Engineering Educators Bringing the World Together
2025 ASEE Annual Conference & Exposition
Palais des congrès de Montréal, Montréal, QC • June 22–25, 2025 ASEE
Paper ID #46964

# Enhancing Data Science Education for Critical Infrastructures with Project-Based Learning

**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 and aboard. Her teaching and research 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, NASA and NSF.

**Dr. Mebougna Drabo, Alabama A&M University**

Dr. Mebougna L. Drabo is currently a professor of Mechanical Engineering at Alabama A&M University (AAMU). He is the chair of the department of Mechanical & Civil Engineering and Construction Management at AAMU. He is also serving as the director of the Alabama EPSCoR Agency for the Department of Energy. He joined AAMU in 2012, leveraging his expertise in teaching and mentoring STEM students while fostering on-campus research and DOE Lab internships. Currently, he directs the DOE/NNSA's Consortium, SPINS and is working on integrating radiation detection systems into cyber manufacturing environments.

His research interests include STEM education, Additive Manufacturing, Thermoelectric Devices for Energy Harvesting, Digital Twinning Technology, Nuclear Radiation Detectors, Nuclear Security and Safety, Small Nuclear Modular Reactors (SMR), Material Characterization (X-ray Photoelectron Spectroscopy & Infrared Microscopy), Nanotechnology, Data Analytics and Visualization, Biofuels Applications, Computational Fluid Dynamics analysis, Heat Transfer, Energy Conservation in building, and Multi Fuel Optimization.

©American Society for Engineering Education, 2025

# Enhancing Data Science Education for Critical Infrastructure Security with Project-Based Learning


## Abstract

A workforce equipped with essential data science skills is crucial for maintaining the United States' competitiveness and strengthening infrastructure security in today's highly interconnected digital world. By analyzing large volumes of data, data science techniques can identify patterns and anomalies that may indicate potential security threats. Enhancing data science education helps students generate data-driven, robust solutions to solving the security problems in many infrastructures including transportation systems, healthcare systems, power plants, etc. Machine learning algorithms, for example, can predict and detect cyber-attacks before they occur. This paper presents the technical approach that aims to boost data science education using the Project-Based Learning (ProjBL) pedagogy. In this study, the data science projects are designed to employ different machine learning models to analyze various datasets related to Internet of Things (IOTs) attacks, power plant operations, and public health, which exposes the students to real-world security challenges. The experimental results obtained from this pilot study show the effectiveness of the approach in enhancing students' ability to understand data, analyze data and develop data-driven solutions to various problems. The student survey also indicates favorable/positive feedback toward this intervention as expected. This 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 science skills is crucial for maintaining the United States' competitiveness and strengthening infrastructure security in today's highly interconnected digital world. By analyzing large volumes of data, data science techniques can identify patterns and anomalies that may indicate potential security threats [2][3]. Enhancing data science education will help students generate data-driven, robust solutions to solving the security problems in critical infrastructures including but not limited to transportation systems, healthcare systems, power plants, etc. Machine learning algorithms, for example, can predict and detect cyber-attacks before they occur. However, current higher education curricula still lack sufficient data science content, especially for cybersecurity at the undergraduate level at MSIs/HBCUs. Hence, there is an urgent need to explore innovative approaches to enhance data science education for infrastructure security.

In this paper, the technical approach that aims to boost data science education using the Project-Based Learning (ProjBL) pedagogy is presented. In this study, three data science projects are

© American Society for Engineering Education, 2025

designed to employ different machine learning models to analyze various datasets related to Internet of Things (IOTs) attacks, power plant operations, and public healthcare systems, which exposes the students to real-world security challenges. 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)

© American Society for Engineering Education, 2024

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 science technologies to increase student interests and exposure to real-world problems in critical infrastructure security. It also aims to help instructors more effectively and efficiently introduce concepts, theory and problem-solving skills to students. This pilot study will support instructors by providing an insightful understanding of the students' successes and challenges when dealing with real-world problems using ProjBL.

A team of faculty members in computer science and mechanical engineering at Alabama A&M University implemented ProjBL instructional practices in CS413/520-"Introduction to Data Science", a co-listed undergraduate/graduate course in Fall 2024. This pilot study has focused on: (1) designing and implementing the data science 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 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].

© American Society for Engineering Education, 2025

After faculty members received training on how to apply ProjBL pedagogies and teaching/learning management tools, they redesigned the data science course by implementing ProjBL.

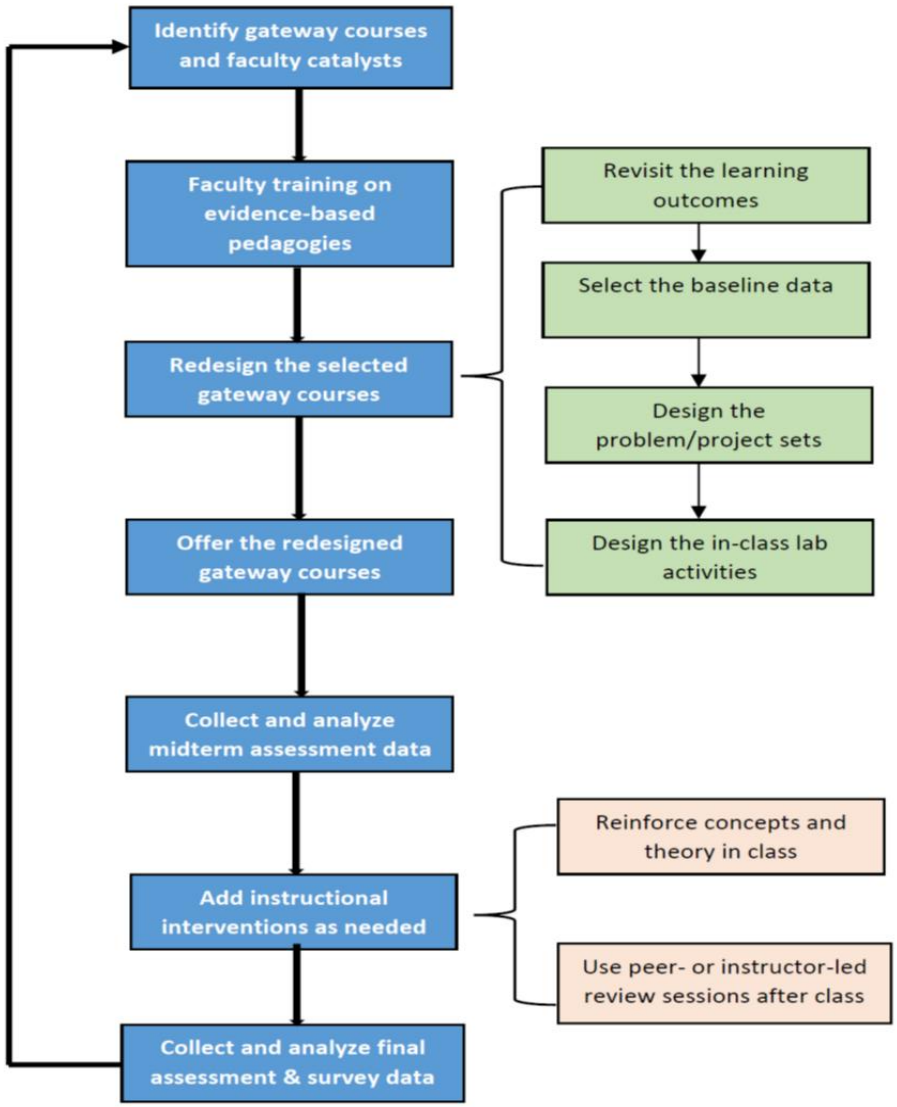



Figure 1. The Logic Model

The ProjBL activities follow a three-phase pedagogical approach in the course: 1) Conceptual Phase (Learning): the students are first introduced to data science fundamentals and features of sample datasets. Then the students are exposed to the Python programming and software development tools such as Microsoft Azure [19] and Google Colab. 2) Experimentation Phase (Practice): The students are given the opportunity to implement and employ one data science algorithm to process one small scale sample dataset. The students are required to vary the parameters and observe the results. 3) Application Phase (Evaluation): The students are asked to

© American Society for Engineering Education, 2024

employ data science algorithms to process real-world large datasets in the team projects. The students then use the accuracy metrics and collect the results for comparison for different algorithm parameters. Students formed the project teams 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.

The team projects focus on applying data science skills to solve the real-world problems in one of the following areas:

a) Analyze the cyber threats in IOT datasets in Kaggle[20]: Detect and predict the presence of Trojan Horses, a type of malware designed to infiltrate and compromise computer systems by disguising itself as legitimate software. Leveraging network traffic data, the goal is to identify patterns and anomalies indicative of Trojan activities. Key Features of the datasets include network activity metrics, timestamps, and labels (Benign vs. Trojan). Students can use classification models in this task.

b) Analyze/detect the anomalies in power plant operation datasets in Kaggle[20]: distinguish the natural events and attacks in the power plants; also predict sensor values to anticipate performance changes. This work supports proactive alert systems and helps operators prepare for, respond to, and prevent security incidents. Students can use classification models to identify the attacks and/or use regression models for senor value predication in this task.

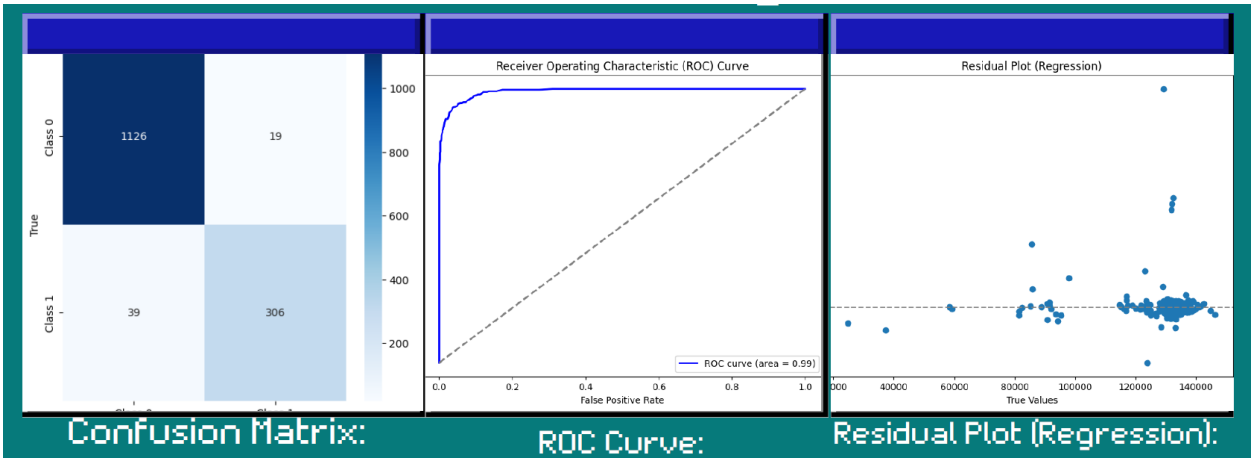
c) Analyze the cyber threats in healthcare system datasets in Kaggle[20]: Analyze the impact and implications from data breaches on healthcare organizations as well as their patients. This dataset contains detailed information on US health data breaches that affected individuals. The key features in this dataset include Breach Submission Date, Type of Breach, Location of Breached Information, Business Associate Present and Web Description. This work helps healthcare professionals to learn more about health care security best practices and prevention against future data-related incidents. Students can use classification models and/or regression models in this task.

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 CS413/CS520 demonstrated weaker understanding on some concepts and skills such as using Python packages in model training & fitting. 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.

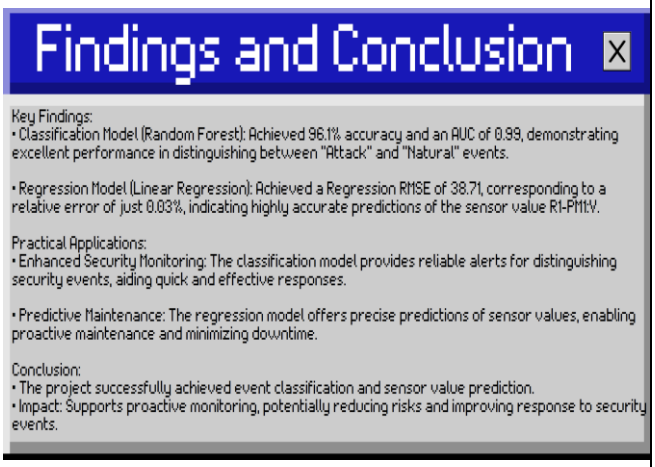
© American Society for Engineering Education, 2025




(a)

```python
# Classification Model: Random Forest Classifier
clf = RandomForestClassifier(n_estimators=100, random_state=42)
clf.fit(X_train, y_class_train)

# Compute ROC curve
fpr, tpr, thresholds = roc_curve(y_class_test, clf.predict_proba(X_test)[:, 1])
roc_auc = auc(fpr, tpr)

# Predict and evaluate regression
y_reg_pred = reg.predict(X_test_reg)
regression_rmse = root_mean_squared_error(y_reg_test, y_reg_pred)
print("Regression RMSE:", regression_rmse)

y_class_pred = clf.predict(X_test)
classification_accuracy = accuracy_score(y_class_test, y_class_pred)
print("Classification Accuracy:", classification_accuracy)

# Classification Model: Random Forest Classifier
clf = RandomForestClassifier(n_estimators=100, random_state=42)
clf.fit(X_train, y_class_train)
```

(b)




(c)

Figure 2. An example of the student project outcomes: (a) The metrics generated from the model; (b) The Python code snippets of the random forest model; and (c) The summery of key findings in the project presentation by students

Figure 2 shows an example of the student project outcomes, which include the implementation code snippets, results, and presentation summary. Classification and regression models are used to analyze the power plant operation datasets downloaded from Keggle [20]. Student projects were implemented on Microsoft Azure or Google Colab. Table 1 includes the student assessment results in CS413/520 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). The base line data in Fall 2023 (without ProjBL) and the new data in Fall 2024 (with ProjBL) 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 2024 compared with the results in the baseline data in Fall 2023 (without PorjBL). Meanwhile, the ABC rates also remained high compared with the baseline data in Fall 2023. In addition, all teams completed their projects on time (i.e., the completion rate is 100%).

© American Society for Engineering Education, 2024

Table 1. Assessment Data in CS413/520- "Introduction to Data Science"

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 | | |
| Fall 2023 (Baseline data, without ProjBL,) | 83 % | 17% | 83 % | 17% | 92% | n.a. |
| Fall 2024 (with ProjBL) | 96% | 4% | 92% | 8% | 96% | 100% |

In addition to the formal assessment, student surveys have been conducted to provide the evaluation and feedback in the course. Table 2 summarizes the student survey results with 18 participants in the survey, which indicated positive feedback and favorable attitude from students toward the data science team projects.

Table 2. 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. | 83% | 17% | 0% | 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. | 89% | 11% | 0% | 0% |
| In the team project, using the data science skills (linear regression, classification, clustering, NN, etc.) and Python programs to analyze the datasets (such as find the patterns, data trends, anomalies) helped me gain the hands-on experience in data science and analytics. | 89% | 11% | 0% | 0% |
| In the team project, using the data science skills (linear regression, classification, clustering, NN, etc.) and Python programs to analyze the datasets (such as find the patterns, data trends, anomalies) helped me better understand the concepts and theory in data science and analytics. | 89% | 11% | 0% | 0% |
| In the team project, solving real-world problems by applying data science skills is helpful and inspiring to my future study and career. | 94% | 6% | 0% | 0% |
| Overall, the engaged learning techniques helped me to learn better. | 89% | 11% | 0% | 0% |

© American Society for Engineering Education, 2025

## Conclusion and Future Work

The data science course in Alabama A&M University has been enhanced by integrating ProjBL pedagogies to equip students with the essential data science skills to tackle the real-world critical infrastructure security problems 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. In addition, positive feedback has been obtained from the student survey data on the interventions, which shows most of the students are in favor of applying data science skills in real-world problem solving in the projects. The intervention provides an opportunity to increase their interests and exposures in data science for infrastructure security areas. Future study will continuously apply the same strategies to enhance data science education using 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 National Science Foundation (Award#2321112) and DOE/NNSA MSIPP. This research project has also benefitted from the Microsoft Accelerating Foundation Models Research (AFMR) grant program.

## 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. Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data, 7*, 1-29.
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.

**© American Society for Engineering Education, 2024**

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.
19. Microsoft Azure: https://azure.microsoft.com/en-us/
20. Kaggle: https://www.kaggle.com/


**© American Society for Engineering Education, 2025**