Increasing the Education Leadership Graduation Rate

An approach to increase the graduation rate of a school district leadership training program.

Presented as the final project for Teachers College HUDK 4050 Educational Data Mining

By Christine Gao, Nicolás Dussaillant, Mina Choi, Robert Abrams

December 21, 2021

# Introduction

Gao, Dussaillant, Choi and Abrams LLC (GDCA), an educational data mining consulting firm, was contacted by Mariner College’s School District Leadership Program to help them find ways to reduce attrition of students in their program and increase their graduation rate. After an initial meeting with the program director to obtain an introduction to their issues, we reviewed the literature to discover programs with similar issues, and to identify potential contributing factors and promising solutions. This report provides short descriptions of the program and available data. An analysis plan is proposed using educational data mining (EDM) techniques. Expected results, implications, challenges and limitations are discussed.

If this analysis plan is of interest, the GDCA team is available to meet to discuss next steps, in-person or via Zoom, depending on Covid comfort levels and scheduling preferences.

# Literature Review: Overview

After identifying our problem, we conducted a literature review to inform our recommended analysis plan. One part of this literature review looked at graduate student attrition from their programs of study. The second part looked at attrition from educational leadership programs specifically. The literature review is presented here in a style convergent with the first stage literature review often used by David Pimentel (who was a member of Robert Abrams’ doctoral committee at Cornell University), where each cited paper has a citation followed by key points in the paper that relate to the researchers’ chosen problem. (Our new-fangled innovation here is linking directly from the literature review to the paper. Back in Dr. Pimentel’s day, we had to walk uphill, in both directions, in the snow to the library’s stacks to find the papers, and then the readers of the literature review also had to walk uphill, in both directions, in the snow to the library if they wanted to read the reviewed papers themselves. We left you one paper to find on your own.)

Based on these studies’ findings, we created a hypothetical version of the common problem and proposed a method for predicting the dropout rate. This could be applied to real educational leadership programs.

In the first, quantitative, phase, five external and internal to the program factors were found to be predictors to students’ persistence in the program: “program”, “online learning environment”, “student support services”, “faculty”, and “self-motivation”. In the qualitative follow up multiple case study analysis four major themes emerged: (1) quality of academic experiences; (2) online learning environment; (3) support and assistance; and (4) student self-motivation. The quantitative and qualitative findings from the two phases of the literature review are discussed with reference to prior research. Implications and recommendations for policy makers are provided.

## Literature Review on Graduate Student Attrition

It is well-known that educational leadership plays a critical role in the development of students’ academic careers throughout their early K-12 experience, so there is no question that it is important for these leaders to receive a high quality education as well. Unfortunately, studies are finding that these programs have high rates of attrition.

To better understand the reasons for this, the researchers first investigated determinants of graduate level enrollment and attrition.

[Understanding the Determinants of Graduate School Enrollment](https://academiccommons.columbia.edu/doi/10.7916/d8-jhtg-hr15) (Mayyasi, 2020)

[Dropping out of master’s degrees: objective predictors and subjective reasons](https://www.tandfonline.com/doi/abs/10.1080/07294360.2020.1799951) (Rotem, Yair & Shustak, 2020)

Findings: Masters students drop out at a rate of 12%. Used a 4-step hierarchical logistic regression model to identify predictors.

Dropout rates were **better predicted by academic performance variables than background variables**

5 factors linked to departure: work obligations, institutional difficulties, family and personal obligations, economic feasibility of degree, and harassment

[I want to quit education: A longitudinal study of stress and optimism as predictors of school dropout intention](https://www.sciencedirect.com/science/article/pii/S0140197114001122?casa_token=3R2-tU7MkX0AAAAA:B1X5pVCfpFpXnd6LY4AEMrIUzzUNHX4vGBfv3K7y_9t-8xBIgdVQSJY37Vsr2zbKJ34nL244Eg) (Eicher, Staerklé & Clémence, 2014)

Findings show that the **levels of stress** and **optimism** affect students’ dropout intentions

A sample of 4K students from in Switzerland

Took into account that **dropout intention may vary or develop over time** instead of looking it as a one-time event

[Predictors of academic efficacy and dropout intention in university students: Can engagement suppress burnout?](https://www.proquest.com/docview/2455816344?pq-origsite=summon) (Marôco, Assunção, Harju-Luukkainen, Lin, Sit, Cheung, et al., 2020)

Surveyed 4k+ university students in 9 different countries

Used Social support, Coping strategies, and school-related variables to measure student engagement and burnout. Then analyzed the effects of student engagement, student burnout, and their interaction to predict Academic Performance and Dropout Intention

An Empirical Study Of Student Relationships And Academic Achievement (Niebuhr, K. E., & Niebuhr, R. E., 1999) *Education*, *119*(4), 679.

## Literature Review on Educational Leadership Programs

[Doctoral Student Attrition and Retention](https://journals.sagepub.com/doi/pdf/10.2190/RAWM-HXTB-M72D-1FFH?casa_token=hYBnvGSnTiYAAAAA:v-xQatmgLiAzHWz_WCeuusG0M-hI5vNu0g8XBi5rvNzgX_bCsPzLYxjdXFtW2i5Dkkk0V4nh5ExVyA) (Pauley, Cunningham, Toth, 1999)

Surveyed students admitted to an educational administration doctorate program between ‘80-’93. Factors found to be positively related to the completion of the doctoral program: level of **financial support** available to the student; familial, peer, faculty, and chairperson **support**; and the student’s **motivation**.

[Innovative Principal Preparation Programs: What works and how we know](https://www.proquest.com/docview/1506940981?parentSessionId=qSlnOr6qcmZQhNF%2FrHOwRii6pQW9gs%2BV5IhBi5k3vk0%3D&pq-origsite=summon&accountid=10226) (Davis & Darling-Hammond, 2012)

Case study of 5 successful (“exemplary”) principal preparation programs with design elements aligned with seven key features of effective leadership preparation programs.

[Students’ Persistence in a Distributed Doctoral Program in Educational Leadership in Higher Education: A Mixed Methods Study](https://link.springer.com/article/10.1007/s11162-006-9025-4) (Ivankova & Stick, 2007)

Identifying factors from program that motivate students to persist in their graduate studies: student support services, faculty, and self-motivation. Follow-up study found major themes, incl.: quality of academic experiences, support and assistance, and student self-motivation.

[Exceptional and Innovative Programs in Educational Leadership](https://journals.sagepub.com/doi/pdf/10.1177/0013161X02382005?casa_token=_f0Q83zYOa4AAAAA:fkPvxyUifFrtyu-4lmI4fYoNbbPG3FBHwhVP_Kmw2Iek9yotq8t6SulfS_Qg7cvWGFW29udw1YcSbg) (Jackson & Kelley, 2002)

Explores the standards of practice of strong preparation programs for school and district administrators.

# School District Leadership Training Program

# Description

Mariner College is a small college located in the West Coast region of the United States with 9,000 undergraduate and 5,000 graduate students total. Their School of Education admits about 800 students per year to 25 degree and certificate programs.

The Mariner College School District Leadership professional program admits 100 students per year. The two year program trains former teachers to become principals and superintendents.

Applicants must have completed a master’s degree. Almost all of the program’s students live off campus. As a result, they all have lives both on and off campus, and have substantial working experience and life perspective.

Leadership students take a total of seven courses, one course per quarter, while working.

One of the program’s big goals is to create and support visionary school leaders.

# The problem

Of the 100 students admitted each year, 30% of the students drop out by the end of Year 1. Out of all students admitted, the program completion rate (graduation) is 60% (or 40% attrition).

The program faculty has asked our team to use educational data mining techniques to find ways to increase their graduation rate to at least 75%.

## Sub-problems

We also know there are sub-problems faced by the program’s students, such as math and statistics anxiety. For instance, some students would rather pay taxes than contemplate the meaning of the Standard Deviation. We do not know, however, if these sub-problems relate directly to the graduation rate problem.

# The reward

Increasing the graduation rate will increase the program’s revenue.

Increasing the graduation rate will provide more highly qualified school district leaders to schools. More highly qualified school district leaders will support schools in the emerging cross-work-practice area of talent operations, both for the new school leaders themselves and for their schools.

More schools with visionary leaders will lead to more students who will learn how to learn, who will succeed in life, and who will contribute to society.

# Available data and ways to collect it

The data fall into three categories: data that already exists in the leadership program’s data systems, and which can be extracted for analysis, data which would have to be collected from leadership program stakeholders via surveys, and data activities which are extra if time is available.

The following data is available for all students (from program data systems). Note that some variables may not be accessible to outside consultants or researchers due to student data privacy laws or concerns. These variables include: Age of student, Years of experience working in a school, Previous leadership experience, Undergraduate major field of study, Ranking (percentile) of their undergraduate graduation cohort, Extent to which a student has used the available support services, Students’ annual income, Course grades in each year of the program, Residential and work zip codes or addresses.

These variables would be collected using surveys administered to students two to three weeks into each quarter: Self-efficacy, Professional interests, Social interests, Career goals, including how they expect the School District Leadership program will enhance their careers, Career obstacles.

Extra data to collect, if there is time, include group concept maps of “School” & “Leadership” - most likely using the CMapTools software. Concept maps can reveal hidden issues and variables, and can also create connections among participants.

# Analysis Plan

The analysis plan has 3 parts: Data preparation and data cleaning, descriptive statistics analysis, and the method for the analysis.

# Data preparation and data cleaning

To prepare and clean the data, the proposal is to do the following:

* **Data format**: One row for each student with one column per variable.
* **ID:** Students identified with an anonymized Student ID and no personal identifiable data.
* **Preparation**:
  + Encode binary variables with 0 and 1 (“Previous leadership experience” and “Use the available support services”).
  + Dummy code the categorical variable (“Undergraduate major field of study”).
  + Set “Zip code” aside for this analysis, Numerical variables with their respective scales.
  + Encode dependent variable (“Drop out”) with 0 and 1.
* **Exceptions**: Rows with missing data (except for Zip code) will be excluded from the analysis.

# Descriptive statistics analysis

Analyze for each:

* Numerical variable:
  + Mean
  + Rank
  + Standard deviation

Some conclusions will be extracted from them to the extent to recognize uniform or dispersed values for each variable.

* Categorical and binary variables:
  + Frequency of occurrence of the DV

Some observations will be made to see disparities inside the sample.

1. Method for the analysis

Using the variables described before with the proposed encoding, we will do the following:

1. **Training and testing samples** (80% - 20% proportion):  
   Split the data randomly in training and testing samples (with an 80-20% proportion).
2. **Principal Components Analysis** (PCA):  
   Do a Principal Components Analysis (PCA) to reduce variables.
3. **Interactions of binary variables over numerical**:  
   Encode interactions between the binary variables over the other numerical variables, but not within the binary variables.
4. **Models**:
   1. Logistic regression
   2. Naive-Bayes
   3. Decision tree
5. **Confusion matrix and performance parameters**:  
   Evaluate the confusion matrix for each model, and performance parameters (selected after the descriptive analysis to account for possible disparities) for both datasets.
6. **Choose best model**:  
   Decide the best model to predict the drop out of a student (in the case of Logistic Regression and Naive-Bayes, having the probability).

# Results and discussion

The section of results and discussion has 4 parts: expected results, potential implications, possible challenges for improvement, and model limitations.

1. Expected results
   1. **Accuracy greater than 85%**  
      One of the important expected results is the performance of the model where we hope we can get an accuracy over 85% of prediction for probabilities above 0.5.
   2. **Consider false positives to be better than false negatives**  
      As the program is intended to support students that are likely to drop out, false positives will be considered to be better than false negatives. Selected performance measurements will need to account for that.
   3. **Analysis of probability predictions**  
      We expect to see the comparison of the probability results for each type of the categorical and binary variables by observing their means and standard deviations. Get conclusions from probability to drop out for students categorized with the different characteristics.
2. Potential implications
   1. **Support**Targeted support for students
   2. **Program**Improve the program accordingly to the characteristics that were found to be more accountable for a good prediction of drop out.
   3. **Selection**  
      Adjust the selection process criteria according to results of the statistical significance of each parameter in the predictive model.
   4. **Effectiveness**  
      Evaluate the effectiveness of available support services by identifying how they impacted past students that either dropped out or ended up in the program.
3. Possible challenges for improvement
   1. **Methods**  
      Try different methods using other predictive methods such as logistic regression with a feature selection algorithm, use SVC, or use Neural Networks.
   2. **Variables**  
      Adding more variables, such as the additional ones listed before that might be collected from now and used in the future to improve the model.
   3. **Clusters**  
      Doing a cluster analysis to identify similar groups, using more variables than the categorical and binary variables.
   4. **Social Network Analysis (SNA)**  
      Developing a Social Network Analysis (and consider Zip code) to see if interactions between students during the first year have an effect on the probability of dropping out.
4. Model limitations
   1. **Significance**  
      Uncertainty about statistical significance and effect size. We rely on the basis that at least one of our models will be statistically significant with a moderate-to-large effect size, which might not be the case.
   2. **Disparities**  
      There may be disparities inside the sample in independent variables and dependent variables could have a significant effect on performance
   3. **Specificity**  
      The model does not suggest actions to be taken, only gives probability of dropping out, and some quick conclusions

# Conclusions

While we can’t guarantee that these analyses will lead to an improved graduation rate, our team is confident that the proposed educational data mining (EDM) analyses will reveal opportunities for program improvement worth discussing with all stakeholders.

The Mariner College School District Leadership program is more important now than ever. Please do reach out to the GDCA team at the enclosed address. We will also reach out to you in the next few weeks.

***Increased\_Graduation\_Rate <- Fortune\_Favors(PreparedMind, EDM, GDCA) + bold(now)***