DEVELOPMENT AND IMPLEMENTATION OF A PREDICTIVE MODEL TO ENHANCE ACADEMIC ADVISING FOR USMLE STEP 1 SUCCESS

Sneha Iyer MBA, MEd, Justine Cameron MPP, Stephen McAllister MA, MBA, Paul Weissburg PhD, Courtney Vengrin MS, PhD

Dartmouth GEISEL SCHOOL OF MEDICINE

INTRODUCTION

DEFINITIONS

- Comprehensive Basic Science Examination (CBSE)
- Comprehensive Basic Science Self-Assessment (CBSSA)
- Pre-clerkship Course Scores: Average of final scores of pre-clerkship courses

Context:
USMLE Step 1 is a pivotal milestone in medical education affecting student progression and residency opportunities.
Early identification of students at risk for failing Step 1 allows for timely academic support and tailored advising interventions.

Current Challenges:
Traditional advising often relies on subjective evaluation and historical data, which may delay identification of at-risk students.

There is a need for objective, data-driven tools to improve the accuracy and timeliness of risk prediction.

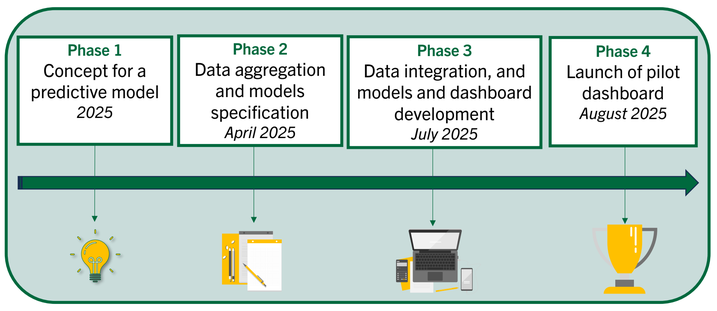
Rationale for Innovation:
Advances in educational data analytics provide opportunities to leverage existing assessment data (CBSE, CBSSA, pre-clerkship course scores) to predict Step 1 outcomes.

Predictive modeling can enhance advising by providing actionable insights earlier in the curriculum.

Purpose:

- Develop and implement a predictive model to estimate the probability of first-attempt Step 1 success using prior academic performance.
- Integrate the model into an interactive dashboard to support timely monitoring and early intervention by advising and student support teams.

Timeline:



QUALITY IMPROVEMENT AIM

- Strengthen evidence-based advising practices.
- Promote efficient allocation of support resources.
- Improve first-attempt Step 1 outcomes while establishing a scalable framework for ongoing refinement and future expansion to additional academic milestones.
- Improve timeliness, accuracy, and consistency of advising decisions compared to traditional approaches that rely heavily on subjective judgement and retrospective indicators
- Enhance organization of this data through use of interactive dashboard.

MODELING APPROACH

Data included CBSE, CBSSA, pre-clerkship course scores, and first-attempt Step 1 results.

Years: 2023 through 2025

A Python-based predictive algorithm was integrated into a Power BI dashboard to enhance end user experience, support interactive visualization, and generate detailed reports.

Seven logistic regression models were developed using single and combined predictors (CBSE, CBSSA, pre-clerkship course scores).

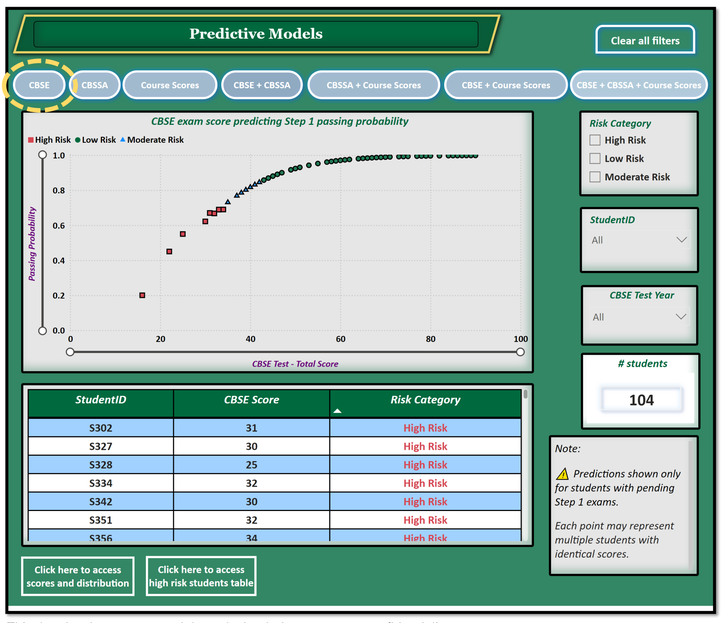
RESULTS

Risk Tier Categorization:
Students were classified into three risk levels based on their passing probabilities: High, Moderate, and Low. The leadership team convened to reach consensus on the categorization of risk tiers.

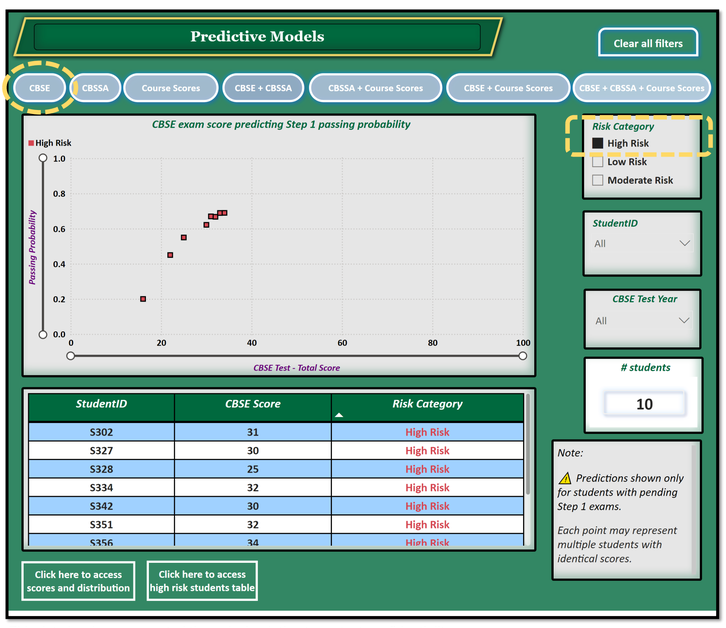
Early Detection:
Using predictive models, advisors can identify students who have a low probability of passing Step 1 on their first attempt.

Dashboard Functionality:
The dashboard enables users to interact with the visuals and select specific risk tiers. Users can also switch between different models to examine the preferred regression model.

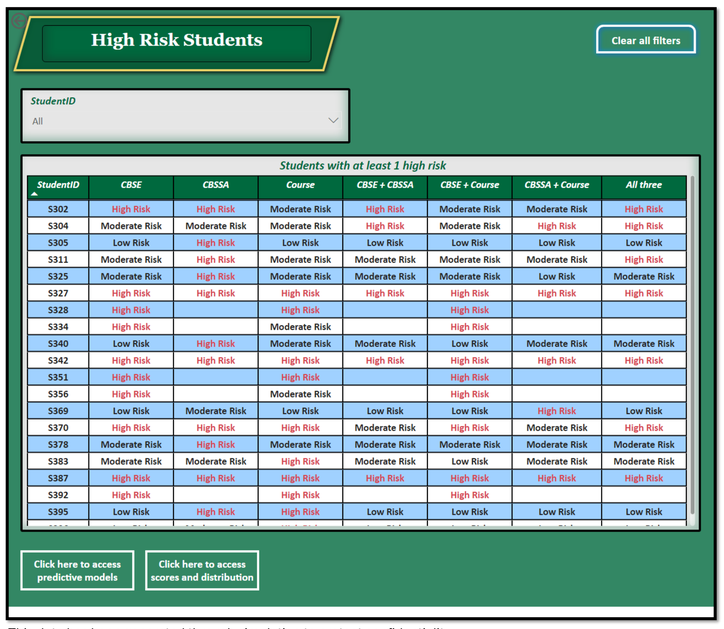
Model Evaluation:
Model performance was assessed through their predictive accuracy. Predictor's importance was determined by their impact on prediction outcomes rather than by statistical significance.



This data has been generated through simulation to protect confidentiality.



This data has been generated through simulation to protect confidentiality.



This data has been generated through simulation to protect confidentiality.

DISCUSSION

- Demonstrates the feasibility and utility of a predictive model using routinely collected academic data to estimate medical students' likelihood of first-attempt Step 1 success.
- Integrates academic data using logistic regression to enhance data-driven advising practices.
- An interactive Power BI dashboard enables timely monitoring student progress, early identification of at-risk learners, and tailored interventions.
- The model demonstrated strong performance in retrospective validation, supporting its use alongside traditional advising methods, which often rely on subjective judgement or incomplete information.
- This approach aligns with the growing focus on evidence-based educational practices and learning analytics in medical education.
- User feedback supports the feasibility and acceptability of integrating the dashboard into routine advising workflows, with potential to improve student support efficiency.
- By sharing the development process and implementation experience, this work contributes to the broader application of predicative analytics in medical education to support student success.
- Provides an adaptable, data-informed advising framework for institutions with similar data infrastructures.
- Future work will focus on prospective validation and exploring integration with other academic and wellness metrics to further refine predictive accuracy and holistic student support.

PROJECT LIMITATIONS

- Developed and validated using retrospective, single institution data, which may limit generalizability to other medical schools with different curricula or assessment structures.
- Relies on available academic metrics, which may omit other influential factors such as student well-being, study habits, or external stressors that affect Step 1 performance.
- Predictive performance depends on the quality and completeness of institutional data, which may vary across cohorts and over time.
- Implementation depends on access to specific data platforms and analytic infrastructure (e.g., OASIS, Power BI), which may not be available at all institutions.

FUTURE DIRECTIONS

- Validate models with incoming student cohorts to assess predictive accuracy and advising impact
- Integrate additional variables such as student wellness, study behaviors, and demographic factors to enhance model precision
- Expand the dashboard to include predictive analytics for other critical milestones (e.g., Step 2 CK, clinical performance) to support longitudinal advising
- Collaborate with other institutions to test model generalizability and adapt the approach to diverse curricular settings
- Evaluate model effectiveness in improving student outcomes through targeted interventions informed by predictive insights

CONCLUSIONS

- Predictive modelling using academic performance data provides a valuable, data-driven tool to enhance advising for the Step 1 success.
- An interactive dashboard delivers timely, objective insights, supporting early identification of at-risk students and facilitating targeted interventions.
- While current validation demonstrates strong predictive capability, ongoing refinement and prospective evaluation will strengthen its utility.
- This innovation highlights the potential of predictive analytics to transform student support and underscores the importance of leveraging institutional data to promote academic success in medical education.

REFERENCES

Keltner C, Haedinger L, Carney PA, Bonura EM. Preclinical assessment performance as a predictor of USMLE Step 1 scores or passing status. *Medical Science Educator*. 2021;31(4):1453-1462. doi:10.1007/s40670-021-01334-7
Saqr M, Fors U, Tedre M. How learning analytics can early predict under-achieving students in a blended medical education course. Medical Teacher. 2017;39(7):757-767. doi:10.1080/0142159x.2017.1309376
Dufault CL, Colson ER, Dallaghan GL, et al. Using dashboards to support continuous quality improvement in undergraduate and graduate medical education. *Journal of General Internal Medicine*. 2024;40(1):171-176. doi:10.1007/s11606-024-09011-2
Manger J, Benedik L, Matott M, O'Malley CB. Why the Delay? Factors Predicting Step 1 Delay. MedSciEduc. 2025;35(3):1215-1219. doi:10.1007/s40670-025-02349-0

Northeast Group on Educational Affairs (NEGEA) Meeting
Scranton, PA
March 2026

sneha.iyer@dartmouth.edu