ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #40885

Infusing Data Analytics Skills into Project-Based Learning for Computer Science and Engineering Education

Dr. Xiang Zhao, Alabama A&M University

Dr. Xiang (Susie) Zhao, Professor in the Department of Electrical Engineering and Computer Science at the Alabama A&M University, has over 20 years of teaching experience in traditional on-campus settings or online format at several universities in US andaboard. Her teaching interests include programming languages, high performance algorithm design, data science, and evidence-based STEM teaching pedagogies. Her recent research work has been funded by DOE, USED and NSF.

Dr. Showkat Chowdhury, Alabama A&M University

Dr. Showkat Chowdhury is a Professor in the Department of Mechanical Engineering at Alabama A&M University in Huntsville, AL. Dr. Chowdhury has extensive background in teaching undergraduate and graduate students in Mechanical Engineering, and performing research in the fields of Computational Fluid Dynamics, Renewable Energy, Nano-Technology, Heat & Mass Transfer, and Combustion. He is managing multi-million dollar external research grants from NSF and DoED as PI. Previously, he worked as a Professor at Bangladesh University of Engineering & Technology (BUET) and at University of Brighton, U.K. He also worked in the Research Division of Corning Inc. He earned his Ph.D. and M.S. in Mechanical Engineering from Clarkson University, New York and B.S. in Mechanical Engineering from BUET.

Prof. Tamara Chowdhury, Alabama A&M University

Tamara Chowdhury is an Assistant Professor of Construction Management Program in the Department of Mechanical & Civil Engineering and Construction Management at Alabama A&M University in Huntsville, AL. Prof. Chowdhury has extensive background in teaching undergraduate students in the Department of Civil Engineering, Construction Management and conducting external funded research. She also worked for a multinational research organization for many years.

Infusing Data Analytics Skills into Project-Based Learning for Computer Science and Engineering Education

Abstract

A workforce equipped with essential data analytics skills is crucial to maintaining the United States' competitiveness and security in the global arena. Data analytics skills are in high demand in order to generate data-driven, robust solutions to solving the biggest challenges that our society is facing today. This paper presents the technical approach that facilitates the infusion of data analytics skills into Project-Based Learning (ProjBL) pedagogies for selected Computer Science and Engineering courses. The experimental results obtained from three pilot courses in computer science, mechanical engineering, and construction management showed the effectiveness of the approach in enhancing students' ability to understand data, analyze data and develop data-driven solutions. The feedback from student surveys also provided a satisfactory outcome as expected. This pilot study intends to share the project team's experience and lessons learned with the STEM education community.

Keywords

Data Analytics, Data Science, Project-Based Learning, STEM Education

Introduction

Data analytics is the process of inspecting, cleaning, transforming, and visualizing data with the goal of discovering insightful and critical information for decision making [1]. The integration of data analytics in STEM education has had a profound impact on the advancement in every sector of industries, government, and academia today. A workforce equipped with essential data analytics skills is crucial to maintaining the United States' competitiveness and security in the global arena. Data analytics skills are in high demand in order to generate data-driven, robust solutions to solving the biggest challenges that our society is facing today [2][3]. However, the current higher education curricula are still lacking in data analytics especially at the undergraduate level at MSIs/HBCUs. Hence, there is an urgent need to explore innovative approaches to infuse data analytics skills into STEM education.

In this paper, a pilot study that includes the course design strategies to integrate data analytics into project-based learning in selected computer science and engineering courses are described. The findings and lessons learned from this study are also presented with the intention to share our experience with the instructors and administrators to advance data science education at MSIs/HBCUs.

Related Work

In the past decade, educators and researchers realized the importance of data analytics in transforming higher education. It was shown by Maier-Hein et al. [4] that incorporating data analytics and exposing students to real-world datasets improved their critical thinking. More impressively, data science education encourages students to explore STEM careers and also

provides a strong foundation for further education and future employment opportunities as studied by Marques et al. [5].

Data Analytics in STEM Education

Brown et al. [6] integrated data analytics in engineering education to address technical requirements from a multicomplex environment perspective concept using data analytics tools such as IBM Watson Analytics. The results obtained from a multi-complex environment have aided students and improved their decision approach to quantify data accuracy and project requirements. The integration of analytics tools fostered the engineering students the ability to forecast requirements and create new methods critical to their engineering design.

Data analytics was also added to a core course on product manufacturing in the industrial engineering curriculum [7]. The pedagogical method was developed by first analyzing and comparing product manufacturing processes and data analytics techniques. Then the result of this analogy was used to develop a teaching and learning method for data analytics. For implementation and validation purposes, a Project Based Learning (ProjBL) approach was adopted, in which students used the methodology to complete real-world data analytics projects. Data from students' grades shows that this approach improved their performance [7].

Recently, Bonfert-Taylor et al. [8] developed multiple data science modules for various engineering and social science courses using R and MATLAB tools. In addition, they offered data science internships and interdisciplinary data science projects as experiential learning opportunities. A student survey tool was developed to measure the student attitude toward data science in four aspects: Interest, Value, Career, and Self-Efficacy.

However, learning fundamentals of data analytics is a form of complex learning, as many concepts and theory are abstract, counterintuitive, and challenging to understand, thus not accessible to learners, even not to instructors at HBCUs due to time and resource constraints. Therefore, research is highly needed to address this gap, which indeed motivates our study.

Project-Based Learning (ProjBL)

Active learning pedagogies have been widely touted as beneficial to student learning [9], retention [10], and engagement. Also, learning outcomes are better when students are active participants in the learning process [11-12], especially for underrepresented students [13].

Teaching is an art of encouraging students to become active learners and awakening their enthusiasm to explore and absorb new knowledge and skills. On the other hand, learning is a dynamic process in which both the teacher and students should actively participate, exchange views, and ask/answer questions in an engaging atmosphere [14]. Student engagement has been shown to be a key factor in student retention in the STEM fields[15]. It has been abundantly demonstrated that pedagogical methods that promote conceptual understanding through interactive engagement of students are far more effective than traditional didactic instructional methods. Almost all of the newly developed methods on teaching and learning have concentrated on student-centered, inquiry-based approaches [16].

One of the successful evidence-based designs for teaching science and engineering courses is **Project-Based Learning (ProjBL)[17].** ProjBL is an instructional methodology that encourages students to learn and apply knowledge and skills through an engaging experience. It provides students the opportunities for deeper learning and for the development of important non-cognitive skills for college and career readiness. Students drive their learning by inquires, research and collaboration toward the completion of the projects. The role of the instructor shifts from a content-deliverer to a facilitator and mentor. In ProjBL, students form a group and work more independently to complete the projects with the instructor providing support only when needed. Students are encouraged to make their own decisions about the project topics and how to complete. One of the main goals of ProjBL is to engage students in deep learning throughout the full project life cycle [17].

Our Approaches

A key objective of our study is to leverage active learning in conjunction with data analytics technologies to increase student engagement and learning and to help instructors more effectively and efficiently teach concepts, theory and problem solving skills. This pilot study will support instructors by providing an insightful understanding of the students' successes and challenges when dealing with real-world problems using data analytics.

A team of faculty members in computer science, mechanical engineering and construction management at Alabama A&M University implemented ProjBL instructional practices in three core courses in STEM curricula by integrating data analytics skills in Fall 2023. This pilot study has focused on: (1) designing the three courses in computer science, mechanical engineering and construction management by adding data analytics components into team projects; (2) incorporating engagement strategies in lectures and laboratory activities that promote active student interaction, critical thinking, and problem-solving; and (3) conducting assessment and surveys to gather feedback from students. This section mainly describes the details of this pilot study.

The team has followed the logic model in Figure 1 that has been established and tested in our previous study for enhancing STEM gateway courses with evidence-based pedagogies [18].

During Fall 2023, faculty catalysts in our project team selected three courses in STEM disciplines to integrate data analytics into ProjBL: CS405 – "Linux with Application Programming", ME360 – Fluid Mechanics, and CMG461 – Capstone Project. The project team intended to test how the strategies are effective in different projects and how students respond to the pedagogies and the data analytics components in the projects.

After faculty catalysts received training on how to apply various evidence-based teaching pedagogies and teaching/learning management tools, they redesigned the selected courses.

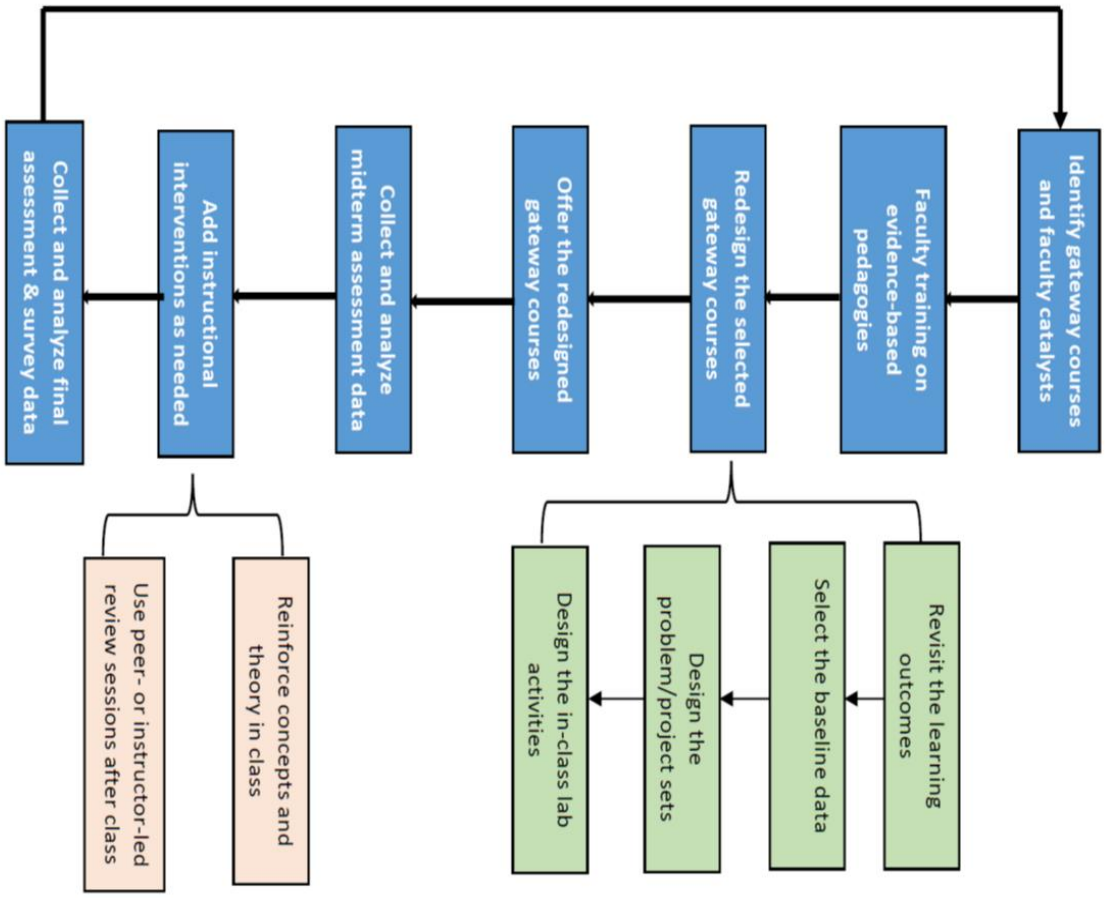

Figure 1. The Logic Model

The ProjBL activities follow a three-phase pedagogical approach in CS405: 1) Conceptual Phase (Learning): the students are first introduced to Linux programming and data analytics fundamentals, and features of sample datasets. Then the students are exposed to the supercomputer Linux environment and software tools. 2) Experimentation Phase (Practice): The students are given the opportunity to implement and employ one data analytics algorithm to process one sample dataset. The students are required to vary the parameters and observe the results. 3) Application Phase (Evaluation): The students are asked to employ both Linux and data science algorithms to process multiple datasets in the team project. The students then use the accuracy metrics and collect the results for comparison for different algorithm parameters. Students formed project team and made their own decisions about how to split the work, when to meet, and how to complete the project tasks. The instructor provided support mostly during the Conceptual Phase and Experimentation Phase, but monitored the progress during the Application Phase. Figure 2 is an example of the student team project in CS405. Students are expected to predict the companies' profit according to their R&D Spend, Admin Cost, Marketing Spend, and Locations.

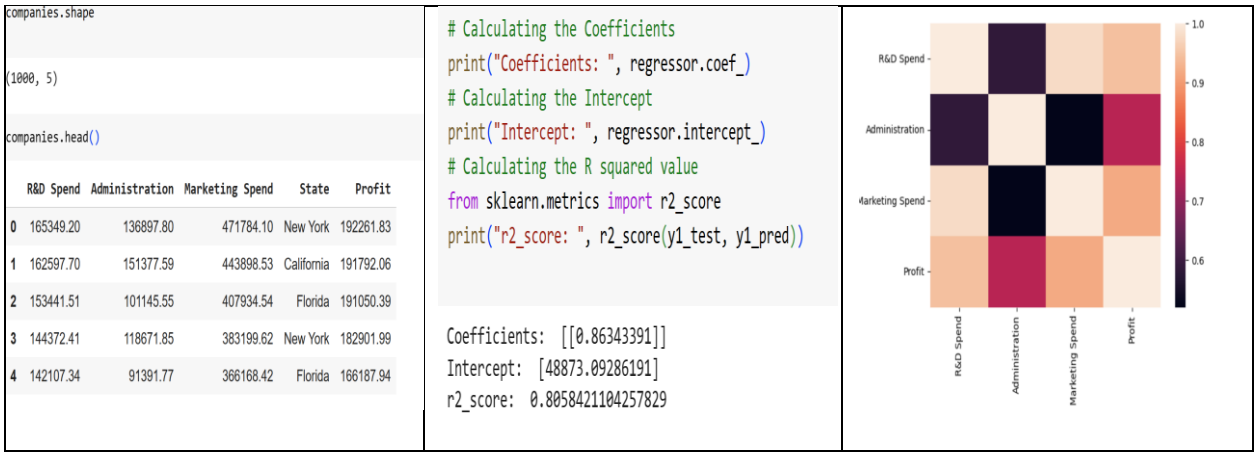

Figure 2. A screenshot of the dataset: 1000 companies' data on R&D Spend, Admin Cost, Marketing Spend, Locations, and Profit(left); An Example of Linear Regression Analysis(middle); and Heatmap to illustrate the correlation between the attributes in the dataset (right)

Similarly, the Project Based Learning activities follow a three-phase pedagogical approach in ME360: 1) Conceptual Phase (Learning): the students are first introduced to the theory required for calculation and analysis of the project performance, for example, for the Design of a Pump System to deliver water at a certain altitude. The students are required to select the pipe materials, pipe diameters, calculate losses and pump capacity required, consider the initial and operating cost, etc. The students are exposed to software tools like Excel for calculation and plotting. 2) Experimentation Phase (Practice): The students are given the opportunity to select one set of design parameters and complete the calculations for one sample dataset. 3) Application Phase (Evaluation): The students are asked to vary the design parameters and analyze the results. They use Excel spreadsheet for calculation by varying design parameters and plot their trend lines for comparison and decision making. Students formed project team and made their own decisions about how to split the work, when to meet, and how to complete the project tasks. The instructor provided support mostly during the Conceptual Phase and Experimentation Phase, but monitored the progress during the Application Phase. Figure 3 shows an example of the team project in ME360.

The Project Based Learning activities follow a three-phase pedagogical approach as well in CMG461: 1) Conceptual Phase (Learning): The students have to select a realistic project based on the previous construction courses with the Instructor's guidance, which is similar to the General Contractors related work. All work to be done by the General Contractors or Subcontractors should be estimated, totaled and marked-up. This can be done on the regular Cost Sheets and Summary Sheets in Excel Spreadsheet or an estimating software named Sage Timberline. 2)Experimentation Phase (Practice): The Students first work on the detailed estimation in Excel spreadsheet in different divisions and sub-divisions. 3)Application Phase (Evaluation): The students then work on the detailed estimation in estimating software named Sage Timberline. Sage Timberline Estimating software is powered by trade-specific cost databases, enabling takeoffs and detailed estimation of each division and sub-division. Students input their own project costs into a prebuilt database or choose a database pre-populated with costs. Since Sage Timberline uses prebuilt database there can be some variations with the calculations in Excel spreadsheet. The students include a discussion section and discuss about

the variation in the two approaches. Figure 4 shows a screenshot of the sample dataset in Excel and Sage Timberline for CMG 461 Project.

	A	B	C	D	E
1	Pipe Diameter (in)	Pump Power (hp)	Pump Cost ($)	Pipe Cost ($)	Total Cost ($)
2	2	20.6	8,000	1,430	9,430
3	3	14.3	6,000	1,850	7,850
4	4	11.3	5,000	2,353	7,353
5	5	11.1	5,000	3,100	8,100
6					

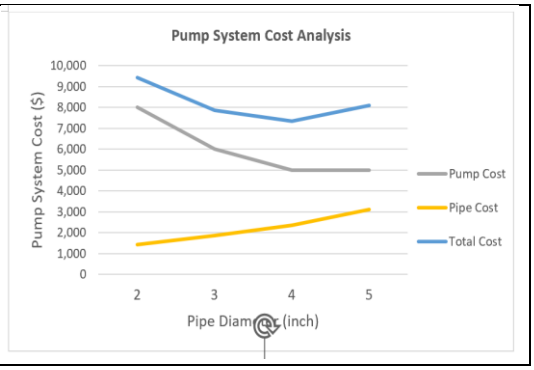

Figure 3. A screenshot of sample Dataset in Excel Spreadsheet for ME360 Project(left) and A sample Excel Plot for ME360 Project Cost Analysis(right)

Software	Labor Cost	Material Cost	Other Cost
Excel Spreadsheet	25 %	45 %	30 %
Sage Timberline	22 %	50 %	28 %

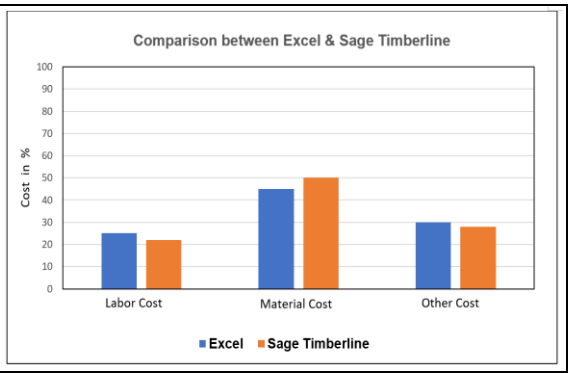

Figure 4. A screenshot of sample Dataset (left) and A sample Excel Plot of Data Comparison for CMG461 Project(right)

The midterm assessments have been conducted to monitor the students' progress and performance, followed by an immediate adjustment of the instructor's intervention as needed. For example, from the tests, class discussion, and midterm exam, students in CS405 demonstrated weaker understanding on some concepts and skills such as using arithmetic expressions and selection statements in shell scripts and vi editor in supercomputer environment. The instructor added in-class lab times to reinforce the related concepts and office hours after class for Q&A. In addition, the data on student project completion rate, exit survey and final exam were collected to evaluate and assess the outcomes of the adopted pedagogies. The pilot study results are presented in the next section.

Results and Discussion

This section summarizes the experimental results obtained from this study. A comparison was also accomplished to verify the effectiveness of the methodologies using the base line data.

CS405 – "Linux with Application Programming" is a core course in the computer science curriculum at Alabama A&M University. Table 1 includes the student assessment results in CS405 regarding the learning outcomes and the ABC rates (only grades A,B, and C are considered as "Pass" according to the computer science curriculum in the university undergraduate bulletin). The base line data in Spring 2023 (without ProjBL) and the new data in Fall 2023 (with ProjBL and data analytics) are compared. The same instructor taught the course using the same syllabus. And the same course learning outcomes have been assessed. It has been observed that the percentages of "Satisfactory" students for both learning outcomes a) and b) have been improved in Fall 2023 compared with the results in the baseline data in Spring 2023 (without PorjBL). Meanwhile, the ABC rates are also improved compared with the baseline data in Spring 2023. In addition, all teams completed their projects on time.

Table 1. Assessment Data in CS405- "Linux with Application Programming"

Learning Outcomes: a) Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline. b) Apply computer science theory and software development fundamentals to produce computing-based solutions.

Semester	Learning Outcome a		Learning Outcome b		ABC Rates	Project Completion Rate
	Satisfactory	Unsatisfactory	Satisfactory	Unsatisfactory		
Spring 2023 (Baseline data, without ProjBL,)	80 %	20%	80 %	20%	80%	n.a.
Fall 2023 (with ProjBL)	92%	8%	92%	8%	92%	100%

In addition to the formal assessment, student surveys have been conducted to provide the evaluation and feedback in three selected courses. Table 2 summarizes the student survey results with 16 participants in CS405, which indicated positive feedback and favorable attitude from students toward the data analytics components in team projects.

Table 2. CS405 Student Survey Summary

Survey Questions	Feedback Scales			
	Strongly agree	Agree	Disagree	Strongly Disagree
Use of engaged learning techniques in this course helped me to understand concepts & principles, apply skills learned to solve given problems, analyze/evaluate possible solutions.	69%	25%	6%	0%
This learning experience helped me to improve my study habits/interest such as reviewing materials, completing work on time, discussing with my peers, understand data, etc.	75%	25%	0%	0%
In the team project, using the Linux skills (commands and scripts) and Python programs (decision-tree, regression) to analyze the data files (such as find the patterns, data trends, classification) helped me gain the experience in data science/data analytics.	75%	25%	0%	0%

Integrating the fundamental skills of data analytics is helpful and impactful to my study and future career.	75%	25%	0%	0%
Overall, the engaged learning techniques helped me to learn better.	75%	25%	0%	0%

Similar assessment results are obtained in ME360 (in Table 3) and CMG461 (in Table 4).

Table 3. Assessment Data vs. Program-Level Learning Outcomes for ME360 with Project Based Learning

Learning Outcomes: a) An ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science and mathematics. b) An ability to design a system, component or process to meet desired needs".

Semester	Learning Outcome a		Learning Outcome b		ABC Rates	Project Completion Rate
	Satisfactory	Unsatisfactory	Satisfactory	Unsatisfactory		
Fall 2022 (Baseline data, without ProjBL,)	80 %	20%	80 %	20%	85%	n.a.
Fall 2023 (with ProjBL)	94%	6%	97%	3%	94%	100%

Table 4. Assessment Data vs. Program-Level Learning Outcomes for CMG 461 with Project Based Learning

Learning Outcomes: a) An ability to understand ethical and professional responsibilities and the impact of technical and/or scientific solutions in global, economic, environmental, and societal contexts. b) An ability to function effectively on teams that establish goals, plan tasks, meet deadlines, and analyze risk and uncertainty.

Semester	Learning Outcome a		Learning Outcome b		ABC Rates	Project Completion Rate
	Satisfactory	Unsatisfactory	Satisfactory	Unsatisfactory		
Fall 2021 (Baseline data, without ProjBL,)	75 %	25 %	80 %	20%	100%	n.a.
Fall 2023 (with ProjBL)	90 %	10 %	95%	5%	100%	100%

Conclusion and Future Work

Three core courses in computer science, mechanical engineering, and construction management have been designed and offered by integrating data analytics into ProjBL pedagogies in this pilot study. The student assessment results indicated the effectiveness of the methodologies employed in this study, especially in prompting critical thinking, enhancing the ability to understand data,

analyze data and develop data-driven solutions to real-world problems, and improving engagement and retention. In addition, positive feedback has been obtained from the student survey data on those courses, which shows most of the students are in favor of learning data analytics skills in the projects. Future study will continuously apply the same strategies to infusing data analytics into ProjBL and collect more data to further verify the effectiveness on engaging students and developing the critical skills for success.

Acknowledgement

This study is sponsored by Department of Education Award # P120A210012 and National Science Foundation Award # 2129961. The project team is thankful to the support of the Program Managers.

References

1. J. Kennedy, P. Abichandani and A. Fontecchio, "An initial comparison of the learning propensities of 10 through 12 students for data analytics education," *IEE Frontiers in Education Conference,* Oklahoma City, OK, pp. 916-918, 2013.
2. Hirsch, D. D. (2013). The glass house effect: Big Data, the new oil, and the power of analogy. *Me. L. Rev.*, *66*, 373.
3. Iqbal, R., Doctor, F., More, B., Mahmud, S., & Yousuf, U. (2020). Big data analytics: Computational intelligence techniques and application areas. *Technological Forecasting and Social Change, 153,* 119253.
4. Maier-Hein, L., Eisenmann, M., Sarikaya, D., März, K., Collins, T., Malpani, A., Fallert, J.,Feussner, H., Giannarou, S., Mascagni, P., Nakawala, H., Park, A., Pugh, C., Stoyanov,D., Vedula, S. S., Cleary, K., Fichtinger, G., Forestier, G., Gibaud, B., … Speidel, S. (2022). Surgical data science – from concepts toward clinical translation. *Medical Image Analysis, 76,* 102306. https://doi.org/10.1016/j.media.2021.102306
5. Marques, L. S., Gresse Von Wangenheim, C., & Hauck, J. C. R. (2020). Teaching Machine Learning in School: A Systematic Mapping of the State of the Art. *Informatics in Education,* 283–321. https://doi.org/10.15388/infedu.2020.14
6. Brown, W., Zhang, L., Sharma, D. K., Jin, Y., Dabipi, I., Zhu, W., & Lawrence, E. (2018, October). The Integration of Data Analytics to Assess Multi-Complex Environments of Research to Practices in Engineering Education. In *2018 IEEE Frontiers in Education Conference (FIE)* (pp. 1-6). IEEE.
7. Aqlan, F., & Nwokeji, J. C. (2018, October). Applying product manufacturing techniques to teach data analytics in industrial engineering: a project based learning experience. In *2018 IEEE Frontiers in Education Conference (FIE)* (pp. 1-7). IEEE.
8. Bonfert-Taylor, P., Ray, L., Pauls, S., Loeb, L., Sankey, L., Busch, J., & Hickey, T. (2022, August). Infusing Data Science into the Undergraduate STEM Curriculum. In *2022 ASEE Annual Conference & Exposition*.
9. Greer, T., Hao, Q., Jing, M., & Barnes, B. (2019). On the Effects of Active Learning Environments in Computing Education. *Proceedings of the 50th ACM Technical*

Symposium on Computer Science Education, 267–272. https://doi.org/10.1145/3287324.3287345
10. Latulipe, C., Rorrer, A., & Long, B. (2018). Longitudinal Data on Flipped Class Effects on Performance in CS1 and Retention after CS1. *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, 411–416. https://doi.org/10.1145/3159450.3159518
11. Hartikainen, S., Rintala, H., Pylväs, L., & Nokelainen, P. (2019). The Concept of Active Learning and the Measurement of Learning Outcomes: A Review of Research in Engineering Higher Education. *Education Sciences*, *9*(4), 276. https://doi.org/10.3390/educsci9040276
12. Shi, Y., Yang, H., MacLeod, J., Zhang, J., & Yang, H. H. (2020). College Students' Cognitive Learning Outcomes in Technology-Enabled Active Learning Environments: A Meta-Analysis of the Empirical Literature. *Journal of Educational Computing Research*,*58*(4), 791–817. https://doi.org/10.1177/0735633119881477
13. Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N.,Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., … Freeman, S.(2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math. *Proceedings of the National Academy of Sciences*, *117*(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117
14. Chang, M. J., Cerna, O., Han, J., & Sáenz, V. The contradictory roles of institutional status in retaining underrepresented minorities in biomedical and behavioral science majors. *The Review of Higher Education*, 31(4), 433-464 (2008).
15. Watkins & Mazur E. (2013) Retaining students in science, technology, engineering, and mathematics (STEM) majors. Journal of College Science Teaching, 42(5), 36–41.
16. Karl A. Smith, "Inquiry-Based Cooperative Learning, Sigma Xi Conference Proceedings, Resha- ping Undergraduate Science and Engineering Education: Tools for Better Learning, p. 53 (1999)
17. Veselov, G.E., Pljonkin, A. P., and Fedotova, A.Y., Project-based learning as an effective method in education, Proceedings of the 2019 International Conference on Modern Educational Technology, pp 54–57 (2019) https://doi.org/10.1145/3341042.3341046
18. Zhao, X., Chowdhury, S., and Chowdhury, T., 2020, "Integrating Evidence-Based Learning in Engineering and Computer Science Gateway Courses", Proceedings of 2020 ASEE Annual Conference and Exposition.