IMPACT OF HIGH SCHOOL CURRICULUM ON UNIVERSITY PERFORMANCE OF STUDENTS
CAPSTONE PROPOSAL

Reshma Elizabeth Roy Kurian,
Graduate student MS in Data Analytics
Rochestor Institute of Technology, Dubai

Advisors: Dr. Yousef Al Assaf,
Khalil Darwish,
Mark Corsi Wentzel (External)

TABLE OF CONTENTS

Abstract

Performance of students in universities and colleges is an important aspect to success in their future life and career. The factors which influence the performance of students could be diverse and could include their academic prowess as well as non-academic socioemotional skills and personalities.

In Dubai, the diversity of students taking undergraduate courses is large not only in their nationalities, but also in the type of high school curriculum they have pursued in school. As students go through 12 + years of learning in a school environment, the school climate could influence the student in his academic and non-academic skill sets.

This capstone project will explore the influence that high school curriculum has on the performance of students in universities and colleges.

The project will be conducted using the data of undergraduate students to identify the differences in academics & behaviors between students coming from different school curricula.

Academic data, from high school & university, and non-cognitive skill assessment data will be mined for patterns & insights on the impact of high school curriculum and student behavior on the student's Grade Point Average (GPA) at university.

In this study, the data of students of Rochester Institute of Technology (RIT), Dubai will be analyzed to achieve the goals of the research as described above.

Statement of the problem

The purpose of this project will be to discover the relationship between high school curriculum and student behavior with student performance in universities using educational data mining techniques to help the education community make better informed decisions to improve the chances of success of students at universities.

Background Information

Dubai has the right population for this project due to its diversity in students.

Close to 70% of all university students in Dubai are residents (Higher Education in Dubai, 2017). Hence it is very important to understand the high school landscape of Dubai.

Education Landscape of Dubai

According to Knowledge and Human Development (KHDA, 2018), parents & students have a wide spectrum of educational options for schools and it is highly privatized with over 89% students attending private schools, offering one of the 17 different curricula as shown in Figure 1.

KHDA is responsible for inspecting all private schools and colleges in Dubai.
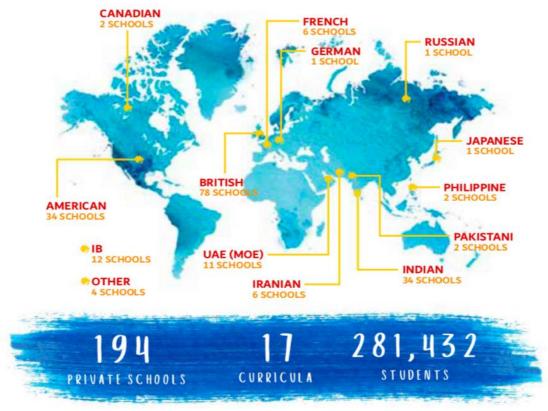
Figure 1. Private school landscape of Dubai, KHDA (2018)

Each of these school curricula mold their students differently according to their respective features & focus areas. These differences in approach, strengths & possible areas of improvement may have an influence on their academic performance & GPAs when they enter universities.

Reliable & objective information about these differences can provide a holistic approach to college readiness of high school students & help universities identify and provide targeted support for students.

This can have a positive effect on undergraduate graduation rates and on time graduation rates.

Admission Process in Rochester Institute of Technology, Dubai

The primary criteria for admission are the high school scores. The minimum scores required to be considered for admission to an undergraduate course is different for different high school curriculum. This and other admission criteria are detailed in Table 1.
The data presented has been received from the admission's director at RIT Dubai, Mr. Mustafa Al Hariri.

The Dubai Ministry of Education policy on cut off marks is followed at RIT for each curriculum based on their difficulty levels.
Physics and Math placement tests are conducted for students to decide which subjects need to be taken by the students. These aren't admission tests.

Table 1. *Undergraduate admission criteria at RIT Dubai*

There isn't a formal interview process or any scoring of students according to their

Scores	Criteria for admission
High school cut off marks	American Curriculum- >70%, Indian curriculum- >50%, Canadian curriculum- >70%, French curriculum: >12/20, Iranian curriculum: >14/20, International Baccalaureate: >24 points, British curriculum -5 subjects "O" level with grades > C, an "A" level at least 1 subject or "A" in 2 subjects if O level grade> D
Physics placement test	Score <11: Student takes Pre-Uni physics Score >12: Student takes Engineering lab, University physics P1
Mathematics placement test	Score<5: Student takes algebra & pre-calculus 6<Score <18: Student takes pre-calculus Score>18: Student starts Calculus
English proficiency test	A minimum score of 5 in IELTS (EmSat score 1100) is required to be enrolled. Score between 5 & 5.5 : Student takes English Language Literacy Score = 6 : Student takes writing 100 Score > 6.5 : Full Acceptance

perceived non-cognitive skills.
A judgment of such skills is subjective and could as it is made based on the personal statement sent by students during the application process and during the interaction of the RIT admissions team with the students during the admission process.

Stakeholders

All the educational decision makers are stakeholders in this project as the insights can improve the quality of education at all levels of the education community.

The following are the main stakeholders who will benefit from the present study:
Educators at Colleges/Universities: College & Universities can provide targeted student support according to needs of the high school curriculum pursued by the student

Educators at Schools: The insights could help schools adapt to improve the college readiness of their students

Education policy makers: The insights will give policymakers a snapshot of the academic or non-academic factors that are key to success at universities. This can help them better the curricula or policies for increasing the success of future students.

Parent-student community: The insights will be useful to the parent-student community as they are the population directly impacted by any improvements in student preparation and performance.

Project definition & Goals

The goal of this project is to provide an insight into the impact of different high school curricula on the university performance of students by identifying the strengths & weaknesses of undergraduate students using educational data mining techniques for Rochester Institute of Technology, Dubai.

This project aims to answer the following 3 research questions:
Q1. What impact does the high school curriculum have on performance and success at university? Success here will be measured by GPA, graduation on time and attrition from university.
Q2. Are there any specific subjects where students from a curriculum face difficulty?
Q3. What is the impact of student behavior on their success & performance at university during their university life?
If time permits, machine learning model will be developed to predict the performance of students based on their high school curriculum and scores.

Literature Review

A systematic literature review has been conducted to examine relevant work done in the past.

Search Strategy
An extensive search for research papers was conducted to try answering the proposed research questions using the keywords:
(Student performance) OR (School curriculum) OR (Educational data mining) OR (GPA prediction) OR (Performance prediction) OR (Performance clustering)

Searched databases include: IEEE Xplore, Google Scholar, Science Direct, ELSEVIER, RIT digital Library.

The topic of academic performance of students has been studied by various researches which shows that it is a critical topic for the education community.

However, a comparison of the different national curricula based on their impact on student performance in post high school context remains a vastly unexplored area.

Many studies have been done using different instruments to measure non-cognitive or socio emotional skills and study their influence on academic performance. The importance of these skills in the success of students is evident from the volume of focused study carried out by many researchers in the past as well as institutions such as The World Bank (Kattan, 2017).

Noftle, E. E., & Robins, R. W. (2007) examined the impact of Big Five personality traits (John & Srivastava, 1999) on GPA and SAT.
According to John & Srivastava, there are five replicable, broad dimensions of personality, and they can be summarized by the broad concepts of Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to experience.
Noftle & Robins found that some aspects of Conscientiousness (achievement-striving, persevering, and self-control) have a directly proportional influence on both

high school and college achievement. The study also found that Openness has a robust relation to SAT verbal scores, but no clear association to GPA.

Horn et al (2001) studied the relationship between the post-secondary success and the level of high school curriculum. The study found that students who had pursued a more rigorous high school curriculum than just the core curricula, developed persistence, staying continuously enrolled in the same institution and staying on track to a bachelor's degree if they transferred. The study doesn't compare different national curricula, as is being proposed in this capstone, but different intensities of high school curricula.

West et al (2016) studied student data from oversubscribed charter schools and open-enrollment district schools in USA for relationship between academic achievement and 4 prominent non-cognitive skills, Conscientiousness, Self-control, grit and growth mindset. The study showed the significant impact schools have on the non-cognitive skills of students as assessed through self-reports. It shed light on the paradox caused by reference bias when comparing the relationship at school levels as oversubscribed charter schools show a lower relationship rate despite higher academic scores and attendance. This study highlighted the need for creating unbiased measures of non-cognitive skills that are robust across a range of school settings.

Educational data mining is a field that exploits statistical, machine-learning, and data-mining algorithms over the different types of educational data (Romero et al, 2010).

Shahiri et al (2015) reviewed the different data mining techniques used in 30 different research papers in the past to predict student performance. The Attributes that have been most used are Cumulative Grade Point Average (CGPA), Internal assessment scores, Students demographic, external assessments, Extra-curricular activities, high school background, Social interaction network and Psychometric factors. The most popular task to predict students' performance is classification with Neural Network & Decision Tree having the highest prediction accuracy of 98% & 91% respectively.

Blanch & Aluja (2013) demonstrated in their study the advantages of using Regression Trees for studying academic performance over regression analysis and structural equation modelling (SEM).

Methodology

In order to analyze the impact of high school curricula on later performance, this project will study high school academic performance & skill sets of the students.

Research paradigm
All the data will be used to find patterns and make inferences about the reality of the influence that high school curricula have on the student performance in undergraduate courses. This data and the findings will be quantitative and empirical

in nature. Hence this knowledge will be produced in a quantitative research paradigm embedded in a **positivist** epistemology approach.

For positivists, knowledge is derived from the facts of experience and observation of the natural world.

Research Population

The project will use student data from RIT Dubai. The research population will include undergraduate student data of the years 2010-2019.

Pending confirmation from the Institutional Research Ethics Board (IREB), non-cognitive skill data will be collected using a self-assessed personality assessment test among current 1st year undergraduate students and new student enrolling for summer 2019 & fall 2019.This will be used as another variable used to support the diagnostics from the academic data. The sample size of the population of this variable will be small as it pertains only to a single cohort of engineering undergraduate students.

Research Design method

For the academic data collected from RIT, this project will follow an experimental quantitative research design method to examine the cause effect relationship of the variables. A deductive approach will be used on the data collected to test the cause effect relationship of the high school curriculum on the GPA variables, dropout rates and timely graduation variables.

IREB approval was sought and granted for the collection of anonymous data of students.

The data that will be collected are:
From High School: High school scores, high school curriculum, 12th grade stream
From RIT: Mathematics placement scores, Physics placement scores, English language Proficiency scores, 1st year GPA, CGPA, High school completion year, Graduation on time data, Dropout data.
Of these variables,

      2 dependent variables are: Dropout rates and Graduation on time rates.
      2 intervening variables are: 1st year GPA and University GPA
      All the remaining variables are Independent variables

For collecting non-cognitive skill data, the survey instrument will need to be identified before seeking approval from IREB.
The various instruments being considered are Cattel's 16 Personality Factor Test (Cattell et al,1972), Big Five Inventory (John & Srivastava, 1999), surveys from Search INSTITUTE and SuccessNavigator (Rikoon et al, 2015).

Educational Data Mining (EDM)

Data mining techniques will be used to answer the research questions and predict the support required for students from different curriculum to succeed in their undergraduate programs at universities.

Multiple regression techniques using regression trees and clustering will be used to analyze the data collected. The relationship between the variables will be studied with few independent variables controlled.

A statistical analytical tool such as MATLAB/R studio/ Tableau / SPSS (Statistical Data Analysis Software) will be used for the statistical analysis.

For machine learning, a subset of the data collected ($1/3^{rd}$) will be reserved for testing of the machine learning model. All the remaining data will be used to train the machine. Various techniques of supervised learning such as regression, k-means clustering, decision trees etc will be employed. The objective of creating the machine learning model is to develop a model which can predict the future university GPA of a high school student from his high school curriculum, scores and non-cognitive skill sets.

Project Evaluation

The project can be evaluated for achievement of each of the project goals.

1. Impact of the high school curriculum have on performance and success at university?
The project should give clear insight into the role played by high school curricula on the performance of students at universities.
3 measures which would indicate the performance of students are:
- GPA
- Graduation on time
- Retention at university

2.Identify specific subjects where students from a curriculum face difficulty
A subject will be considered as difficult if more than 20% of students of a certain curriculum score lower than the average score. The data used will be the placement test scores in Mathematics, Physics and English proficiency scores.

3. Impact of student behavior on their success & performance at university during their university life?
Student behavior will be measured as the personality type of the student as assessed in the psychometric instrument chosen.
The student performance will be measured in terms of the GPA, credit scores and Retention at university.

Note: The results of this research question are subjective to the sample size of the data that will be collected.
The non-cognitive assessment will be a consent-based test and will be sent out only to current students limiting the sample size. Hence the reliability of the result will depend on the percentage of positive respondents.

4. Machine learning model to predict the performance of students based on their high school curriculum and scores.

2/3rd of the data population will be used to train the model. The remaining 1/3rd population will be used to test the model.

The input that will be given into the predictive model will be high school curriculum, high school scores and non-cognitive skill set.

The prediction will be made in terms of range of university CGPA of the same student.

An accuracy of 60% and above should be considered successful.

The accuracy of the predictive model will be dependent on the volume of training data.

The project should propose the academic or non-academic support at the high school and university levels that would help improve the college readiness of the student.

Project deliverables

Final capstone project with descriptive analytics giving insights into the impact of school curricula on university success of students using visual analytics.

If time permits, a prototype of the predictive model that will predict the college GPA from high school scores, curriculum and non-cognitive skills.

Project Timeline

The project is expected to be completed by the 10th of December 2019. The different phases of the project life cycle are given in Figure 2.

There are 2 critical steps which will determine the success of the project and the possibility of developing a machine learning predictive model.

These are:
1. Collection of academic data for Rochester Institute of Technology, Dubai
2. Sample size of population for non-cognitive skill assessment.

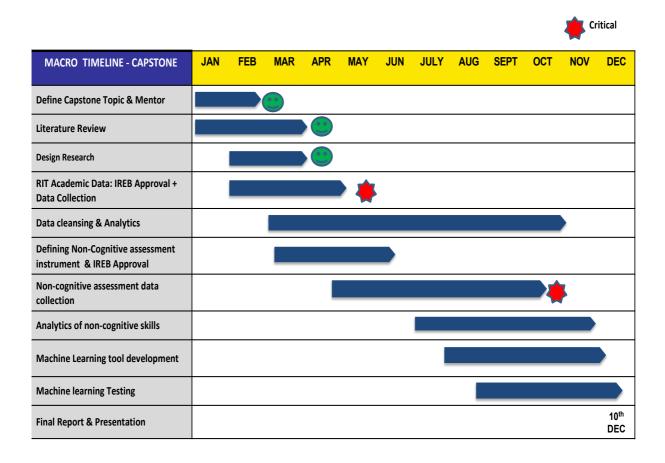| MACRO TIMELINE - CAPSTONE	JAN	FEB	MAR	APR	MAY	JUN	JULY	AUG	SEPT	OCT	NOV	DEC
Define Capstone Topic & Mentor												
Literature Review												
Design Research												
RIT Academic Data: IREB Approval + Data Collection												
Data cleansing & Analytics												
Defining Non-Cognitive assessment instrument & IREB Approval												
Non-cognitive assessment data collection												
Analytics of non-cognitive skills												
Machine Learning tool development												
Machine learning Testing												
Final Report & Presentation												10th DEC

Figure 2. Macro timeline of the proposed capstone project

Project Budget

The expected expenditure for the project would include

1. Cost of non-cognitive skill assessment instrument like SuccessNavigator : less than 10$ per student (if a commercial test instrument is chosen).
2. Cost of SPSS license – 89$

As the project benefits RIT Dubai, a request would be made for funding.

References

- Horn, L., Kojaku, L. K., & Carroll, C. D. (2001). High school academic curriculum and the persistence path through college: Persistence and transfer behavior of undergraduates 3 years after entering 4-year institutions. *Washington, DC: U.S. Dept. of Education, National Center for Education Statistics.*

- Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big Five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology.* 93(1), 116-130. doi:10.1037/e514412014-495

- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *In L. A. Pervin & O. P. John (Eds.), Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford

- West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., & Gabrieli, J. D. (2016). *Promise and Paradox. Educational Evaluation and Policy Analysis*,38(1), 148-170. doi:10.3102/0162373715597298

- Blanch, A., & Aluja, A. (2013). A regression tree of the aptitudes, personality, and academic performance relationship. *Personality and Individual Differences*,54(6), 703-708. doi:10.1016/j.paid.2012.11.032

- Rikoon, S. H., Liebtag, T., Olivera-Aguilar, M., Steinberg, J., & Robbins, S. B. (2015). Anticipating College Enrollment: AdaptingSuccessNavigator®for High School Students. *ETS Research Report Series*,2015(2), 1-26. doi:10.1002/ets2.12084

- Romero, C., & Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* 40, 601-618.

- Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A Review on Predicting Students Performance Using Data Mining Techniques. *Procedia Computer Science*,72, 414-422. doi:10.1016/j.procs.2015.12.157

- Kattan, R. B. (2017, May 09). Non-cognitive skills: What are they and why should we care? *Retrieved from http://blogs.worldbank.org/education/non-cognitive-skills-what-are-they-and-why-should-we-care*

- Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1972). 16 PF: Sixteen personality factor questionnaire. Champaign, IL: *Inst. for Personality and Ability Testing.*

- Knowledge and Human Development Authority (2018). *Dubai Private Schools. A decade of growth. Key Findings 2008-2018.* Dubai, UAE: KHDA

- Knowledge and Human Development Authority (2017). *Higher Education in Dubai.* Dubai, UAE: KHDA