovery in chemical engineering," *AIChE J.,* vol. 62, *(5),* pp. 1402-1416, 2016.

[3] C. Ashraf *et al*, "Data science in chemical engineering: Applications to molecular science," *Annual Review of Chemical and Biomolecular Engineering,* vol. 12, pp. 15-37, 2021.

[4] J. Hachmann *et al*, "Building and deploying a cyberinfrastructure for the data-driven design of chemical systems and the exploration of chemical space," *Molecular Simulation,* vol. 44, *(11),* pp. 921-929, 2018.

[5] S. Chandran and R. Kasat, "Aligning data analytics and strategy in the chemical industry," *Aligning Business Strategies and Analytics: Bridging between Theory and Practice,* pp. 195-208, 2019.

[6] L. Chiang, B. Lu and I. Castillo, "Big data analytics in chemical engineering," *Annual Review of Chemical and Biomolecular Engineering,* vol. 8, pp. 63-85, 2017.

[7] P. M. Piccione, "Realistic interplays between data science and chemical engineering in the first quarter of the 21st century, part 2: Dos and don'ts," *Chem. Eng. Res. Design,* vol. 169, pp. 308-318, 2021.

[8] P. M. Piccione, "Realistic interplays between data science and chemical engineering in the first quarter of the 21st century: Facts and a vision," *Chem. Eng. Res. Design,* vol. 147, pp. 668-675, 2019.

[9] B. Braun *et al*, "Data science challenges in chemical manufacturing," *IFAC Preprints,* 2020.

[10] J. Gossage *et al*, "Integrating best practice pedagogy with computer aided modeling and simulation to improve undergraduate chemical engineering education," in *2001 Annual Conference,* 2001, .

[11] S. E. Brownell and K. D. Tanner, "Barriers to faculty pedagogical change: Lack of training, time, incentives, and… tensions with professional identity?" *CBE—Life Sciences Education,* vol. 11, *(4),* pp. 339-346, 2012.

[12] C. Henderson, "The challenges of instructional change under the best of circumstances: A case study of one college physics instructor," *American Journal of Physics,* vol. 73, *(8),* pp. 778-786, 2005.

[13] I. Sadler, "The challenges for new academics in adopting student-centred approaches to teaching," *Studies in Higher Education,* vol. 37, *(6),* pp. 731-745, 2012.

[14] K. Bennie and K. Newstead, "Obstacles to implementing a new curriculum," in *Proceedings of the National Subject Didactics Symposium,* 1999, .

[15] R. Tang and W. Sae-Lim, "Data science programs in US higher education: An exploratory content analysis of program description, curriculum structure, and course focus," *Educ. Inf.,* vol. 32, *(3),* pp. 269-290, 2016.

[16] 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, .

[17] P. Sánchez-Holgado *et al*, "Formation in data science in secondary education: Big data as a transversal competence," in *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality,* 2020, .

[18] R. Israel-Fishelson *et al*, "Preparing K-12 students to meet their data: Analyzing the tools and environments used in introductory data science contexts," in *Proceedings of the 2023 Symposium on Learning, Design and Technology,* 2023, .

[19] I. Song and Y. Zhu, "Big data and data science: Opportunities and challenges of iSchools," *Journal of Data and Information Science,* vol. 2, *(3),* 2017.

[20] P. Bonfert-Taylor *et al*, "Infusing data science into the undergraduate STEM curriculum," in *2022 ASEE Annual Conference & Exposition,* 2022, .