Leveraging Predictive Analytics on ABET Assessment Data to Forecast Attainment of Complex Civil Engineering Student Outcomes in an HBCU

Obiageli Gertrude Adaku Nwachukwu, Morgan State University

Obiageli Nwachukwu is a Graduate Research Assistant at the Centre for Equitable Artificial Intelligence and Machine Learning Systems, Morgan State University, where she specializes in data-driven modeling, machine learning, and applied bioinformatics. Her research spans predictive analytics, digital health adoption, and the development of equitable AI frameworks for underserved communities. She has contributed to NTIA-funded projects focused on broadband access, telehealth readiness, and community-centered digital equity initiatives. With a background in pharmacy and advanced training in computer science (bioinformatics), she integrates domain expertise in health systems with computational methods to build predictive models that support informed decision-making in public health and technology adoption. Her broader interests include clinical trial outcome prediction, translational data science, and AI-enabled solutions for improving health access and outcomes.

Julius Ogaga Etuke, Morgan State University

Julius Ogaga Etuke is a seasoned civil engineer and Ph.D. Candidate in Civil Engineering at Morgan State University, where his research focuses on sustainable and resilient infrastructure engineering. With over 15 years of consulting and project leadership experience, primarily in Nigeria's infrastructure landscape, he brings a robust, practical perspective to academic inquiry. His doctoral research is at the cutting edge of transportation resilience, integrating virtual and augmented reality (VR/AR), climate adaptation modeling, and data-driven solutions to improve disaster preparedness and risk mitigation in complex transportation systems. As a COREN-registered engineer and member of the Nigerian Society of Engineers, his work emphasizes impactful research and consultancy that bridges structural design principles with urgent sustainability and resilience challenges.

Dr. Oludare Adegbola Owolabi P.E., Morgan State University

Dr. Oludare Owolabi, a professional engineer in Maryland, joined the Morgan State University faculty in 2010. He is the director of the Sustainable Infrastructure Development, Smart Innovation and Resilient Engineering Research Lab, Department of Civil Engineering at Morgan State University.

Leveraging Predictive Analytics on ABET Assessment Data to Forecast Attainment of Complex Civil Engineering Student Outcomes in an HBCU

Abstract

Attainment of complex student outcomes, particularly teamwork and experimentation and data analysis, is essential for Accreditation Board for Engineering and Technology (ABET) accreditation and for preparing civil engineering graduates for professional practice. However, many ABET assessment systems rely on lagging indicators derived from end-of-term evaluations, limiting the ability of programs to implement timely instructional and advising interventions under ABET Criterion 4 Continuous Improvement. This study investigates the feasibility of using outcome-level predictive analytics to forecast end-of-term attainment of complex civil engineering student outcomes using early, course-embedded assessment data.

Using a multi-year dataset spanning Fall 2021 through Spring 2025, comprising over 31,000 rubric-level ABET outcome assessment records aggregated to the student-term level, supervised machine learning models were developed to predict attainment of teamwork (E5) and experimentation and data analysis (E6) outcomes. Predictor features were derived exclusively from outcome-level submissions available during the first 30 percent of each academic term. Logistic Regression and Random Forest classifiers were evaluated using five-fold stratified cross-validation, with performance assessed through accuracy, Area Under the Receiver Operating Characteristic Curve (AUC), and class-specific F1-scores.

Results demonstrate strong and consistent predictive performance across outcomes. For experimentation and data analysis, models achieved AUC values above 0.92, indicating robust discrimination under class imbalance. For teamwork outcomes, AUC values of 1.000 were observed under the present data conditions. Across both outcomes, Random Forest models provided modest gains in overall accuracy, while Logistic Regression offered competitive discrimination with greater interpretability. Exploratory fairness evaluations using demographic parity and equal opportunity criteria across available cohort partitions yielded minimal observed disparity in model behavior, supporting responsible and equity-aware application of predictive analytics within the HBCU context.

The findings indicate that early, outcome-level assessment data contain meaningful predictive signal and can strengthen ABET Criterion 4 continuous improvement processes. Because the methodology relies exclusively on data elements already collected by most ABET-accredited programs, the proposed predictive analytics framework is scalable, transparent, and adaptable across civil engineering curricula, enabling earlier intervention and more equitable student support.

Keywords: ABET Accreditation, Continuous Improvement, Predictive Analytics, Machine Learning, Student Outcomes Assessment, Civil Engineering Education, Learning Analytics.

1. Introduction

1.1 Background of Study

Accredited civil engineering programs are required to demonstrate that graduates attain defined student outcomes and that assessment results are systematically used to inform continuous program improvement, as required by the Accreditation Board for Engineering and Technology (ABET). These requirements are formalized through ABET Criterion 3, which specifies student outcomes, and Criterion 4, which emphasizes the use of assessment data to guide evidence-based decision-making and curricular enhancement [1]. Together, these criteria place increasing emphasis not only on outcome attainment, but also on the timeliness, rigor, and effectiveness of assessment processes used by engineering programs.

Among the ABET student outcomes, teamwork and experimentation and data analysis represent particularly complex competencies. These outcomes require students to demonstrate skills that extend beyond technical knowledge, including collaboration, communication, experimental design, data interpretation, and professional judgment. In civil engineering curricula, such competencies are typically assessed through laboratory-intensive courses, team-based projects, and design experiences that span multiple courses and academic terms. Consequently, attainment of these outcomes is influenced by both early and sustained student engagement across the curriculum.

Despite their importance, many ABET-compliant assessment systems rely predominantly on end-of-term or end-of-program measures, such as final project evaluations, capstone assessments, and summative rubric scores. While these approaches satisfy accreditation documentation requirements, they function primarily as lagging indicators and provide limited opportunity for early identification of students who may struggle to attain specific outcomes. As a result, instructional and advising interventions are often reactive rather than proactive, constraining the effectiveness of ABET Criterion 4 continuous improvement processes [2], [3].

Recent advances in learning analytics and predictive modeling offer a pathway to enhance ABET assessment practices by leveraging existing course-embedded data to generate earlier insight into student performance. Prior research in engineering education and educational data mining has demonstrated the feasibility of using machine learning techniques to predict student success, retention, and course-level outcomes based on assessment and activity data [4], [5]. However, much of this work has focused on aggregate academic success measures rather than on forecasting attainment of specific ABET student outcomes, particularly complex outcomes related to teamwork and experimentation and data analysis.

An outcome-level predictive approach is especially valuable for civil engineering programs, where assessment data are already collected at scale through learning management systems and outcome-aligned rubrics. When analyzed appropriately, these data can serve not only as documentation for accreditation, but also as actionable inputs to support early intervention, advising, and instructional refinement. Importantly, this approach aligns directly with the intent of ABET Criterion 4 by strengthening the feedback loop between assessment, analysis, and improvement.

At the same time, the application of predictive analytics in engineering education raises important ethical and equity considerations. Predictive models that are not carefully evaluated may inadvertently reinforce existing disparities or be misinterpreted as deterministic judgments rather than decision-support tools. These concerns are particularly salient in Historically Black Colleges and Universities and other minority-serving institutions, where responsible use of student data and transparency in analytics are essential.

Incorporating fairness-aware evaluation and explicitly framing predictive outputs as supportive rather than punitive is therefore critical for ethical deployment [6], [7].

Motivated by these considerations, this study presents a data-driven approach for forecasting attainment of complex ABET student outcomes using early, course-embedded assessment data. Rather than introducing new assessment instruments, the approach leverages existing ABET-aligned outcome records to identify early performance patterns associated with later attainment. Using a multi-year dataset from a civil engineering program, predictive models are developed and empirically evaluated for outcomes related to teamwork and experimentation and data analysis. The study demonstrates how outcome-level predictive analytics can strengthen ABET continuous improvement practices while remaining scalable, transparent, and adaptable across civil engineering programs.

1.2 Objectives of the Study

The primary objective of this study is to evaluate the effectiveness of predictive analytics models in forecasting attainment of complex ABET student outcomes using early, course-embedded assessment data within a civil engineering program. Rather than introducing new assessment instruments, the study focuses on extracting actionable insight from existing ABET-aligned outcome records to support earlier and more targeted continuous improvement actions.

Specifically, this study seeks to:

1. Develop and empirically evaluate predictive models that use early outcome-level assessment data to forecast end-of-term attainment of teamwork and experimentation and data analysis outcomes in civil engineering education.
2. Assess the feasibility of using early course-embedded outcome indicators as leading signals for ABET Criterion 4 continuous improvement, including their potential to inform advising, instructional refinement, and program-level decision-making.
3. Examine model performance across outcome types and modeling approaches, with attention to accuracy, discrimination capability, and class-sensitive performance metrics relevant to early identification of students at risk of non-attainment.
4. Evaluate ethical and equity considerations associated with outcome-level predictive analytics, including exploratory assessment of model behavior across demographic groups and discussion of responsible interpretation and use within an HBCU context.

1.3 Research Questions

Guided by these objectives, the study addresses the following research questions:

RQ1: To what extent can predictive analytics models trained on early, course-embedded ABET assessment data accurately forecast end-of-term attainment of complex civil engineering student outcomes related to teamwork and experimentation and data analysis?

RQ2: How does predictive performance vary between different modeling approaches, and what tradeoffs emerge between overall discrimination performance and identification of students at risk of non-attainment?

RQ3: How effective are early outcome-level performance indicators as leading signals for strengthening ABET Criterion 4 continuous improvement processes in civil engineering programs?

RQ4: What ethical and equity considerations arise when applying outcome-level predictive analytics in an HBCU context, and how can model results be interpreted and used responsibly to support student success?

2. Methodology

This section describes the data source, outcome definitions, feature construction process, predictive modeling approach, validation strategy, and fairness considerations employed in this study. The methodology is designed to be transparent and replicable by civil engineering programs seeking to leverage existing ABET assessment data to strengthen continuous improvement processes.

2.1 Data Source and Study Context

The analysis utilizes a multi-year dataset derived from a civil engineering program's ABET-aligned outcomes assessment system at a Historically Black College and University. The dataset spans Fall 2021 through Spring 2025 and consists of course-embedded outcome assessment records exported from the learning management system. These records include outcome scores, points possible, submission timestamps, course identifiers, and anonymized student identifiers.

The study focuses on two complex student outcomes aligned with ABET Criterion 3: Teamwork and Experimentation and Data Analysis. These outcomes are commonly assessed through laboratory activities, team-based projects, and design-oriented coursework within the civil engineering curriculum. All student identifiers were removed prior to analysis to ensure confidentiality and compliance with ethical research practices.

Because the dataset does not explicitly label assessment timing, early-term evidence was operationally defined using submission timestamps. For each academic term, assessments submitted within the first four weeks of the term were classified as early-term and used for feature construction.

2.2 Outcome Aggregation and Attainment Definition

Outcome assessment records were originally captured at the assignment or rubric-item level. To support predictive modeling and align with ABET reporting practices, records were aggregated to the student-term–outcome level. When outcome scores were reported on different point scales, values were normalized to a 4-point scale using the recorded points possible.

Outcome attainment was operationalized using a threshold consistent with common ABET assessment practice. A student-term outcome was classified as attained if the mean normalized score was greater than or equal to 3.0, and not attained otherwise. This binary classification enabled consistent modeling across courses and academic terms.

2.3 Early Outcome-Level Feature Construction

To evaluate early prediction capability, predictor variables were constructed using only outcome assessment data available during the early portion of each academic term. For each student-term–outcome record, the first 30 percent of chronologically ordered outcome submissions were identified using submission timestamps.

From these early submissions, a set of descriptive features was computed, including early mean outcome score, score variability, minimum and maximum early scores, number of early submissions, number of distinct courses contributing early outcome data, and temporal coverage of early submissions. These features capture early performance patterns observable to instructors and assessment coordinators prior to end-of-term outcome evaluation.

2.4 Predictive Modeling Approach

Two supervised classification models were evaluated to forecast outcome attainment:

1. **Logistic Regression**, selected for its interpretability and suitability for binary outcome prediction in educational assessment contexts.
2. **Random Forest**, selected for its ability to model nonlinear relationships and interactions among predictors through ensemble learning.

The use of both models enables comparison between an interpretable baseline approach and a higher-capacity ensemble method, supporting balanced evaluation of predictive performance and practical usability for academic stakeholders.

2.5 Model Training and Validation

Model performance was evaluated using five-fold stratified cross-validation to preserve outcome class distributions across folds and to maximize efficient use of available data. To prevent information leakage across repeated student outcome records, cross-validation folds were constructed using term-level separation rather than random row level partitioning. For each fold, models were trained on four partitions and evaluated on the remaining partition, with performance metrics averaged across folds. Because students may appear across multiple academic terms, term-level fold separation does not fully prevent the same student from contributing observations to both training and test partitions within a given fold. To assess the sensitivity of results to this potential source of leakage, a supplementary student-grouped cross-validation was conducted in which all records belonging to the same student were assigned to the same fold, ensuring complete separation of students across training and test sets. Results from both validation strategies are reported and compared.

Performance was assessed using accuracy, Area Under the Receiver Operating Characteristic Curve (AUC), and F1-score. AUC was included to provide threshold-independent evaluation under class imbalance, while F1-score was examined for both attained and non-attained classes to assess the models' ability to identify students at risk of non-attainment [8].

2.6 Fairness and Ethical Considerations

To support responsible application of predictive analytics, model performance was examined across available demographic attributes using commonly adopted fairness criteria, including demographic parity and equal opportunity. These analyses were conducted to identify potential disparities in predictive behavior rather than to support automated decision-making.

Because subgroup sample sizes vary and are limited for certain demographic categories, fairness results are interpreted cautiously and presented as exploratory indicators. Model outputs are explicitly framed as decision-support tools intended to inform advising, instructional refinement, and program-level assessment discussions, rather than as deterministic evaluations of individual students.

2.7 Replicability

The modeling pipeline relies exclusively on data elements commonly available in ABET assessment systems and learning management platforms. As such, the methodology is readily adaptable by other civil engineering programs seeking to enhance ABET Criterion 4 continuous improvement through data-driven analysis without introducing new assessment instruments. All analyses were conducted using standard implementations of Logistic Regression and Random Forest classifiers with default parameter settings, and results are reported as averages across five-fold stratified cross-validation to support reproducibility across institutional contexts.

2.8 Predictive Analytics Framework for ABET Continuous Improvement

This study adopts an outcome-level predictive analytics framework designed to enhance ABET Criterion 4 continuous improvement processes using existing course-embedded assessment data. The framework integrates early outcome evidence with supervised learning models to support timely, data-informed decision-making while maintaining transparency, scalability, and ethical safeguards.

The framework consists of five interrelated stages. First, ABET-aligned data capture leverages outcome-level assessment records collected through learning management systems and rubric-based evaluations. These data are mapped directly to ABET Criterion 3 student outcomes and aggregated across courses and academic terms.

Second, early indicator extraction identifies performance signals available during the early portion of the academic term. In this study, early indicators are derived from the first 30 percent of chronologically ordered outcome submissions and summarize early achievement patterns through descriptive statistics such as mean performance, variability, and submission coverage.

Third, predictive modeling applies supervised classification techniques to forecast end-of-term outcome attainment. Logistic Regression and Random Forest models are employed to balance interpretability and predictive capacity, enabling comparison of modeling tradeoffs relevant to academic stakeholders.

Fourth, performance and fairness evaluation assesses model effectiveness using accuracy, Area Under the Receiver Operating Characteristic Curve, and class-specific F1-scores, alongside exploratory fairness checks across demographic groups. This stage ensures that predictive outputs are evaluated not only for technical performance but also for responsible and equitable behavior.

Finally, decision support for continuous improvement translates predictive insights into actionable inputs for advising, instructional refinement, and program-level assessment. Rather than serving as deterministic judgments, predictive results are intended to inform supportive interventions and strengthen the feedback loop central to ABET Criterion 4.

By relying exclusively on data elements commonly available in ABET assessment systems, the proposed framework is readily adaptable across civil engineering programs and provides a structured pathway for extending traditional assessment practices toward earlier and more effective continuous improvement.

Figure 1 summarizes the outcome-level predictive analytics framework adopted in this study, illustrating the flow from ABET-aligned, course-embedded assessment data to decision support for ABET Criterion 4 continuous improvement. The framework highlights how early outcome-level indicators are extracted from existing assessment records, integrated into supervised predictive models, evaluated for performance and fairness, and translated into actionable inputs for advising, instructional intervention, and curriculum refinement.

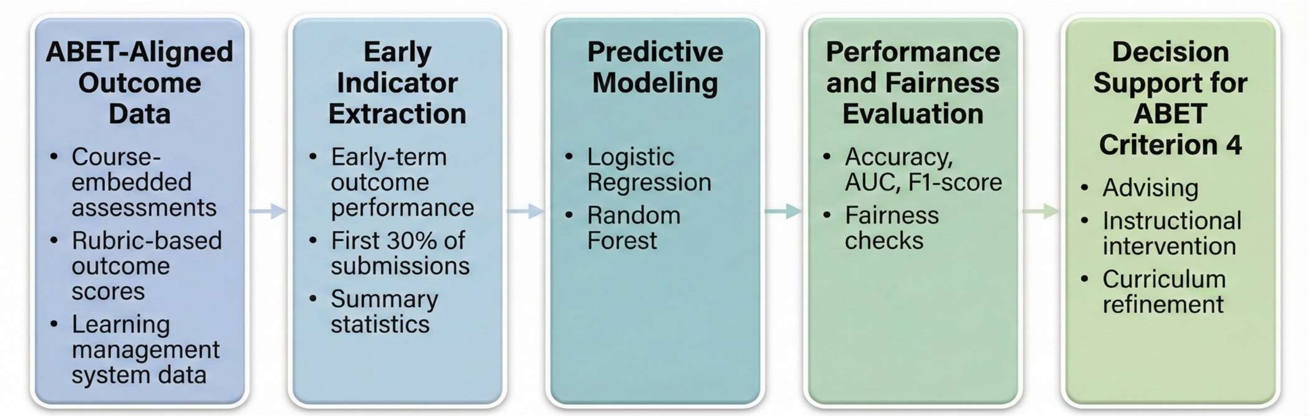

Figure 1. Outcome-Level Predictive Analytics Framework for ABET Continuous Improvement.

3. Results

3.1 Dataset Overview and Outcome Attainment

The final dataset used for analysis spans Fall 2021 through Spring 2025 and includes 31,311 outcome-level assessment records derived from course-embedded ABET assessments. After aggregation to the student-term level and filtering for valid submission dates, two outcomes were retained for analysis: E5 (Teamwork) and E6 (Experimentation and Data Analysis). Using a normalized 4-point rubric scale and an attainment threshold of 3.0, the resulting samples consisted of:

- E6 (Experimentation and Data Analysis): n = 281 student-term observations, with an attainment rate of 92.2%.
- E5 (Teamwork): n = 96 student-term observations, with an attainment rate of 89.6%.

Both outcomes exhibit class imbalance favoring attainment, which is a common characteristic of ABET outcome datasets. Accordingly, threshold-independent and class-sensitive evaluation metrics were used to assess predictive performance.

The substantial difference between the total record count and the final analytical sample warrants clarification. The 31,311 raw records represent individual rubric-item-level observations exported from the learning management system, where each row corresponds to a single rubric criterion scored for one student on one assignment. Multiple rubric items map to the same ABET outcome within a given assignment, and a student may accumulate dozens of such records across a term. The aggregation pipeline proceeded through the following steps: (1) all records were filtered to retain only those associated with E5 or E6 outcome groups, yielding 875 and 11,410 records respectively; (2) records lacking valid submission timestamps were removed, retaining 79.4% of E5 records and 95.6% of E6 records; (3) records were aggregated to the student-term level by computing mean normalized scores per student per term; and (4) student-term observations were retained only where sufficient early-term submissions existed to construct predictor features. This pipeline produced 96 student-term observations for E5 and 281 for E6, representing the number of unique student-term enrollments with complete early assessment data, not the number of individual rubric scores. The primary source of attrition for E5 was missing submission timestamps in 20.6% of records, likely reflecting manually entered scores or bulk imports that bypassed the timestamp logging system. This limitation is acknowledged and represents an avenue for data quality improvement in future implementations.

3.2 Predictive Performance for Experimentation and Data Analysis (E6)

Table 1 summarizes the predictive performance of Logistic Regression and Random Forest models for forecasting attainment of experimentation and data analysis (E6) and teamwork (E5) outcomes using early, course-embedded ABET assessment data. Performance is reported in terms of accuracy, Area Under the Receiver Operating Characteristic Curve (AUC), and F1-scores for both the attained and at-risk classes, averaged across five-fold stratified cross-validation. These metrics provide complementary perspectives on overall classification accuracy, discrimination capability under class imbalance, and identification of students who may be at risk of non-attainment.

For E6, Logistic Regression achieved an average accuracy of 0.919 and an AUC of 0.968, indicating strong discrimination between attainment and non-attainment cases. The Random Forest model exhibited comparable accuracy (0.910) and AUC (0.925), reflecting consistent predictive capacity across both modeling approaches. F1-scores for the at-risk class ranged from 0.440 to 0.441, indicating that identification of non-attainment cases remains more challenging under class imbalance. These results suggest that early outcome-level performance contains substantial predictive signal for experimentation-related outcomes and that both modeling approaches are suitable for early forecasting, with modest trade-offs between interpretability and marginal performance gains.

Table 1. Summary of Predictive Model Performance

Outcome	**Model**	**Accuracy**	**AUC**	**F1-score (Attained)**	**F1-score (At-Risk)**
E6 – Experimentation and Data Analysis	Logistic Regression	0.919	0.968	0.956	0.441
E6 – Experimentation and Data Analysis	Random Forest	0.910	0.925	0.950	0.440
E5 – Teamwork	Logistic Regression	0.877	1.000	0.931	0.400
E5 – Teamwork	Random Forest	1.000	1.000	1.000	1.000

Across both outcomes, Logistic Regression and Random Forest demonstrated comparable discrimination performance, as reflected in similar AUC values reported in Table 1. While Random Forest provided marginal gains in overall accuracy and attained-class F1-scores, Logistic Regression offered competitive performance with greater interpretability. This pattern suggests that both approaches are viable for early outcome forecasting, with model selection depending primarily on institutional priorities related to transparency and deployment.

3.3 Predictive Performance for Teamwork (E5)

Predictive performance for the teamwork outcome is also reported in Table 1. Both models achieved very high accuracy and discrimination metrics for this outcome. This strong performance reflects, in part, the structured and rubric-driven nature of group-based teamwork assessments, which produce stable early signals of outcome alignment rather than direct measures of latent teamwork skill.

For E5, Logistic Regression achieved an average accuracy of 0.877 and an AUC of 1.000, while Random Forest achieved an accuracy of 1.000 with an identical AUC of 1.000. For Logistic Regression, F1-score was 0.931 for the attained class and 0.400 for the at-risk class. For Random Forest, F1-scores were 1.000 for both classes.

While these results indicate exceptionally strong predictive performance, they should be interpreted with caution. The effective sample size for teamwork is smaller than for experimentation, and early and final assessments for teamwork outcomes are often closely aligned within the same course activities. As such, high performance may reflect strong coupling between early and summative assessments rather than generalizable predictive separation across diverse instructional contexts.

3.4 Feature Importance Analysis

To provide actionable insight into which early-term behaviors are most predictive of outcome attainment, feature importance was assessed for both modeling approaches. For Logistic Regression, standardized coefficients indicate the direction and relative magnitude of each feature's contribution to the predicted log-odds of attainment. For Random Forest, Gini impurity-based importance reflects each feature's average contribution to reducing classification uncertainty across all decision trees. Both analyses were conducted on the full analytical sample to characterize overall predictor structure rather than fold-specific performance.

Table 2 presents the feature importance results for both outcomes and both models. Across E5 and E6, early mean outcome score emerged as the dominant predictor, with the highest Logistic Regression coefficient and Random Forest Gini importance in both cases. For E5, early minimum score ranked second in importance under both approaches, suggesting that the lowest early performance observation carries particular signal for final teamwork attainment. For E6, score variability and number of early submissions contributed meaningfully alongside mean score, reflecting the more distributed assessment structure of experimentation-related coursework across multiple laboratory courses.

These findings offer concrete guidance for faculty and assessment coordinators. Early mean outcome score, defined as the average rubric score across the first 30 percent of term submissions, is the single most informative early indicator for both outcomes. This suggests that programs seeking to implement early identification systems can prioritize consistent rubric-based scoring of initial assignments and laboratory reports as the foundation of a practical early warning signal. The relatively low importance of the number of distinct courses and the number of early submissions indicates that score quality is more informative than submission volume or coverage breadth during the early term.

Table 2. Feature Importance for Early Outcome-Level Prediction

Feature	E5 LR Coeff	E5 RF Gini	E6 LR Coeff	E6 RF Gini
Early Mean Score	+1.471	0.477	+2.077	0.469
Early Std Dev	-0.680	0.023	+0.089	0.152
Early Min Score	+2.189	0.362	+1.447	0.097
Early Max Score	+1.062	0.129	+0.591	0.137
No. Early Submissions	-0.062	0.008	+0.139	0.114
No. Distinct Courses	+0.011	0.000	+0.719	0.030

Note: LR coefficients reflect models trained on the full analytical sample. RF Gini importance values sum to 1.0 within each outcome. Models were trained using scikit-learn (Python) with default hyperparameter settings.

To further examine the relationship between early-term features, full-term mean score, and attainment label, Pearson correlation matrices were computed for both E5 and E6. As shown in Figure 2, stronger correlations between early and final mean scores are observed for E5 than for E6, providing additional evidence of structural coupling in teamwork assessments.

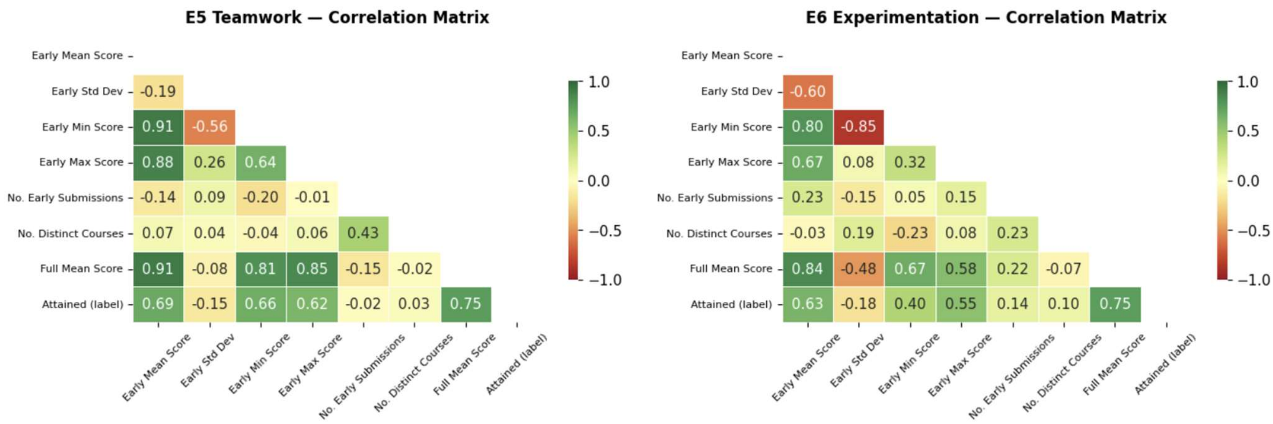

Figure 2. Pearson Correlation Matrix Between Early Features, Full-Term Mean Score, and Attainment Label for E5 Teamwork (left) and E6 Experimentation and Data Analysis (right).

3.5 Comparative Model Performance

As shown in Table 1, both models achieve AUC values near 0.90 for experimentation outcomes, indicating strong and stable discrimination. For teamwork outcomes, AUC values of 1.000 were observed for both models, reflecting perfect separation under the present data conditions. As discussed in Section 4.1, this result is best interpreted as a reflection of structural coupling between early and summative assessments rather than generalizable predictive capability.

Across both outcomes, Random Forest models exhibit marginally higher accuracy and attained-class F1-scores, while Logistic Regression provides competitive discrimination performance with greater transparency in feature effects. This balance suggests that either approach could be viable depending on institutional priorities related to interpretability, transparency, and deployment.

3.6 Summary of Key Findings

Overall, the results demonstrate that:

1. Early, course-embedded outcome assessment data can reliably forecast end-of-term attainment for complex ABET student outcomes.
2. Experimentation and data analysis outcomes exhibit strong but realistic predictive performance, supporting their suitability for early intervention use.
3. Teamwork outcomes show very high predictive performance, though results are sensitive to sample size and assessment structure.
4. Multiple modeling approaches yield consistent conclusions, reinforcing the robustness of the outcome-level predictive signal.

These findings provide empirical support for the use of outcome-level predictive analytics as a complement to traditional ABET assessment practices and motivate further exploration of responsible deployment within continuous improvement workflows.

4. Discussion and Implications

4.1 Interpretation of Key Findings

The results demonstrate that early, course-embedded ABET assessment data contains substantial predictive signal for forecasting end-of-term attainment of complex civil engineering student outcomes. For experimentation and data analysis (E6), both Logistic Regression and Random Forest models achieved strong discrimination performance, with AUC values near 0.90. This level of performance indicates that early outcome-level indicators can reliably distinguish between students likely to attain the outcome and those at risk of non-attainment, even under class imbalance. These findings support the feasibility of using existing assessment records as leading indicators rather than relying solely on summative measures. In the context of this civil engineering program, early laboratory reports from courses such as Fluid Mechanics and Geotechnical Engineering served as key predictors of final E6 attainment, demonstrating that foundational laboratory experiences provide meaningful early signals of experimentation-related competency development.

For teamwork (E5), both models achieved exceptionally high discrimination, with AUC values of 1.000 under term-level cross-validation. Before interpreting this as evidence of strong predictive capability, however, we examined whether the high performance might reflect structural coupling between early and final assessments rather than genuine prediction of an independent future outcome. Correlation analysis indicated stronger association between early and final mean scores for E5 than for E6, suggesting that early and summative teamwork assessments may largely capture the same underlying performance within the same instructional sequence.

It is important to clarify that this does not represent data leakage in the classical sense. Early submissions are temporally prior to final assessments, and no future information entered model training. What it does reveal is that the near-perfect AUC for E5 is better understood as a reflection of assessment consistency than as evidence of the model's ability to predict a truly independent future outcome. In contrast, the moderate coupling observed for E6 ($r = 0.839$) supports a more defensible interpretation: early laboratory performance provides meaningful, non-redundant information about end-of-term experimentation outcomes. We therefore encourage programs to consider this framework for teamwork outcomes to first examine whether their own assessment structures exhibit similar coupling, as this will significantly affect how predictive results should be interpreted and used. This strong coupling, rather than representing a limitation solely, reinforces the critical importance of intentionally designing early team activities, as they are directly indicative of final performance. For teamwork, the value of this predictive framework may lie less in forecasting failure and more in diagnosing early team dynamics, enabling just-in-time interventions such as conflict resolution workshops or team restructuring to support student success.

Across outcomes, Random Forest models provided modest gains in overall accuracy and attained-class F1-scores, while Logistic Regression offered competitive discrimination performance with greater interpretability. This tradeoff suggests that institutions may select modeling approaches based on priorities related to transparency, ease of communication with faculty, and deployment constraints, recognizing that both approaches demonstrated consistent and practically useful predictive behavior across the analytical sample.

4.2 Implications for ABET Criterion 4 Continuous Improvement

These findings have direct implications for strengthening ABET Criterion 4 continuous improvement processes. Traditional ABET assessment systems often function as retrospective reporting mechanisms, identifying gaps only after outcomes have been assessed at the end of a term or program cycle. By contrast, outcome-level predictive analytics enable earlier identification of potential non-attainment, allowing faculty and program administrators to intervene while instructional adjustments and student support remain feasible.

In practice, predictive outputs derived from early assessment data can inform targeted advising, supplemental instruction, laboratory support, or team-based skill development. At the program level, aggregate predictive trends can highlight courses or curricular sequences where early performance systematically predicts later challenges, thereby guiding evidence-based curriculum refinement. Importantly, this approach does not replace existing ABET assessment practices but enhances them by extending the temporal reach and actionable value of assessment data.

4.3 Pedagogical and Advising Implications

From a pedagogical perspective, the strong predictive role of early outcome performance underscores the importance of intentionally designed early assignments, laboratories, and team activities. Importantly, the predictive models developed in this study are intended to function as decision-support tools that inform instructional and advising judgment, rather than as automated or deterministic evaluators of individual student performance. When early assessments are aligned with the cognitive and professional demands of targeted outcomes, they not only support learning but also generate actionable data for instructional decision-making.

For advising and student support within an HBCU context, operationalizing predictive outputs requires intentional integration into existing student success structures. Practically, this may involve academic advisors or faculty receiving early-term flags for students whose outcome-level performance falls below the predictive threshold, prompting proactive outreach, such as scheduled check-ins, referrals to supplemental instruction, or targeted laboratory tutoring, before deficiencies compound. For experimentation and data analysis outcomes, where early mean score and minimum score were identified as the strongest predictive features, instructors can use the first laboratory reports or data analysis assignments as structured early diagnostic tools, interpreting low scores as actionable signals rather than simply gradebook entries.

Example Intervention Workflow Using Early Outcome Prediction:

- **Week 4:** The predictive model flags Student X for E6 (experimentation and data analysis) at-risk status based on early laboratory report scores.
- **Week 5:** The academic advisor receives the automated flag and initiates a structured check-in meeting to discuss laboratory coursework challenges.
- **Week 6:** Student X is connected to supplemental instruction sessions focused on laboratory report writing, data interpretation, and experimental documentation.
- **End of Term:** Student X demonstrates improved performance on subsequent laboratory assignments and successfully attains the E6 outcome.

This illustrative sequence demonstrates how outcome-level predictive analytics can function as a timely intervention catalyst within civil engineering programs, transforming assessment data into proactive academic support rather than retrospective evaluation.

In teamwork contexts, where early and summative assessments are closely coupled, interventions may focus less on prediction and more on intentional early team formation support, conflict resolution resources, and structured peer evaluation practices. Critically, at an HBCU where students may face compounding structural barriers, including financial pressures, first-generation status, and limited access to professional networks, predictive flags must be paired with genuine support infrastructure rather than surveillance. The observed F1-scores for the at-risk class indicate that false positives and false negatives remain possible; consequently, predictive signals should always be used to initiate supportive conversations and expand resource access, never to restrict opportunity or assign academic penalties.

4.4 Ethical and Equity Considerations

The application of predictive analytics in engineering education raises important ethical considerations, particularly within HBCU and minority-serving institution contexts. A key data limitation must first be acknowledged: the Canvas LMS export used in this study does not include student demographic attributes such as race, gender, or ethnicity. Direct demographic-group fairness analysis was therefore not computable from this dataset alone and would require IRB-approved linkage to institutional enrollment records. Future work should prioritize IRB-approved research that links this predictive framework to actual demographic data, enabling comprehensive fairness evaluation and validation of these exploratory findings across race, gender, and other protected attributes.

As a structural proxy, model performance was evaluated across academic term season (Fall vs. Spring), capturing potential cohort-level variation in student population and instructional context. For each outcome and subgroup, performance metrics including AUC, F1-score (Attained), and F1-score (At-Risk) were computed. Differences in predicted positive rates and true positive rates were summarized as Demographic Parity and Equal Opportunity deltas, respectively. Table 3 presents these subgroup-level performance metrics and associated disparity measures. Subgroup performance metrics were computed by applying the trained models to the full analytical sample and then evaluating prediction performance within Fall and Spring partitions for diagnostic comparison. Separate models were not retrained within each subgroup due to limited subgroup sample sizes. Accordingly, subgroup AUC values reflect within-group discrimination performance of the overall model rather than independently cross-validated subgroup models.

Table 3. Exploratory Fairness Evaluation by Term Season Proxy Group

Computed Fairness Values:

Outcome	Group	n	Actual	Pred	AUC	F1-Att	F1-Risk
E6	Fall	144	88.2%	93.1%	0.987	0.973	0.741
E6	Spring	137	96.4%	94.9%	0.999	0.992	0.833
	Δ (Parity / Opportunity)	–	–	0.018*	–	0.019*	–
E5	Fall	55	83.6%	83.6%	1.000	1.000	1.000
E5	Spring	41	100.0%	100.0%	N/A†	1.000	0.000
	Δ (Parity / Opportunity)	–	–	0.164*	–	0.000*	–

† AUC and F1-Risk not computable for Spring E5, all observations belong to the attained class.

For experimentation and data analysis (E6), evaluation across Fall (n = 144) and Spring (n = 137) cohorts yielded a Demographic Parity Delta of 0.018 and an Equal Opportunity Delta of 0.019, indicating minimal disparity in model behavior across cohorts. AUC values were 0.987 for Fall and 0.999 for Spring, reflecting consistently strong discrimination across both groups. For teamwork (E5), the Spring cohort achieved 100% actual attainment, making a meaningful parity comparison unavailable for that outcome.

While these proxy-based results do not constitute a full demographic fairness audit, they provide an initial quantitative foundation consistent with responsible AI practices. Transparent reporting of model limitations, ongoing monitoring of performance across demographic groups as data become available, and clear communication of model intent are essential to responsible deployment. Models developed in this study are intended to support equitable student success by enabling earlier and more targeted support, not to label or track students in ways that constrain opportunity. Embedding predictive analytics within a broader culture of care, advising, and continuous improvement is critical to realizing their potential benefits.

Beyond subgroup performance stability, we also examined the possibility that students appearing across multiple terms could introduce information leakage into term-level cross-validation. A student-grouped sensitivity analysis was therefore conducted. Among the analytical samples, 12 of 83 E5 students and 97 of 143 E6 students contributed observations in more than one term. Under student-grouped cross-validation, all records from a given student were assigned to the same fold, ensuring that no student appeared in both training and test sets.

For E6, performance remained highly stable, with AUC values of 0.964 and 0.909 for Logistic Regression and Random Forest, respectively, compared to 0.968 and 0.925 under term-level separation, representing differences of less than 0.02. For E5, results remained at 1.000 under both strategies, consistent with the structural coupling between early and final assessments. As illustrated in Figure 2, the strong correlation between early and full-term mean scores for E5 provides additional evidence of this coupling. These findings indicate that E6 results are robust to student-level separation and are not materially inflated by cross-term student overlap.

4.5 Generalizability and Practical Adoption

A key strength of the proposed approach is its reliance on data elements already collected by most ABET-accredited civil engineering programs. Because the methodology leverages existing outcome-level assessment records, it is readily adaptable across institutions without requiring new assessment instruments or major infrastructure investments. Programs can tailor feature construction, thresholds, and modeling choices to align with local assessment practices while preserving the core analytic framework.

Taken together, the findings suggest that outcome-level predictive analytics represent a practical and scalable enhancement to ABET assessment systems. When implemented thoughtfully, such approaches can improve the timeliness, effectiveness, and equity of continuous improvement efforts in civil engineering education.

5. Conclusion and Future Work

This study demonstrates the feasibility and value of applying outcome-level predictive analytics to strengthen ABET assessment and continuous improvement processes in civil engineering education. By leveraging early, course-embedded ABET assessment data, the proposed approach enables reliable forecasting of end-of-term attainment for complex student outcomes related to experimentation and data analysis and teamwork. The results indicate that early performance indicators contain meaningful predictive signal and can support timely, evidence-based decision-making by faculty and program administrators.

Empirical evaluation across multiple academic terms indicates strong and consistent predictive performance for experimentation-related outcomes, with particularly high discrimination observed for teamwork outcomes, though interpretation of these results should account for potential structural coupling between early and summative assessments. While differences in performance across modeling approaches highlight tradeoffs between interpretability and marginal gains in accuracy, both Logistic Regression and Random Forest models demonstrate practical utility for early identification of students at risk of non-attainment. Importantly, these findings underscore that predictive analytics can complement, rather than replace, traditional ABET assessment practices by extending the temporal reach and actionable value of assessment data.

Beyond technical performance, this work emphasizes the responsible and equitable application of predictive analytics within ABET-accredited programs. Ethical considerations, including fairness evaluation, transparency, and non-punitive use of predictive outputs, are essential to ensuring that analytics serve as tools for student support and continuous improvement. These considerations are particularly salient in HBCU and minority-serving institution contexts, where equity, trust, and student-centered decision-making are foundational to effective educational innovation.

Because the methodology relies on data elements already collected by most engineering programs, the proposed framework is readily transferable across civil engineering curricula without requiring new assessment instruments or major infrastructure investments. Future work will focus on prospective validation using new student cohorts, integration of predictive outputs into advising and assessment dashboards, incorporation of IRB-approved demographic linkage for comprehensive fairness evaluation, and extension of the approach to additional ABET student outcomes. Collectively, this study contributes a scalable, data-driven pathway for enhancing ABET continuous improvement while promoting timely and equitable student success in civil engineering education.

REFERENCES

[1] ABET, *Criteria for Accrediting Engineering Programs, 2024–2025*, Baltimore, MD, USA: ABET, 2024. [Online]. Available: https://www.abet.org/accreditation/accreditation-criteria/

[2] G. Rogers, "Continuous improvement in engineering education: Closing the loop," *International Journal of Engineering Education*, vol. 18, no. 1, pp. 36–42, 2002.

[3] J. E. Froyd, P. C. Wankat, and K. A. Smith, "Five major shifts in 100 years of engineering education," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1344–1360, May 2012.

[4] R. S. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.

[5] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601–618, Nov. 2010.

[6] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning: Limitations and Opportunities*, Cambridge, MA, USA: MIT Press, 2023.
[Online]. Available: https://fairmlbook.org

[7] D. Selbst, A. Boyd, S. Friedler, S. Venkatasubramanian, and J. Vertesi, "Fairness and abstraction in sociotechnical systems," in *Proc. ACM Conf. Fairness, Accountability, and Transparency (FAT*)*, Atlanta, GA, USA, 2019, pp. 59–68.

[8] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, Jun. 2006.