



A POLYTECHNIC INSTITUTION

Predictors of Student Success

Analysis of School of Business First Term Students

**Prepared for:
BCIT School of Business**

June, 2004



The path you choose can make all the difference.

Executive Summary

Objective of the Report

The objective of this report is to provide information to the School of Business regarding possible indicators that may influence the success of students. The report examines student characteristics and how those characteristics influence academic success, as defined by a student's grade point average (GPA), for their first term at BCIT. The data for this report consisted of just over 1500 student records selected from the fall terms of 1998, 2000, 2001 and 2002, where the student record contained sufficient information.

Purpose of the Report

The purpose of this report is to establish correlations between student successes as measured by their grades and/or courses completed with explanatory variables, such as previous grades and various demographic data collected by BCIT. This report builds from a previous effort which both similar data and analytical approaches were used but yielded mixed results. The analytical approaches used in this report consist of basic statistical techniques and more comprehensive data mining techniques.

Summary of Analyses

- 85% of students cite their objective for attending BCIT's School of Business as being career or job driven, whereas the remaining 15% cite their objective as being non-work related.
- 64% of School of Business students have some level of post secondary education or work related experience, whereas the remaining 36% of students are admitted directly from high school.
- Gender distribution is relatively equal in most programs (50.1% female, 49.9% male). Some programs indicate a self selection bias, such as Human Resources Management have more females compared to males, whereas the International Trade and Transportation and Operations Management programs have more males compared to females.
- The largest cohort of students attending the School of Business are between the ages of 20-24 years of age, whereas the smallest cohort of students are 19 years or younger.
- The data analysis inferred a correlation between the age and average GPA of all School of Business students during their first term. Students who were 19 years or younger had an average GPA of 65%, whereas students over 30 years had an average GPA of 74%; thus indicating a 9% difference between these age groups.
- School of Business students have an average overall GPA of 70%. Human Resources Management scored the highest at 78%, whereas International Trade and Transportation had the lowest GPA at 64%.
- Males and females perform relatively equal with respect to average GPA during their first term at BCIT (69.0% and 70.3%, respectively).
- Microeconomics was indicated as having the highest failure rate, at 18%, for School of Business first term students. Business Math had the second highest failure rate at 14%.
- The success rate, defined as having no grade lower than 50%, of students with respect to core courses was 75%. Human Resources Management had the highest success rate (91%), whereas International Trade and Transportation and Management Systems had the lowest success rates (62% and 64%, respectively).
- The data inferred that younger students, 19 years and under, failed more core courses when compared to other age groups. Students over the age of 30 years had the lowest failure rate of core courses.
- When analyzing factors related to students overall GPA, the previous education of these students was indicated as having the highest impact on their performance. The data mining indicated the higher the level of previous education the student had, the higher the average overall GPA they achieved. The second most distinguishing factor for determining overall GPA was English 12 scores.
- When assessing success of students in their core courses, their previous math grades were determined as having the most significant impact on their average core GPA. The data mining indicated a correlation between a higher math grade and a higher core GPA. Further analysis indicated that depending on the student's previous math grade, three factors (age group, previous education or English 12 scores) were related to their success.

Table of Contents

INTRODUCTION	1
PROJECT PURPOSE	1
BACKGROUND RESEARCH	2
REPORT STRUCTURE	2
CONSTRAINTS.....	3
MISSING DATA	4
INCONSISTENT DATA ENTRY	4
LACK OF DATA	4
I: DEMOGRAPHICS OF SOB FIRST TERM STUDENTS.....	5
A. OBJECTIVES FOR ATTENDING BCIT.....	5
<i>Figure 1.1: Student Objectives for Attending BCIT</i>	<i>5</i>
<i>Figure 1.2: Student Objectives for Attending BCIT by Age.....</i>	<i>6</i>
B. PRIOR EDUCATION OF FIRST TERM STUDENTS.....	6
<i>Figure 1.3: Students Previous Education Level Prior to Attending BCIT.....</i>	<i>6</i>
<i>Figure 1.4: Education Level Prior to Attending BCIT by Program</i>	<i>7</i>
C. AGE AND GENDER PROFILES OF FIRST TERM STUDENTS	7
<i>Figure 1.5: Gender Breakdown by Programs</i>	<i>8</i>
<i>Figure 1.6: Average Age of Programs.....</i>	<i>8</i>
<i>Figure 1.7: Age Group Breakdown of Programs</i>	<i>9</i>
II: STUDENT PERFORMANCE ANALYSIS OF FIRST TERM IN SCHOOL OF BUSINESS	9
<i>Figure 2.1: Average GPA by Age Group</i>	<i>10</i>
<i>Figure 2.2: Average GPA by Program</i>	<i>10</i>
<i>Figure 2.3: Program Average GPA by Age Group</i>	<i>11</i>
<i>Figure 2.4: Program Average GPA by Gender.....</i>	<i>12</i>
III: STUDENT PERFORMANCE ANALYSIS OF FIRST TERM COMMON CORE COURSES... 	12
<i>Figure 3.1: Failure Rate of Common Core Courses</i>	<i>13</i>
<i>Figure 3.2: Success Rate of Programs in Core Courses</i>	<i>13</i>
<i>Figure 3.3: Failure Rate of Core Courses by Age Group.....</i>	<i>14</i>
<i>Figure 3.4: Failure Rate of Core Courses by Gender.....</i>	<i>15</i>

IV: RESULTS FROM DATA MINING.....	16
OVERVIEW.....	16
UNDERSTANDING DATA MINING SOFTWARE.....	16
<i>Figure 4.1: A Sample Node from Scenario.....</i>	<i>17</i>
SCENARIO AND BCIT.....	17
CHARACTERISTICS OF STUDENTS.....	18
<i>Figure 4.2: 4.0 Point Scale.....</i>	<i>18</i>
<i>Figure 4.3: Regions by Code.....</i>	<i>18</i>
<i>Figure 4.4: Student Objectives for Attending BCIT.....</i>	<i>19</i>
<i>Figure 4.5: Age Group Categories.....</i>	<i>19</i>
<i>Figure 4.6: Previous Education Level Categories.....</i>	<i>19</i>
FACTORS THAT INFLUENCE STUDENT’S OVERALL GPA.....	20
<i>Figure 4.7: Results of Data Mining for Overall GPA.....</i>	<i>20</i>
FACTORS THAT INFLUENCE STUDENT’S CORE GPA.....	22
<i>Figure 4.8: Results of Data Mining for Core GPA.....</i>	<i>22</i>
CONCLUSION.....	25
APPENDIX1: PROGRAM AREA SPECIFIC RESULTS.....	29
APPENDIX 2: RECODING OF BCIT’S FIRST TERM GPA.....	35
<i>GPA Data Frequencies by Program / Option.....</i>	<i>40</i>
APPENDIX 3: RECODING PROVINCIAL ENGLISH AND MATH MARKS.....	41
TOEFL SCORE.....	42
COMP SCORE.....	42
TWE SCORE.....	42
<i>English Class Code.....</i>	<i>43</i>

Introduction

One of the most debated yet understudied academic issues has been the setting of pre-requisites and how that translates into student success at BCIT. That is to say what is the proper set of pre-requisites that balances effort with effectiveness in ensuring that the students who are accepted into programs have the set of characteristics that maximizes their chances of successes. The balance is to produce a set of pre-requisites that is neither too restrictive nor not restrictive enough.

The objective of this report is to provide information for the School of Business with respect to possible indicators that may influence student performance. The report examines student characteristics and how they influence academic success as defined as the student's grade point average (GPA) while completing their first term at BCIT. The number of students records included in this analysis was just over 1500 taken from the following fall terms, 1998, 2000, 2001 and 2002.

The main analysis tool used in this report is the user-friendly Cognos data mining tool which determines relationships among a set of data characteristics or variables. Other analyses are also presented which help explain the data and in some cases, help focus the analysis and perhaps point to some issues, which may require further debate, by project stakeholders.

Project Purpose

The core purpose of this study is to establish correlations between student successes as measured by grades and/or courses completed with explanatory variables, such as past grades, and various demographic data collected by the Institute.

This work builds from a previous effort in which both similar data and analytic approaches were used, and which yielded mixed results. The first attempt was more a focused effort to understand and link the various sources of student data available for analysis. The analysis that was done was secondary in that a correlation package provided by *Cognos* was used to determine "best predictors".

In previous phases, missing data has proved key in what variables could be included in the model due to degrees of freedom problems. As well, a brief literature search was conducted on post-secondary research in this field as a means of validating the methodology incorporated into this analysis models and more importantly, as a means to find clues as to which variables we may be missing in comparison with these other efforts.

Background Research

The ability to forecast and predict academic performance has always been of interest to educators and administrators. Keim and Stevenson¹ developed a survey tool for determining causes and trends in student attrition within the North Carolina college system. Their survey tool incorporated both academic factors and non-intellective factors including demographic characteristics, personality variables, motivation and financial need. The objective was to create a survey that could be used at the institutional level as an Applicant Survey Questionnaire.

Levin and Wyckoff² used a logistic regression model to predict persistence and success in the engineering program at Pennsylvania State University. For pre-enrolment they found that high school GPA, algebra score, gender, chemistry score, and the reason for choosing engineering were strong predictors of success. They went on to use data from the students' freshman year to forecast success in sophomore year engineering. Here they found calculus, physics and chemistry scores to be strong indicators of success.

Sadler, Cohen, and Kockesen³ used a logistic regression model to identify students at risk of retention in first year university. The study looked at 2209 freshmen students for the fall of 1994 and fall of 1995 at NYU. The variables used included pre-college schooling, family background, first semester academic and social integration, and financial considerations. Their report concluded that their model for identifying students at risk would allow for the development of an optimum intervention strategy.

Report Structure

The first section of this report presents some basic results about the data itself. Some figures are presented which describe key variables including dependent variables, such as first term GPA. This section gives the reader a better understanding of the data in a user-friendly manner. Even as a stand-alone section, there are many interesting aspects of the data, which are discussed, that are illuminating and important to point out.

The second section provides the reader with an introduction into the data mining approaches and techniques used in for this project. Subsequent to the data mining introduction is the presentation of the results of this exercise. Following each of the graphic results are written analyses of what the data mining software indicated as having impacts on the dependent variables, first term GPA and core GPA.

¹ Keim, H. David; and Stevenson, Alice F. "Model for Determination of Student Attrition: Causes and Trends," (13 Nov 1980) ERIC ref #: ED 224 374

² Levin, James; and Wyckoff, John. "Identification of Student Characteristics That Predict Persistence and Success in an Engineering College at the End of the Sophomore Year: Informing the Practice of Academic Advising" (Feb 1990) ERIC ref # ED 319 355

³ Sadler, William E.; Cohen, Frederic L.; and Kockesen, Levent "Factors Affecting Retention Behavior: A Model to Predict At-Risk Students" AIR 1997 Annual Forum Paper. ERIC ref #: ED 410 885

The third aspect of this report deals with conclusions based on the data analysis. The conclusion section is divided into two aspects when presenting its findings. The first section contains conclusions that are of a more general basic nature and is related to both the School of Business and its programs. The second section contains conclusions that are based on the statistical analysis and relate mostly to the School of Business as a whole, but contain some insight with respect to its programs.

The attachments section of this report provides further details into the dependent variables, first term GPA and core GPA, and presents more detailed looks of that data at the program level. As well, some interesting alternative analytical approaches are discussed as well. These were attempted but not formally written up.

The table below provides an overview of each of the programs used in the data analysis.

Table 1: Project Cohort by Program

Program	Option Code	Number of Students	Share	Credential
Financial Management	FMGT	543	33.8%	2 year diploma
Human Resource Management	HRMG	48	3.0%	2 year diploma
Integrated Management Systems	IMST	36	2.3%	2 year diploma (combined)
Intl. Trade & Transportation	INTT	103	6.4%	2 year diploma
Information Tech. Management	ITMG	57	3.5%	2 year diploma (combined)
Management Systems	MGTS	142	8.8%	2 year diploma (combined)
Marketing Management	MKTG	585	36.4%	2 year diploma
Operations Management	OPMT	93	5.8%	2 year diploma
Total		1607	100%	

Constraints

Before approaching this project, it is important to note that there were a few constraints that may have had an effect on the direction and development of this analysis. It is possible that the constraints limited the development of a model that could provide further analysis beyond the previous attempts. By noting the constraints, this will allow for a deeper understanding as to why some variables were not used or treated in the manner they were. Some constraints the project met included missing data, inconsistent data entry and lack of student information.

Missing Data

Missing data and information consisted of blank fields that related to the various student characteristics used in this project. Blank fields were consistently found in all aspects in the data, ranging from student course grades to gender. Being that the information was not available it should be noted that these students were not included in the analysis, as they could not offer information to enhance the model. By eliminating these students, this lowered the sample population to be used for analysis.

Inconsistent Data Entry

Certain aspects of the data indicated inconsistent data entry, which resulted in the data having to go through an extensive cleaning exercise before it could be used in the analysis. One aspect in the cleaning of data involved the coding of student prerequisites to provide a consistent measure of their grades, as post-secondary data was entered as either a letter grade or a percentage in a free formatted text field. It should be noted that the provincial grades, especially with respect to math providing a range of levels, making it difficult to define.

Lack of Data

The scope of the student characteristics used was limited to the data provided by the BCIT Banner Database and the Entrance Student Survey. Based on research to determine what student characteristics could be used to further enhance the model, it was noted that currently BCIT was not able to provide student indicators beyond the general characteristics used in this model. For future reference, by expanding the student characteristics to include more complete and standardized information, such as standard provincial post-secondary exam data, may allow for stronger results in predicting the factors that contribute to a student's performance.

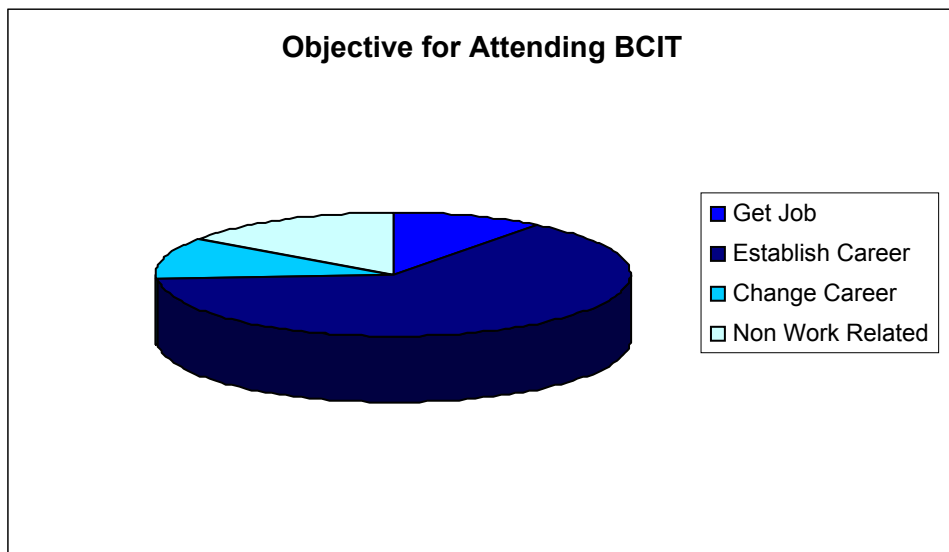
I: Demographics of SOB First Term Students

A. Objectives for Attending BCIT

School of Business Students Objective for Attending BCIT

85% students attending BCIT's School of Business cite their objectives for enrolment as being career or job driven, with 75% of students in this segment stating they wanted to establish a career. Whereas the remaining 15% of students stated their objective for attending BCIT was non-work related.

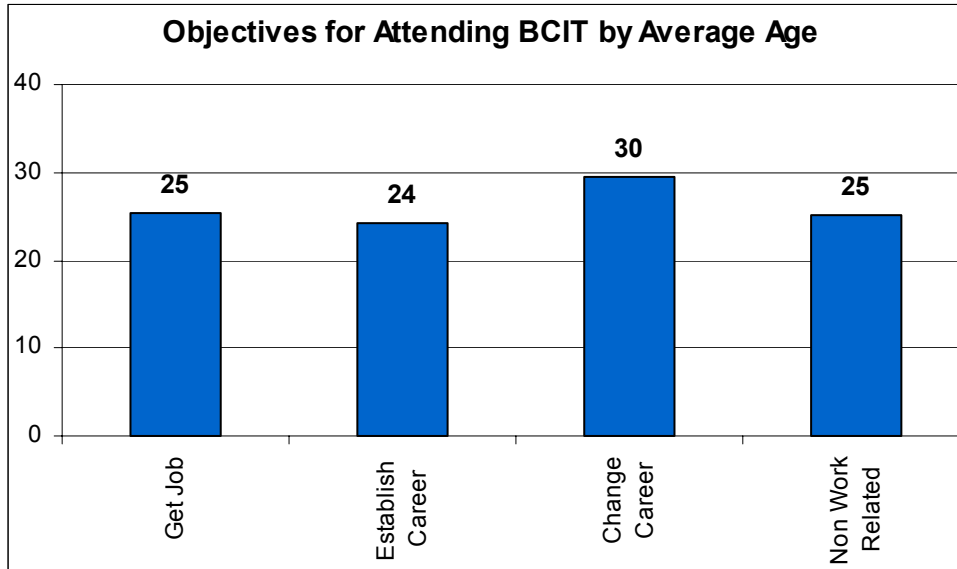
Figure 1.1: Student Objectives for Attending BCIT



Objective for Attending BCIT by Age

In looking at the motivation of students in attending BCIT, the average age is fairly equal across three of the four categories taken from the entrance survey. Those interested in “getting a job” or “establishing a career” were approximately 25 years of age. Not surprisingly, those interested in a “career change” were slightly older at an average age of 30 years.

Figure 1.2: Student Objectives for Attending BCIT by Age

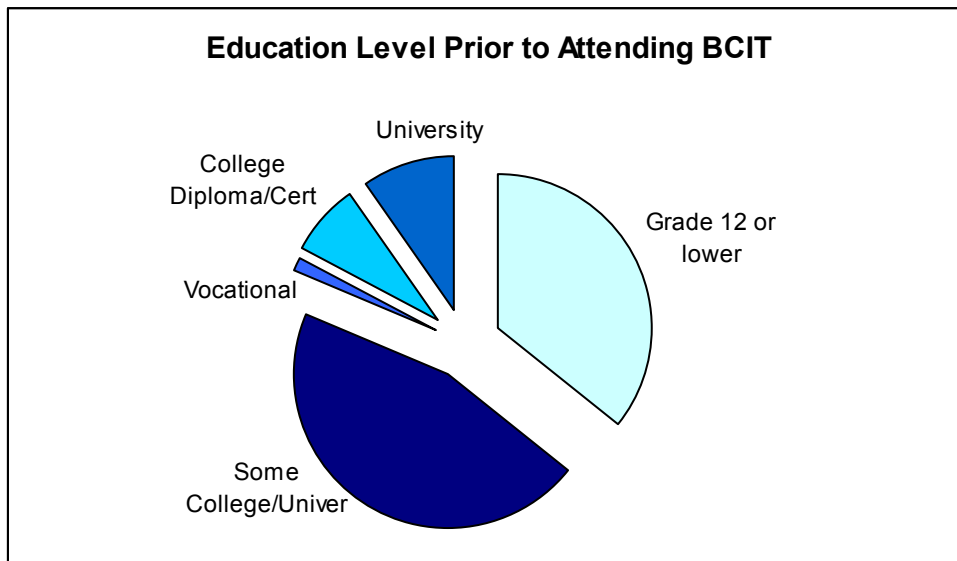


B. Prior Education of First Term Students

Previous Education Level of School of Business Students

Of all students entering first term programs in the School of Business, 64% had some form of education or training beyond high school. 46% of all entrants had some university or college and 10% had completed a degree while 7% had some type of college diploma or certificate. A small group of students (1%) had prior vocational training.

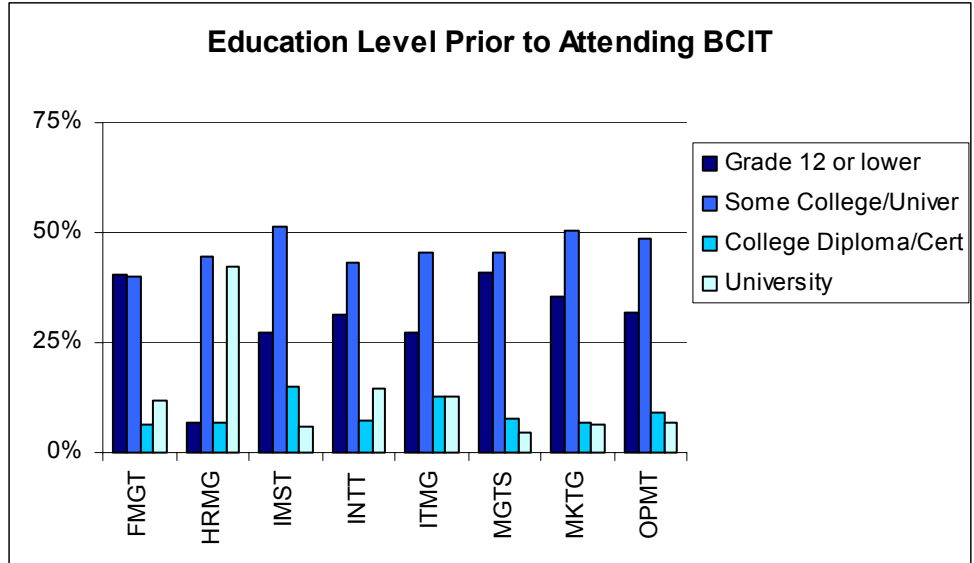
Figure 1.3: Students Previous Education Level Prior to Attending BCIT



Prior Education Level by Program

Most of the programs mirrored the aggregate statistic with the exception of the Human Resource Management Program, where 42% of first term students had previously completed a university degree. The programs with the highest percentage of students entering directly from high school were Financial Management and Marketing Management.

Figure 1.4: Education Level Prior to Attending BCIT by Program

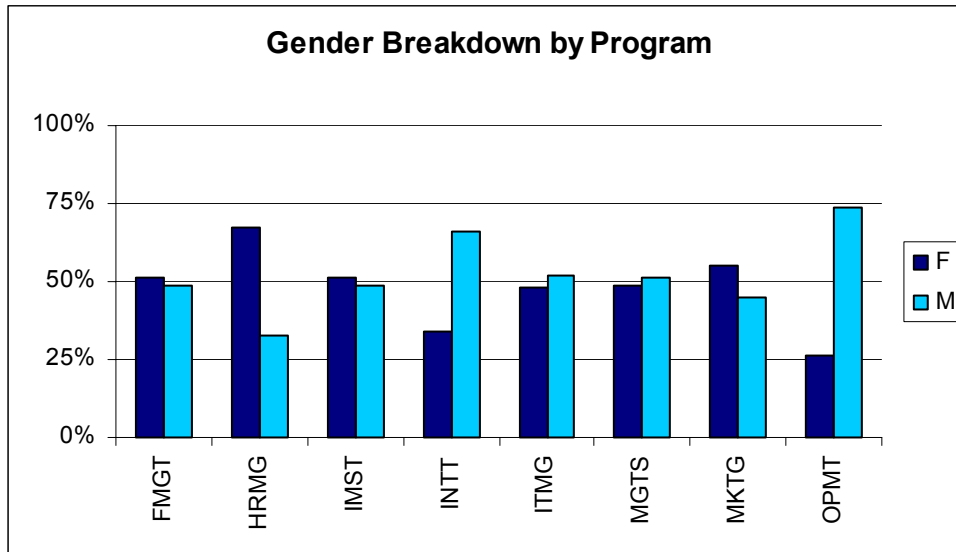


C. Age and Gender Profiles of First Term Students

Gender Breakdown by Program

Overall in the School of Business, the gender representation is almost equal (50.1% female, 49.9% male). Some programs show a gender based self-selection bias; in particular Human Resources Management has a high proportion of females to males. Whereas International Trade and Transportation and Operations Management tend to have a larger proportion of males to females relative to other programs.

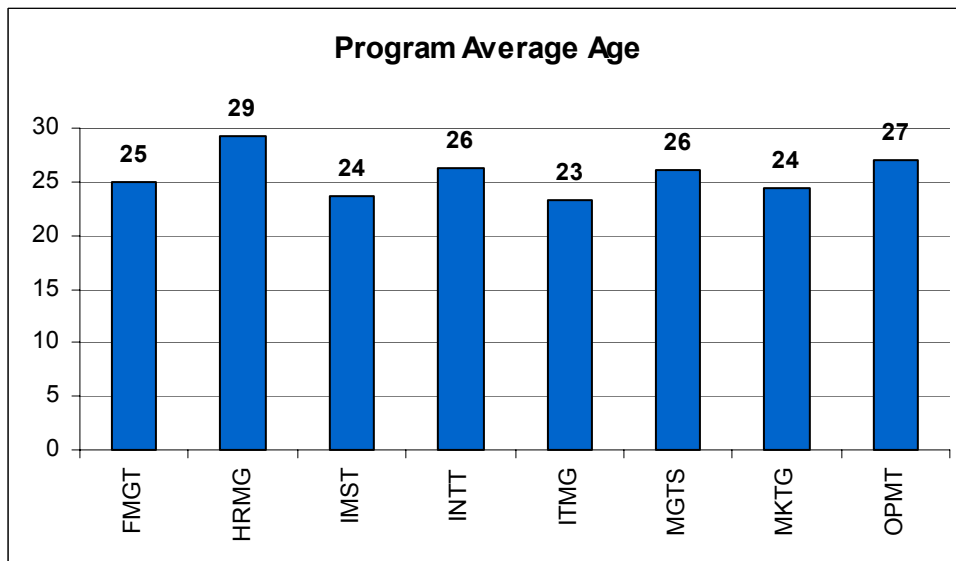
Figure 1.5: Gender Breakdown by Programs



The Average Age of Students by Program

The data indicated that the average age in the School of Business as 25 years, with a range of ages from 18 to 55 years. Many of the programs are relatively close to the average age of students, with the exception of Human Resources Management, which has an average age of 29 years. The youngest students (18 years of age) were in Financial Management, International Trade and Technology, Information Technology Management and Marketing Management. Interestingly, Marketing Management was also noted as having the oldest students at 55 years of age.

Figure 1.6: Average Age of Programs



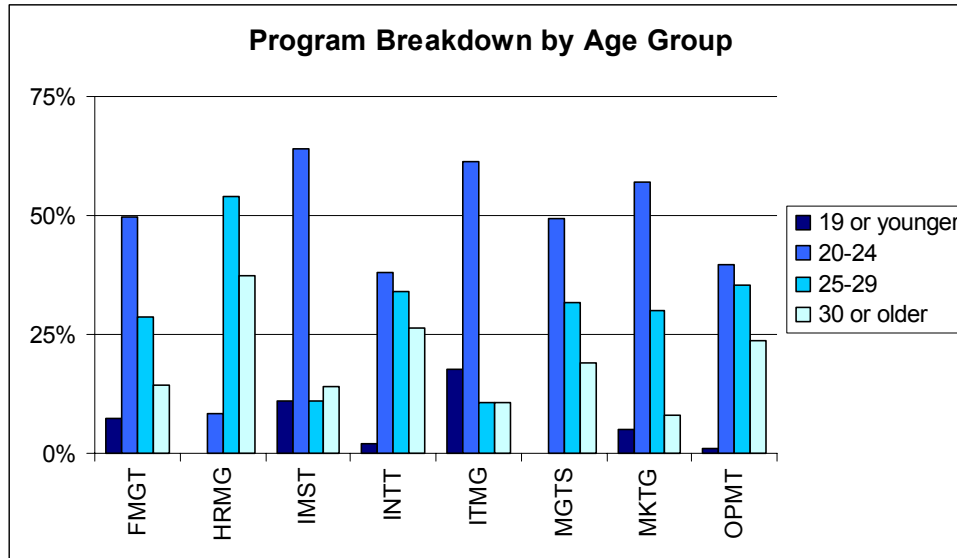
Program Age Group Breakdown

The age group breakdown for the School of Business is as follows:

- 6% of students are 19 or younger
- 51% are between 20-24 years
- 30% are between 25-29 years
- 14% are 30 or older.

The vast majority of all students in the School of Business are between the ages of 20-24 years, which is reflected in all of the programs, except for Human Resources Management, which has a large proportion of students between the ages of 25-29 years and no students under the age of 19. The remaining programs mirror the aggregate statistic, with the exception of Integrated Management Systems and Information Technology Management, which has a higher proportion of students aged 19 or under and have equal numbers of students in the age groups 25-29 years and 30 or older.

Figure 1.7: Age Group Breakdown of Programs

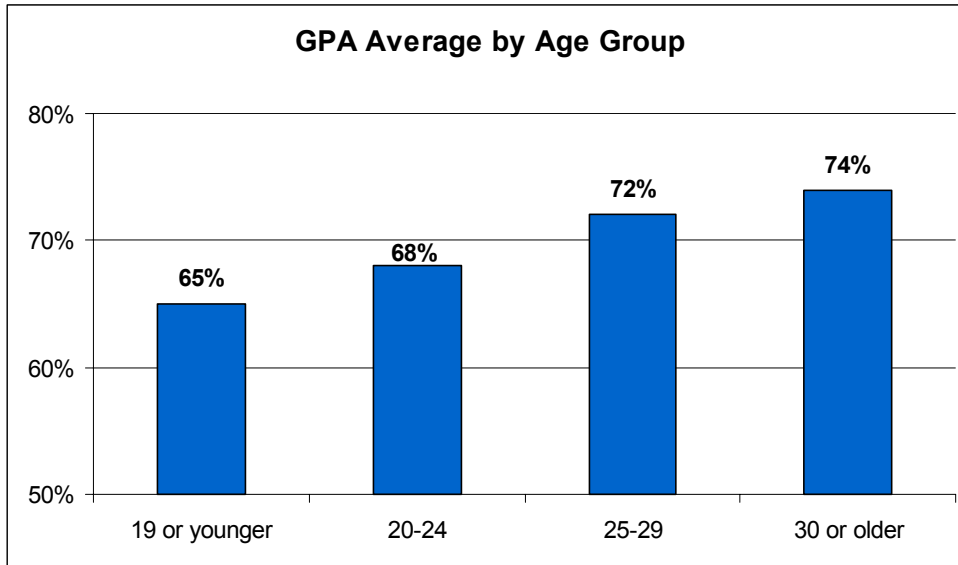


II: Student Performance Analysis of First Term in School of Business

Average GPA by Age Group

The data indicated a correlation between the age and average GPA of all School of Business students during their first term at BCIT. As noted in the graph below, there is a 9% difference in the students average GPA of the youngest cohorts of 19 years and younger (65%) when compared with the oldest cohorts of 30 years or older (74%).

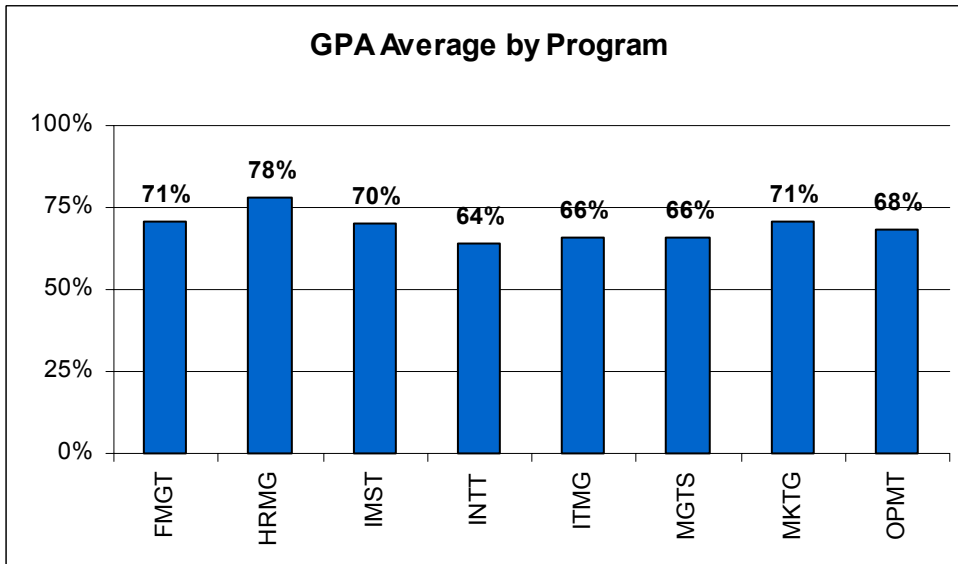
Figure 2.1: Average GPA by Age Group



Average GPA by School of Business Programs

School of Business students scored an average GPA of 70% during their first term of classes. Many of the programs indicated having an average GPA close to the School of Business average. The data indicated that students in the International Trade and Technology scored the lowest at 64%, whereas students in the Human Resources Management program scored the highest at 78%. Students in the Human Resources Management program may have a tendency to score higher due to the requirement of having an average GPA of 75% to enter the second year of the program.

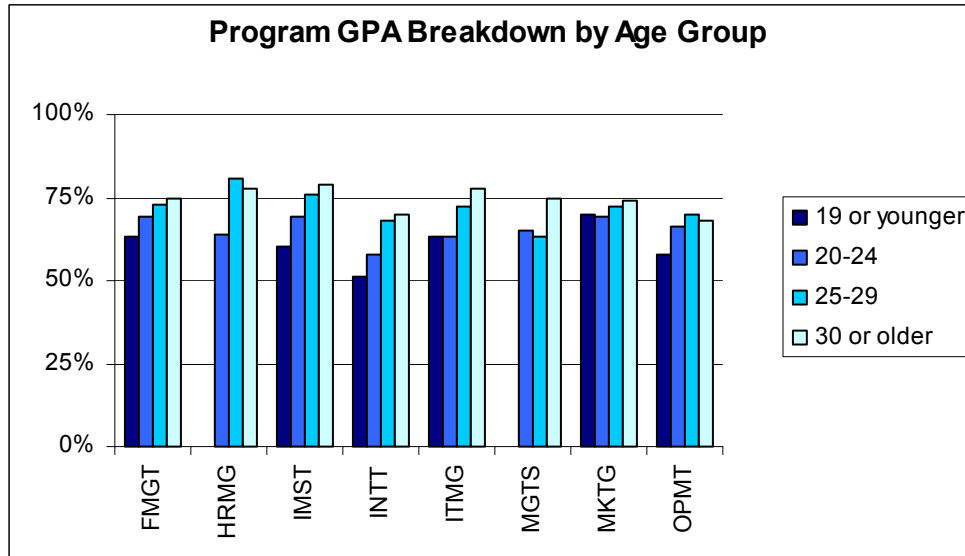
Figure 2.2: Average GPA by Program



Program GPA by Age Group

Many of the programs mirrored the statistical findings of a correlation between age and GPA. As noted in most programs, there is a rise in GPA by the increasing years of the age groups, with the exception of Human Resources Management and Operations Management where students in the age group 25-29 years were indicated to have the higher average GPA, with a difference of about 3%, when compared with their 30 or older cohorts.

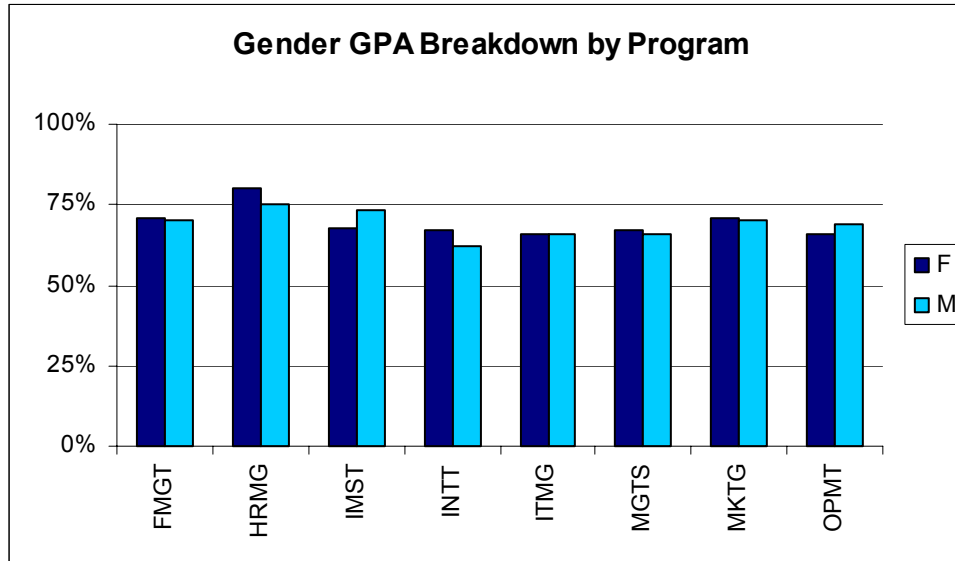
Figure 2.3: Program Average GPA by Age Group



Program GPA by Gender

When comparing the performance by gender, it was determined their performance was relatively equal (70.3% females, 69.0 males). This statistic is reflected in many of the School of Business programs, with the exception of Operations Management and Integrated Management Systems, where males scored about 3.5% higher in their average GPA when compared to their female counterparts. Females were noted as scoring a higher average GPA (relatively about 5%) than their male counterparts in Human Resources Management and International Trade and Transportation.

Figure 2.4: Program Average GPA by Gender



III: Student Performance Analysis of First Term Common Core Courses

All first term School of Business programs have certain courses in common. These are sometimes referred to as the “common core.” In particular, all first term students take Communications, Economics, Business Math and Accounting.

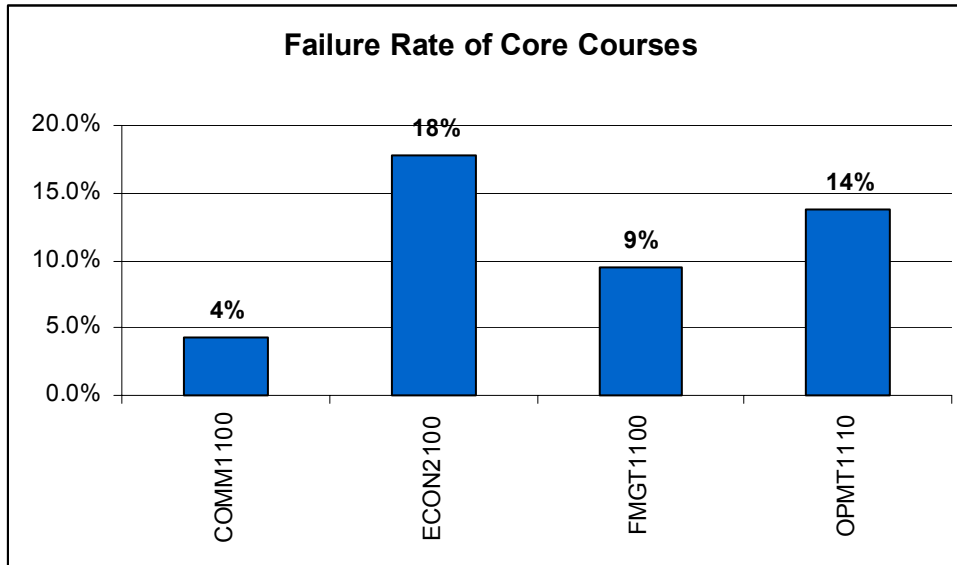
This section focuses on the performance across programs within these common core courses. Historically, these courses have been the ones most likely to present problems or challenges for first term students and are believed to have the highest rate of failure.⁴ This data includes transfer credits, which were treated as a grade of 75%. All failing grades were earned and does not include withdrawals.

Failure Rate of Common Core Courses

The following graph depicts number of fails as a percentage of students in first semester, full time programs. Microeconomics (ECON 2100) has the highest failure rate at 18%. Communications (COMM1100) has the lowest failure rate at 4%.

⁴ A failing grade at BCIT is usually less than 50%. In some programs a different benchmark may be chosen. For example, Financial Management requires a grade of 65% in Introductory Accounting for a student to continue in the program. For the purpose of this study, a grade less than 50% was considered a fail.

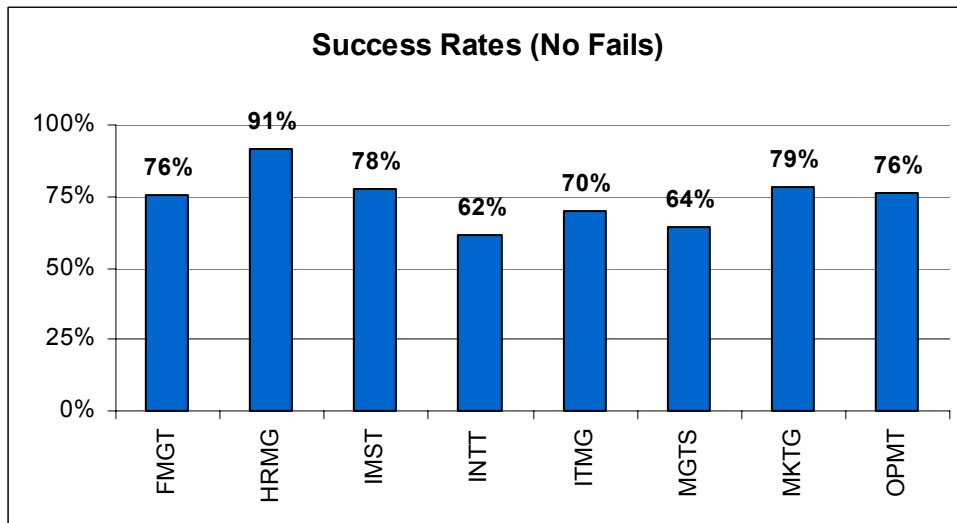
Figure 3.1: Failure Rate of Common Core Courses



Program Success Rates in Core Courses

Success rate was defined as completing first term with no grade lower than 50% in the core courses. Overall, 75% of students in the School of Business successfully completed all core courses. The following graph shows the breakdown of success rates by program. Most programs had success rates in the 70% to 79% range. Noted exceptions are the Human Resource Management with the highest success rate at 91%, whereas both International Trade and Transportation and Management Systems were below 70% (62% and 64%, respectively).

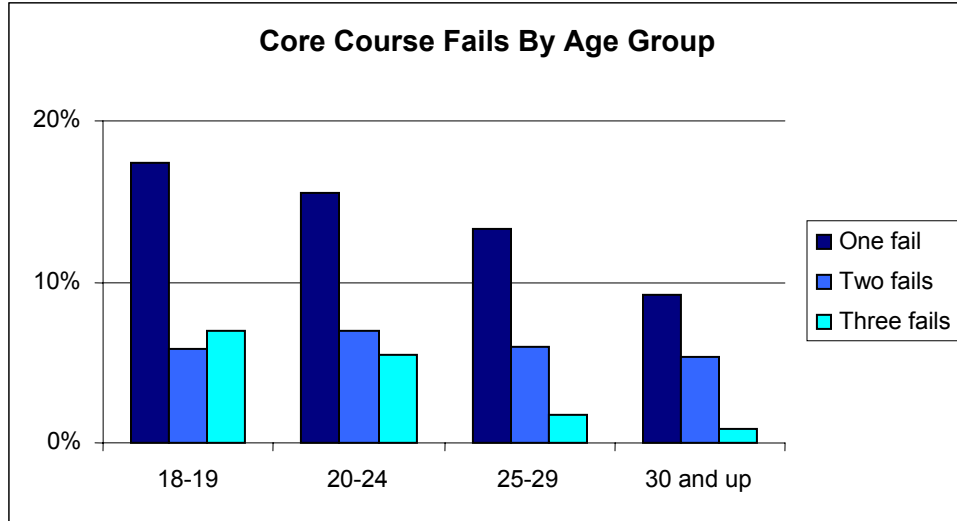
Figure 3.2: Success Rate of Programs in Core Courses



Core Courses Failed by Age Group

To provide a comparison of performance by age, the number of courses failed was a percentage of an age group, not as a percentage of the total body. The graph illustrates that an increase in the age group correlates with a decrease in the number of courses failed. In all age categories, the majority failed a single course. In the age groups 18-19 and 20-24, it can be noted that students who failed two courses were relatively equal to students who failed three courses.

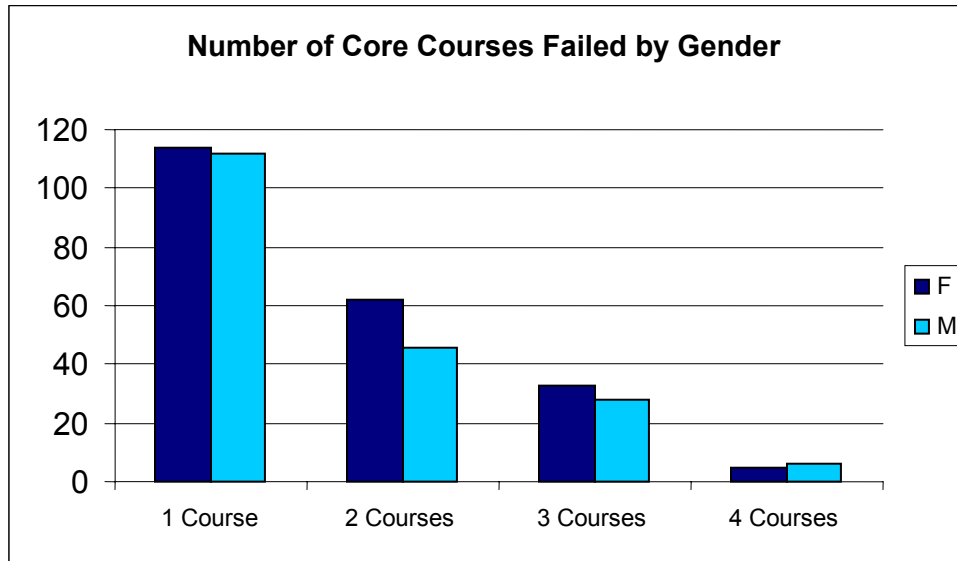
Figure 3.3: Failure Rate of Core Courses by Age Group



Failure Rate by Gender

As the following graph illustrates, there is no noticeable difference in failure rates when analyzed by gender.

Figure 3.4: Failure Rate of Core Courses by Gender



IV: Results From Data Mining

Overview

Data mining is a process, which a computer program analyzes data sets to determine linkages between variables and commonalities found within the data. Data mining is guided by the user, who sets the parameters to focus the analysis in a particular area of interest. A particular data-mining program called *Scenario* was applied to the BCIT Student database. The focus of the analysis was to find common characteristics of students based on their overall GPA and core GPA, which was made up of Economics, Communications, Accounting and Business Math. Such characteristics included age, gender, previous education level, previous Math grades and English 12 marks.

The student data gathered from Banner was organized, cleaned and formulated into spreadsheets to meet the requirements of *Scenario*. Cleaning of the existing data consisted of checking for missing information, removal of unnecessary information and redundancies. It was important to correct and remove any irrelevant information to ensure that *Scenario* was able to configure valid possible factors or outcomes. When inputting information into the spreadsheets, coding of information was included to help enhance the data mining process.

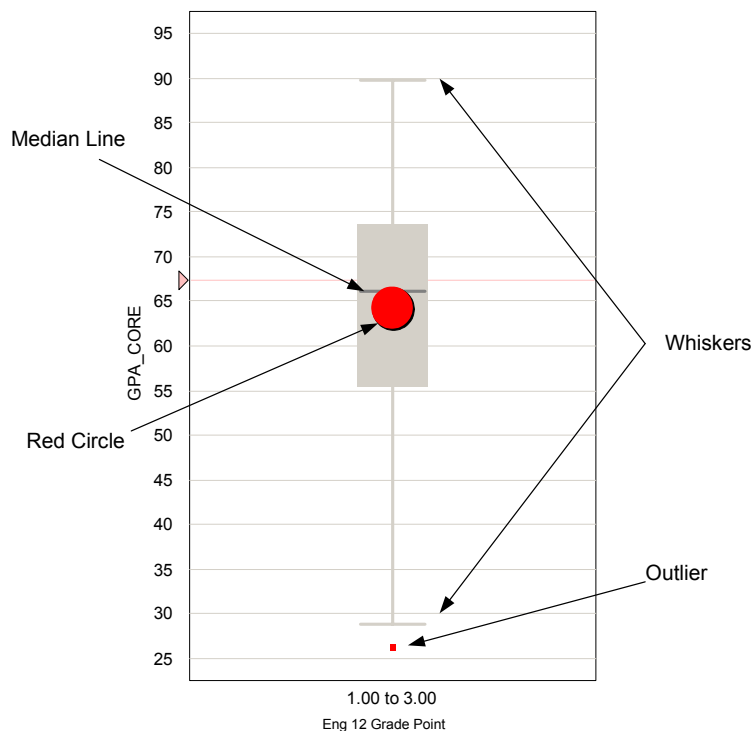
Understanding Data Mining Software

Scenario explores data looking for factors or segments that explain any characteristics that are associated with changes in the target. A target variable is selected at the beginning of the analysis process from within the data set. *Scenario* then determines what characteristics or factors are associated with the target, using selected variables chosen from the overall dataset. This allows for an analysis of particular areas of interest or concern. Once the target and the factors have been selected, *Scenario* tests the data to determine what impact or affect the factors have on the target. When it is determined that factors are available, *Scenario* provides an analysis. This analysis includes the importance of the factors and their impact on the target. Further, *Scenario* produces a rank order of influence that each variable has on the target. The program then continues to mine the factors to determine how the various factors may impact on each other. The end result is a chain effect that allows the user to track how the various influences work back to the target via the other factors. *Scenario* provides both a graphic and statistical summary of the results of the data mining, allowing for concise and simple interpretation.

When viewing the graph view or tree view, it is important to understand the various characteristics the graphic representation of the data mining results. Figure 4.1 illustrates the common features of each node – or subset – found in the tree diagrams. Each node contains (as labelled) a red circle, median line, shaded box, “whiskers”, and outliers. The red circle in each segment indicates the relative size of the segment and the vertical position of the circle indicates the average target value. The median line, located in the shaded box, determines if the data in the segment is normal or skewed. If the median line

appears in the center of the box, the data is normally distributed around the mean, but if the line appears above or below the center, then the data is skewed. The shaded box represents the data variation for each segment. The reliability of the segment in relation to the factor is determined by the size and overlap of the box relative to other segments. If the boxes do not overlap then the segments are said to be strongly related to changes in the target. The whiskers show the largest and smallest values in each segment that do not appear unusual. The maximum distance of the whiskers is 1.5 times the height of the shaded box. If a value or record is outside these whiskers, they are considered outliers. Outliers consist of records that are unusually high or low values.

Figure 4.1: A Sample Node from Scenario



When the information provided in the segments is qualitative (recorded as text) the data will be represented in a bar graph format. The bar graph consists of a legend, indicating which colour corresponds with each bar. The segments are divided into a set of graphs to provide an overview of all the relative segments in the relation to the target. The segment headings are provided along the bottom of the bar graphs.

Scenario and BCIT

It was decided that the areas of focus for analyzing student success would be on their first term overall GPA and core GPA. The overall GPA consisted of the student's marks from all courses taken during the first term of their program. The core GPA was calculated to establish the student's GPA from Economics, Communications, Accounting and Business Math, as it was determined that these courses were relevant across all programs and were considered to be the most difficult during first term.

Characteristics of Students

The variables or characteristics used to predict student success are discussed below to provide an understanding of why variable was selected, as this is an integral aspect in providing meaningful information when using the data-mining tool.

Higher Math was a dummy variable that captured those students who had math greater than Math 11. A 0 indicated that the student did not have math higher than grade 11, whereas a 1 indicated the student had a higher level math course.

Math Grades and **English 12** were coded using a numeric 4.0 point scale. The tables below provide a reference to the percentage or letter grade associated with the 4.0 point scale.

Figure 4.2: 4.0 Point Scale

Percentage	4.0 Scale	Letter Grade	4.0 Scale
0.95	4.33	A+	4.33
0.90	4.00	A	4.00
0.85	3.67	A-	3.67
0.80	3.33	B+	3.33
0.75	3.00	B	3.00
0.70	2.67	B-	2.67
0.65	2.33	C+	2.33
0.60	2.00	C	2.00
0.55	1.67	C-	1.67
0.50	1.00	D	1.00
0.00	0.00	F	0.00

Location was a coded variable that used a numeric value to correspond to the different regions. The table below provides the number associated with the region.

Figure 4.3: Regions by Code

Value	Region
1	Van/Burn
2	Fraser Valley
3	Coquitlam
4	Richmond/Delta
5	Other region in BC
6	West/North Vancouver
7	Other Province
8	Other Country

Related Job Experience was a dummy variable that captured students who had work experience that was considered relevant to their program. This information was taken

from the Entrance Student Survey. A 1 indicated the student had no related work experience, whereas a 2 indicated that the student had relevant work experience.

Gender was a coded variable that was taken from the Banner Database. Students coded as 1 were female, and students coded as 2 were male.

Objective for Attending BCIT was a coded variable that used a numeric value to correspond to the different reasons as to why students attend BCIT. This data was collected from the Entrance Student Survey. The table below provides the numeric value that relates to the student's objective.

Figure 4.4: Student Objectives for Attending BCIT

Value	Objective
1	Get a job
2	Establish a career
3	Change/get better job or career
4	Non work related reasons

Age Group was a coded variable that distinguished the age group that the student fell into. The age of the student was calculated from the Banner Database. The following table provides the numeric value that corresponds with the specified age group.

Figure 4.5: Age Group Categories

Value	Age Group
1	19 or younger
2	20-24 years
3	25-29 years
4	30 or older

Previous Education Level was a coded variable that used to determine the level of education a student had prior to attending BCIT. The previous education level of students was gathered from the Entrance Student Survey. The table below provides the value that relates to the student's education before attending BCIT.

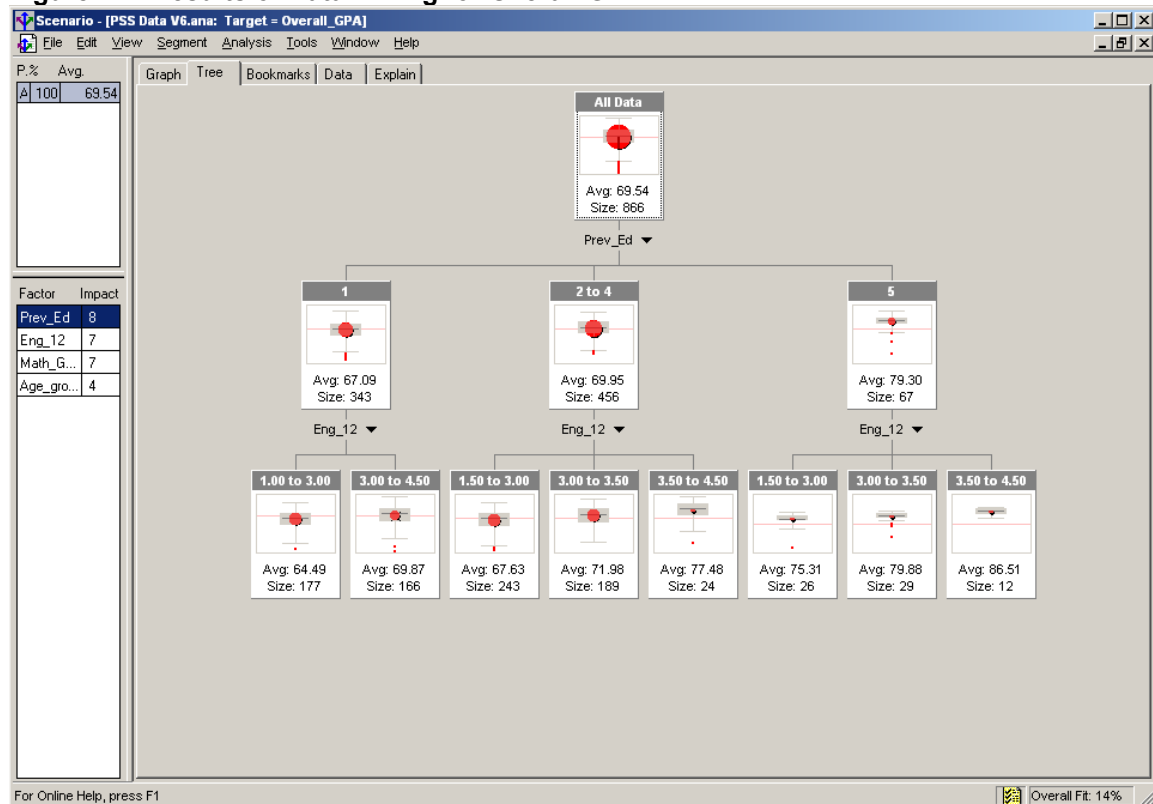
Figure 4.6: Previous Education Level Categories

Value	Education Level
1	Grade 12 or less
2	Some college or university
3	Vocational
4	College level certificate or diploma
5	University degree

Factors that Influence Student's Overall GPA

The first target variable analyzed was overall GPA, which is defined here as the student's GPA in their first term of BCIT studies. Results of this data mining analysis are shown in Figure 4.7. The average overall GPA for the School of Business was 69.5%, which is based on a sample of 866 student records, deemed valid by the data mining process. The first factor to be analyzed is previous education of the students. The data was then drilled down producing three segments, which are the natural resulting clusters of the data based on common characteristics derived by the Cognos software.

Figure 4.7: Results of Data Mining for Overall GPA



Segment one was based on students having a previous education before attending BCIT of Grade 12 or lower. The average overall GPA for this segment was 67% and contained 40% of the sample. There are four students whose average overall GPA were between 27% to 36%, which was outside the overall GPA range of 38% to 93%, which the software subsequently considered to be outliers.

Fifty three percent of the School of Business students were contained in segment two. This group was defined by their previous education levels, which were as follows:

- Some college or university
- Vocational
- College level certificate or diploma.

The average overall GPA in for this segment was nearly 70%. The range of overall GPA's for segment two was from 40% to 92%. There were two students in this segment who had overall GPA's of 35% and 38%, which was outside the allowable range.

The smallest percentage of students were found in the third segment, for a total of eight percent. Segment three was classified as students having a previous education that consisted of a university degree. The average overall GPA in this segment was 79%. Three students in this segment had overall GPA's of 35%, 51% and 63%, which was outside the "allowable" overall GPA range of 64% to 93%.

Each of the three segments were then drilled down to provide further analysis. For each of the three segments, it was determined that student's English 12 grades had the second largest impact, following previous education, on their overall GPA than any of the other explanatory variables that were included in the dataset.

It was revealed that students in the School of Business with a previous education of Grade 12 or lower, that their performance was enhanced by their English 12 grades. The English 12 grades were split into two segments. The first segment contained students whose English 12 GPA scores ranged from 1.00 to 3.00. Segment one had an average overall GPA of 64.5%, and contained of 20% of the sample. The second segment contained students whose English 12 GPA scores ranged from 3.00 to 4.50. The average overall GPA for this sample was nearly 70% and contained 19% of the students.

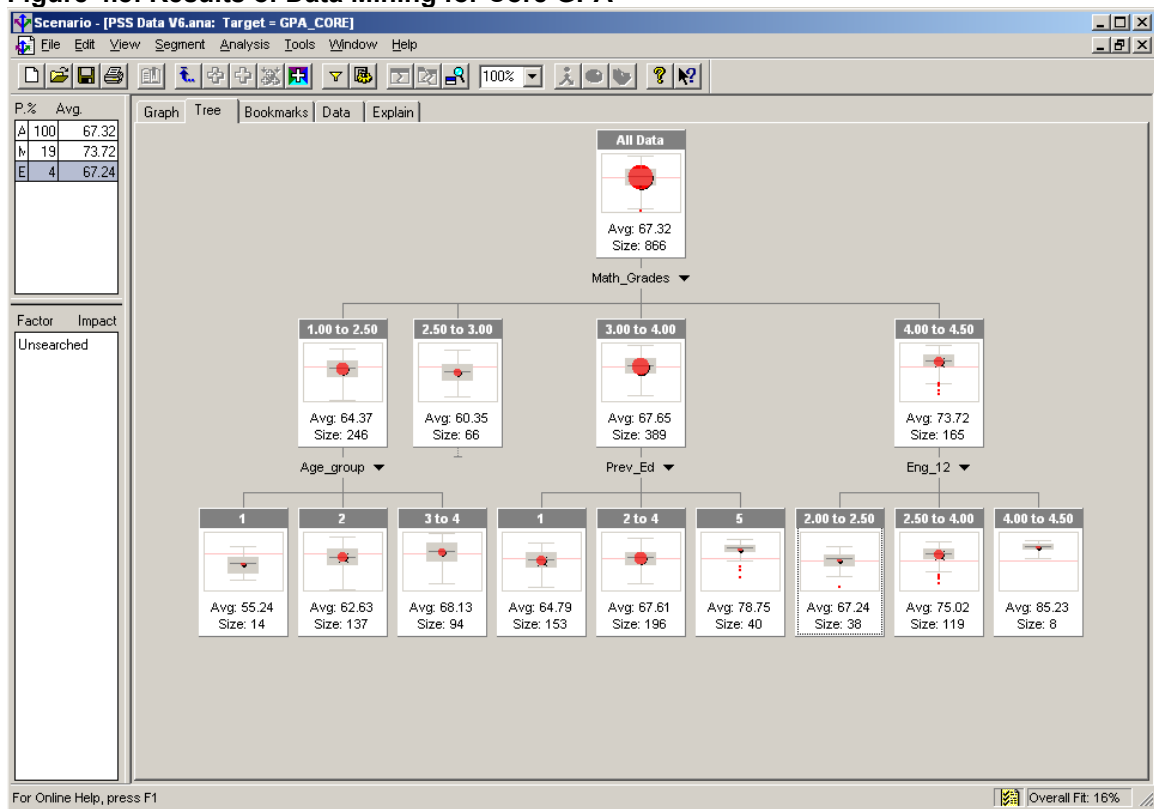
Students who had a previous education of some college or university, vocational or college level certificate or diploma were separated into three segments based on their English 12 grades. The English 12 scores for the first segment ranged from 1.50 to 3.00. The largest proportion of students were found in this segment, which contained 28% of the sample. Students contained in this segment had an average overall GPA of nearly 68%. The second segment contained students with English 12 grades of 3.00 to 3.50. The average overall GPA for segment two was 72% and contained 22% of the student sample. The third segment contained students with English 12 grades of 3.50 to 4.50. The average overall GPA for segment three was 77.5%, which was made up of the lowest sample of students, for a total of three percent.

Students who had a university degree were grouped into three segments based on their English 12 grades. The first segment was made up of students with English 12 grades of 1.50 to 3.00. The students in this segment had an average overall GPA of 75% and contained three percent of the sample. Segment two was based on students having an English 12 score of 3.00 to 3.50. The average overall GPA in segment two was nearly 80%, and consisted of about three percent of the students. The third segment had the lowest percentage of students, at about one and a half percent. Segment three contained students with English 12 scores ranging from 3.50 to 4.50, with the students found in this segment having an average overall GPA of 86.5%.

Factors that Influence Student’s Core GPA

The second target variable analyzed was the School of Business students core GPA. The core GPA was calculated using the students’ percentage grades for first term in Economics, Accounting, Business Math and Business Communications. The results of this analysis are shown in Figure 4.8. The average core GPA was 67%, which is based on a sample of 866 student records that the data mining processes deemed as valid. The first factor to be analyzed is students Math Grades before attending BCIT, which the data mining process determined as having the highest impact on core GPA. The data was then drilled down, producing four segments. These four segments are the results of natural clustering found within the data that share common characteristics derived by the Cognos software.

Figure 4.8: Results of Data Mining for Core GPA



The first segment was based on students having previous Math Grades of 1.00 to 2.50. The average core GPA for segment one was 64%, with 28% of the sample being found in this segment. The range of the core GPA was from 26% to 90%, with all students this segment falling between the specified range.

The smallest percentage of students was located in segment two, for a total of eight percent, and had an average core GPA of 60%. The Math Grades for this segment were from 2.50 to 3.00. The range of the core GPA for this segment was from 30% to 87%, with all students’ scores being within the range.

45% of the students were located in the third segment, which contained the largest proportion of the sample. These students had an average core GPA of nearly 68%. Segment three was contained students having previous Math Grades from 3.00 to 4.00. The range for segment three was 33% to 93%, and all students in this segment fell within the designated range.

Segment four was based on students having previous Math Grades of 4.00 to 4.50. 19% of the students were located in this segment and had an average core GPA of nearly 74%. There were six students having core GPA scores ranging from 33% to 46%, which was outside the segment's "allowable" range of 47% to 94%.

The four segments of the Math Grades were then drilled down to provide deeper analysis. Each of the four segments yielded different results, ranging from no further analysis to various other explanatory variables, such as age, previous education and English 12 grades.

Students, who had previous Math Grades ranging from 1.00 to 2.50, were assessed to have Age Group influence their performance. Based on this Math Grade variable, Age Group was split into three segments. The first segment was based on students being 19 years of age or younger. Two percent of the students were located in this segment and they had an average core GPA of 55%. Segment two consisted of students in the Age Group of 20 to 24 years of age, and contained of 16% of the sample. The average core GPA for the second segment was nearly 63%. The third segment contained students in the Age Groups 25 to 29 years and 30 years and older, these had an average core GPA of 68%. Eleven percent of the sample were located in segment three.

The data mining process determined that students who had previous Math Grades ranging from 2.50 to 3.00 did not have any further factors that influenced their performance.

It was determined that students with Math Grades of 3.00 to 4.00 were further influenced by previous education. The data mining process split previous education of students into three segments. The first segment contained students having a previous education of grade 12 or lower. The average core GPA for segment one was nearly 65%, with 18% of the sample being located in this segment. Students who had some college or university, vocational training or a college level certificate or diploma as their previous education were grouped together in the second segment. 23% of the students were located in segment two, with these students had an average core GPA of nearly 68%. Segment three contained students who had a previous education that consisted of a university degree. These students had an average core GPA of nearly 79%, with five percent of the sample located in the third segment.

Students with Math Grades of 4.00 to 4.50 were concluded to have been further influenced by their English 12 grades. The data mining process split English 12 grades of the students into three segments. The first segment consisted of students having English 12 scores of 2.00 to 2.50. Four percent of the sample was located in this segment, with

these students having an average core GPA of 67%. The second segment consisted students having English 12 scores of 2.50 to 4.00, with 14% of students being located in this segment. The average core GPA for students in segment two was 75%. Segment three contained students having English 12 grades of 4.00 to 4.50. The smallest sample of students was located in the third segment for a total of one percent, with students having an average core GPA of 85%.

Conclusion

From the analysis of the BCIT student data, several conclusions can be made with respect to determining student success at the Institute. Due to the structure of the report, there seemed to be two natural divisions into which the conclusions could be grouped. The first section provides conclusions relating to general characteristics of students within the School of Business, whereas the second section contains conclusions relating to the explanatory statistical analysis of student success within the school.

Based on the analysis of a more general basic nature, the following conclusions were inferred regarding students within the School of Business and its programs over a four-year period:

- 85% of students cite their objective for attending BCIT's School of Business as being career or job driven, whereas the remaining 15% state their objective for attending the school as non-work related.
- 64% of School of Business students have some level of post secondary education or work related experience, whereas the remaining 36% of students are admitted directly from high school.
- Gender distribution is relatively equal in most programs (50.1% female, 49.9% male). Some programs indicate a self selection bias, such as Human Resources Management where there are more females compared to males, whereas International Trade and Transportation and Operations Management have more males as opposed to females.
- The largest cohort of students attending the School of Business are between the ages of 20-24 years of age, whereas the smallest cohort of students are 19 years or younger.
- The data analysis inferred a correlation between the age and average GPA of all School of Business students during their first term. Students who were 19 years or younger had an average GPA of 65%, whereas student over 30 years had an average GPA of 74%; thus indicating a 9% difference between these age groups.
- School of Business has an average overall GPA of 70%, which is reflected in most programs. Human Resources Management scored the highest at 78%, whereas International Trade and Transportation had the lowest GPA at 64%.
- Regardless of gender, males and females perform relatively equal with respect to average GPA within their first term at BCIT (69.0% and 70.3%, respectively).

- Microeconomics was indicated as having the highest failure rate, at 18%, for School of Business first term students. Business Math had the second highest failure rate of first term students at 14%.
- The success rate, defined as having no grade lower than 50%, of students within the School of Business with respect to the core courses was 75%. Human Resources Management had the highest success rate at 91%, whereas International Trade and Transportation and Management Systems were below 70% (62% and 64%, respectively).
- The data analysis inferred that younger students, 19 years or younger, failed more core courses when compared with the other age groups. Students who were 30 years or older had the lowest failure rate of core courses.
- Gender does not provide any noticeable difference with respect to the failure of first term core courses.

The data mining exercise yielded some interesting statistical analysis results with respect to the School of Business, as a whole, and each of the programs within the school. These conclusions may support the fact that there are explanatory characteristics or variables that aid in determining student success. The conclusions relating to the statistical analysis are discussed below:

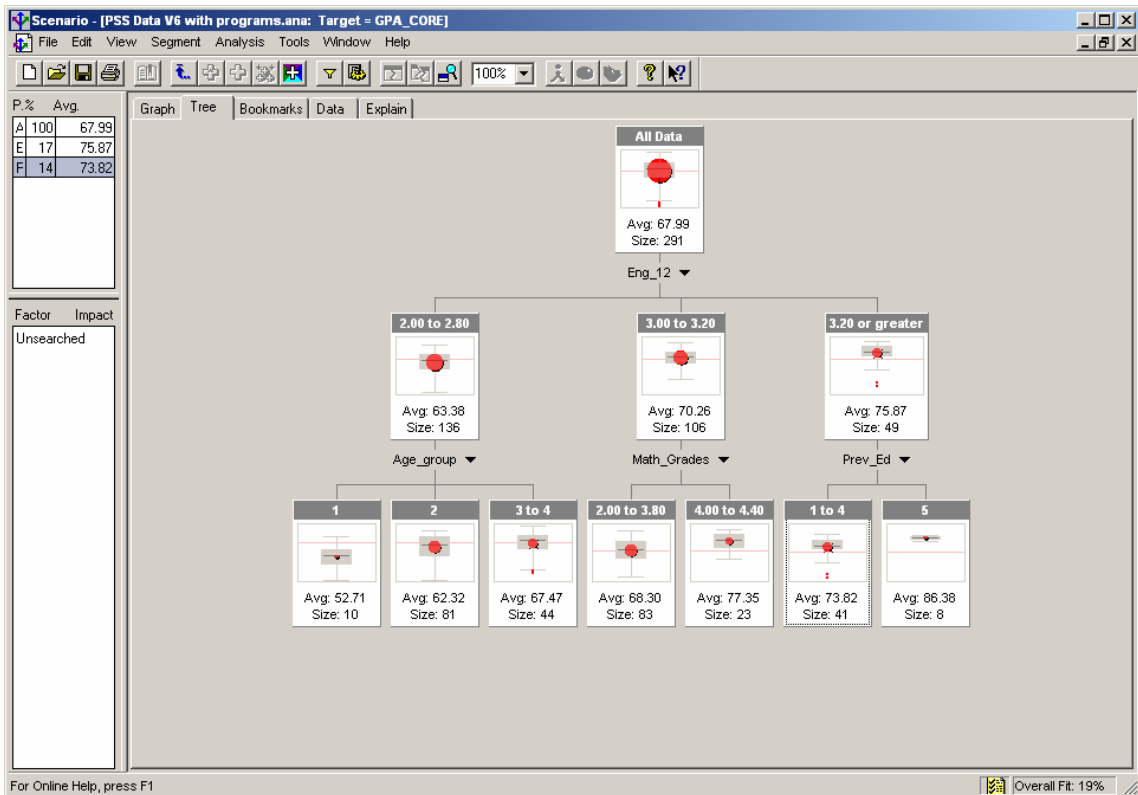
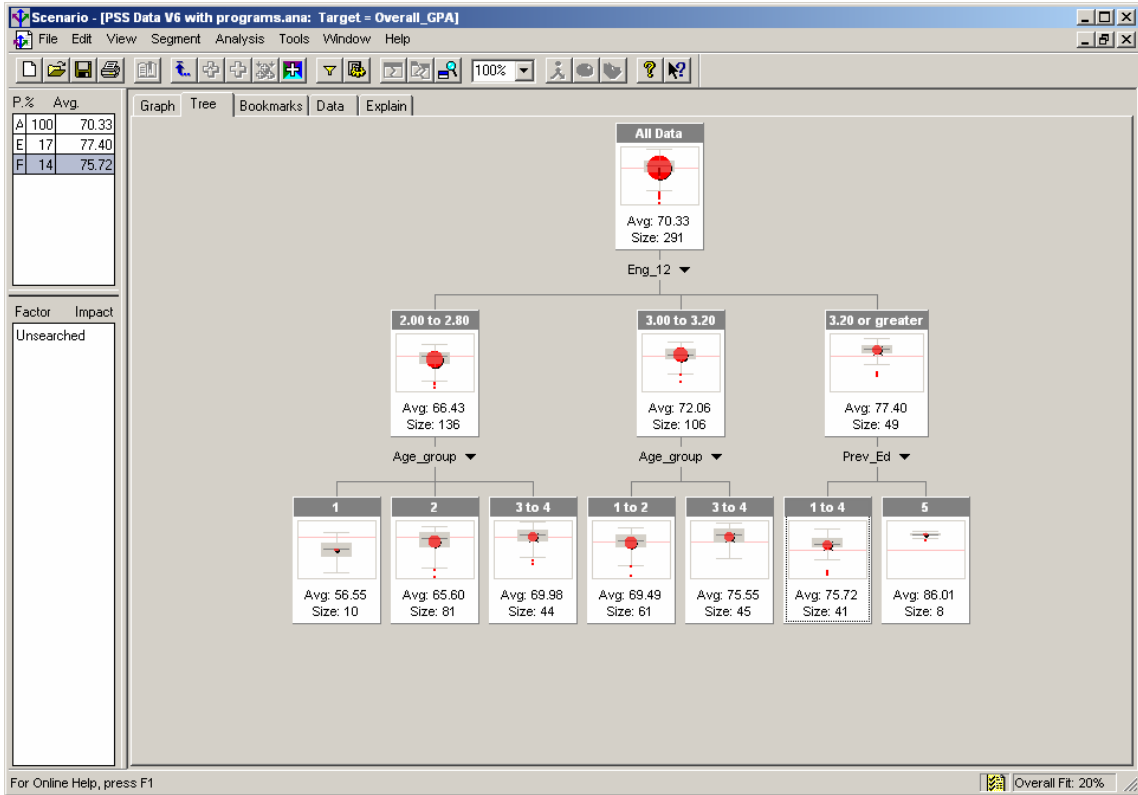
- When analyzing factors related to School of Business students' overall GPA, previous education of these students was indicated as having the highest impact on their performance. It was noted that the higher the level of previous education the student had, the higher the average overall GPA they achieved. The data mining inferred a 12% difference in average overall GPA of students who have an education of Grade 12 or lower (67.0%), when compared with students who had a university degree (79.3%). The second most distinguishing factor for determining the GPA was English 12 scores. The data mining suggested a strong correlation between scores in English 12 and average overall GPA.
- When assessing the success of School of Business students in their core courses, their previous math grades were determined as having the most significant impact on their average core GPA. The data mining indicated a correlation between a higher math grade and higher core GPA. There was a 9% difference between students having math scores of 1.0 to 2.5 (64.3%), when compared with students having math scores of 4.0 to 4.5 (73.7%). With further analysis, it was determined that depending on the students' math grades, there were three factors (age group, previous education and English 12) that were related to their success. It was noted that as students increase in age, their average core GPA also rises. As stated earlier, the higher the level of previous education, the greater the student scored with respect to their average core GPA. The same also applies for English 12 scores, the higher the student scored in English 12, the higher the average core GPA the student achieved.

- Most programs, when analyzed individually, yielded results similar to the School of Business. The International Trade and Technology program indicated that the age groups of its students were the most distinguishing characteristic. Students in this program who were 19 years and younger and 20-24 years had an average overall GPA of 58.6%, whereas students who were 25-29 years and 30 years and older had an average overall GPA of 67.9%, indicating a 9% difference in GPA amongst these two segments. Operations Management yielded that related job experience had the most significant impact on its students GPA. Students in this program who had relevant work experience scored 10% higher in their average overall GPA, when compared to students who did not have any work experience.

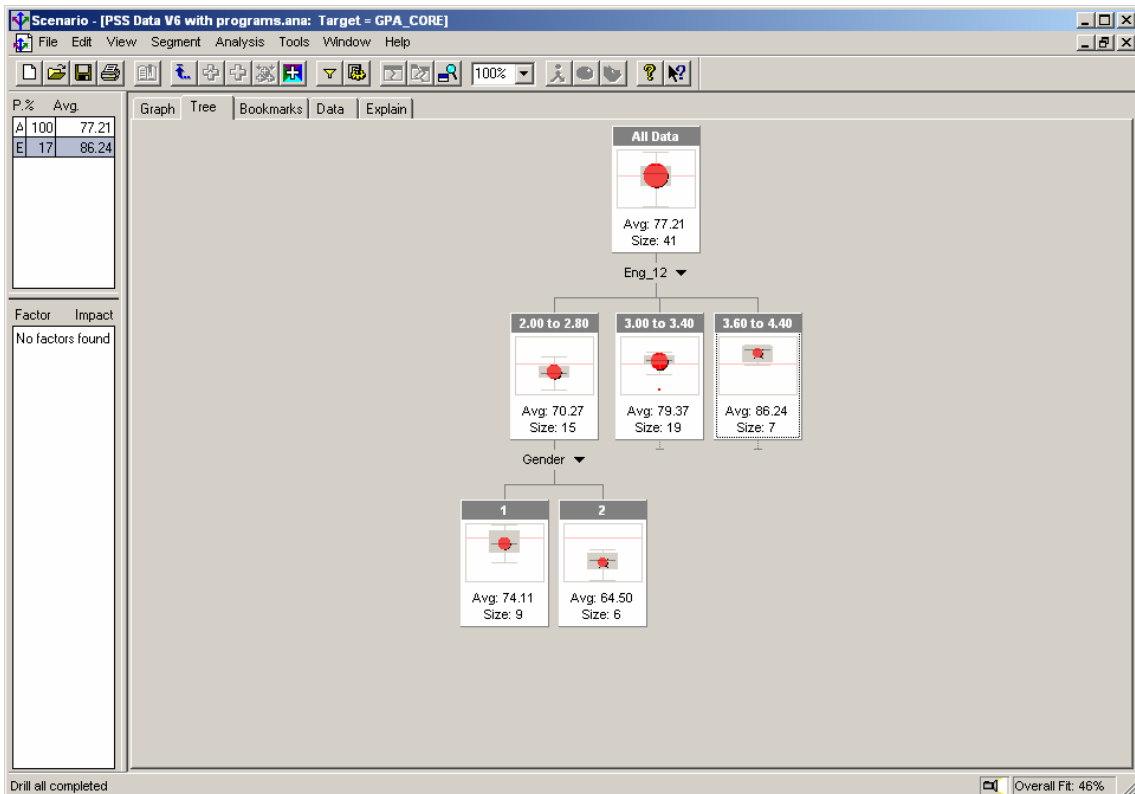
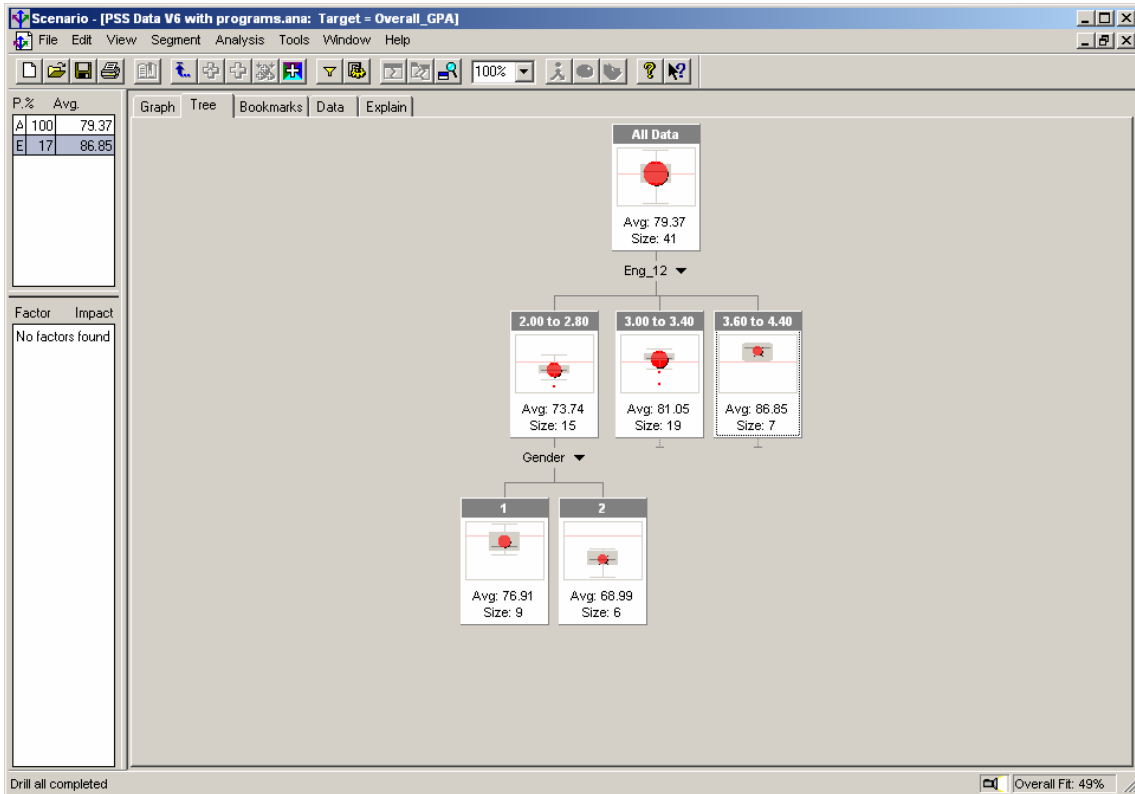
Appendices

Appendix1: Program Area Specific Results

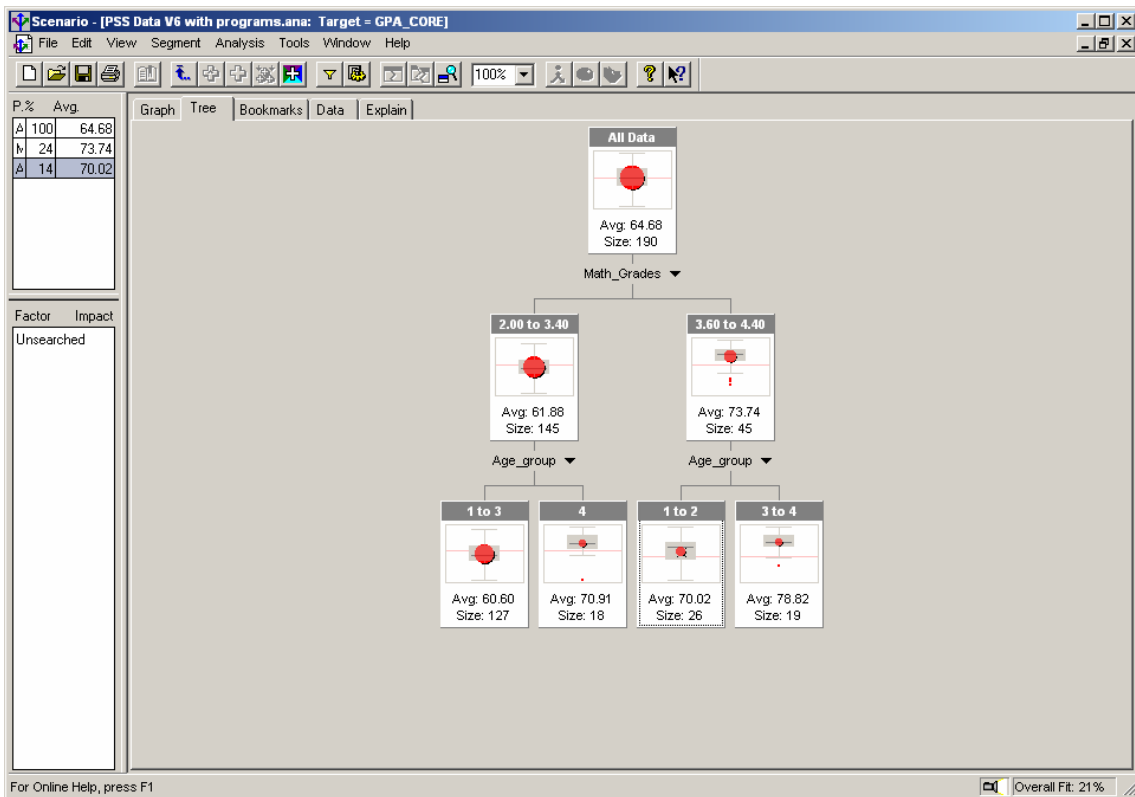
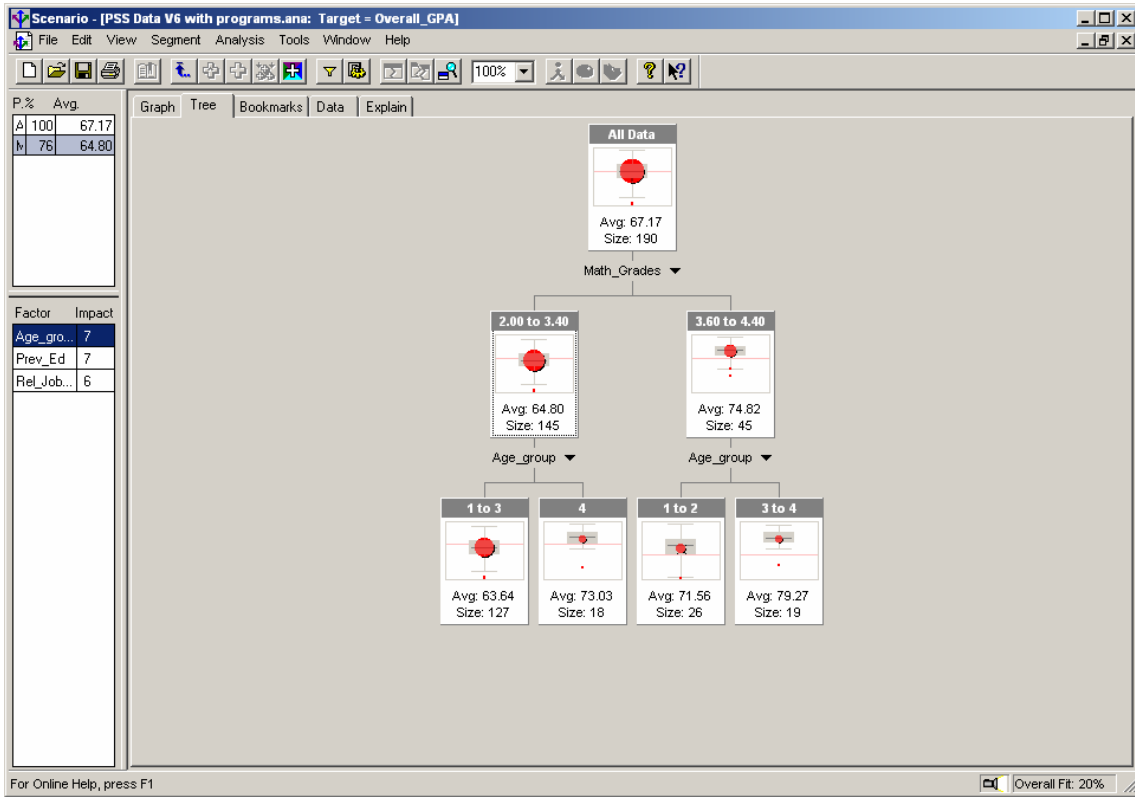
1. Financial Management (Overall & Core GPA)



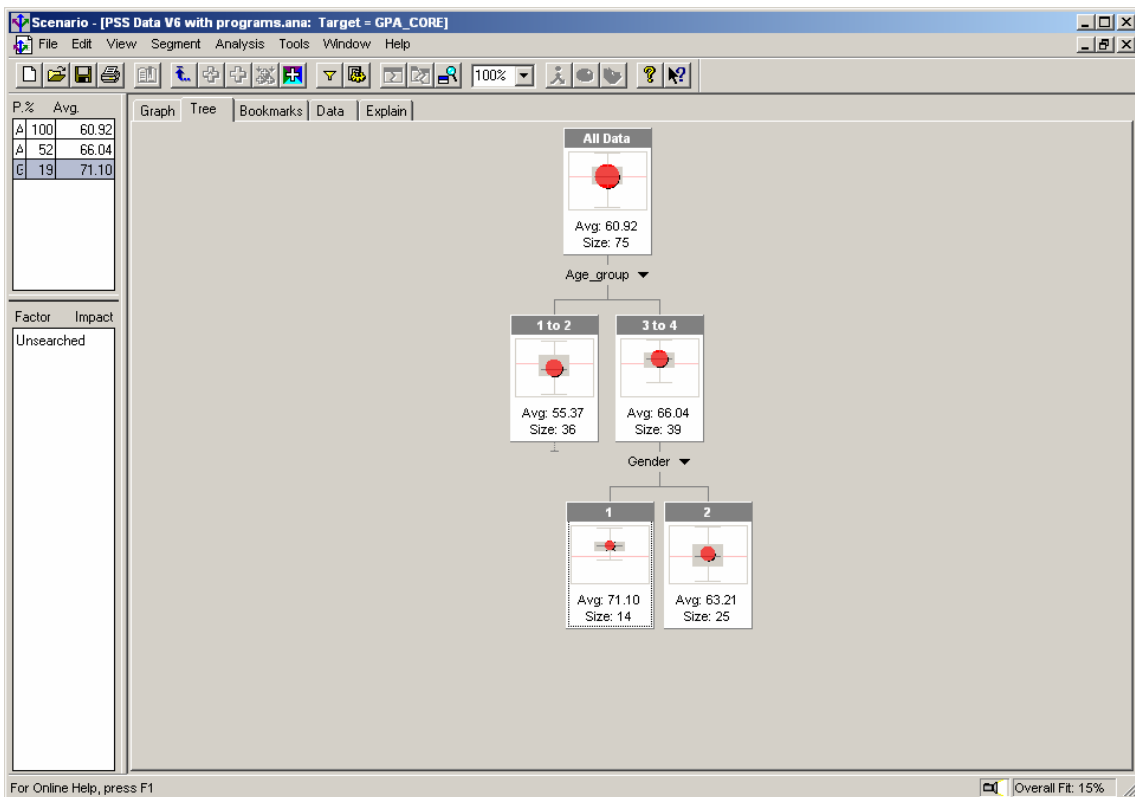
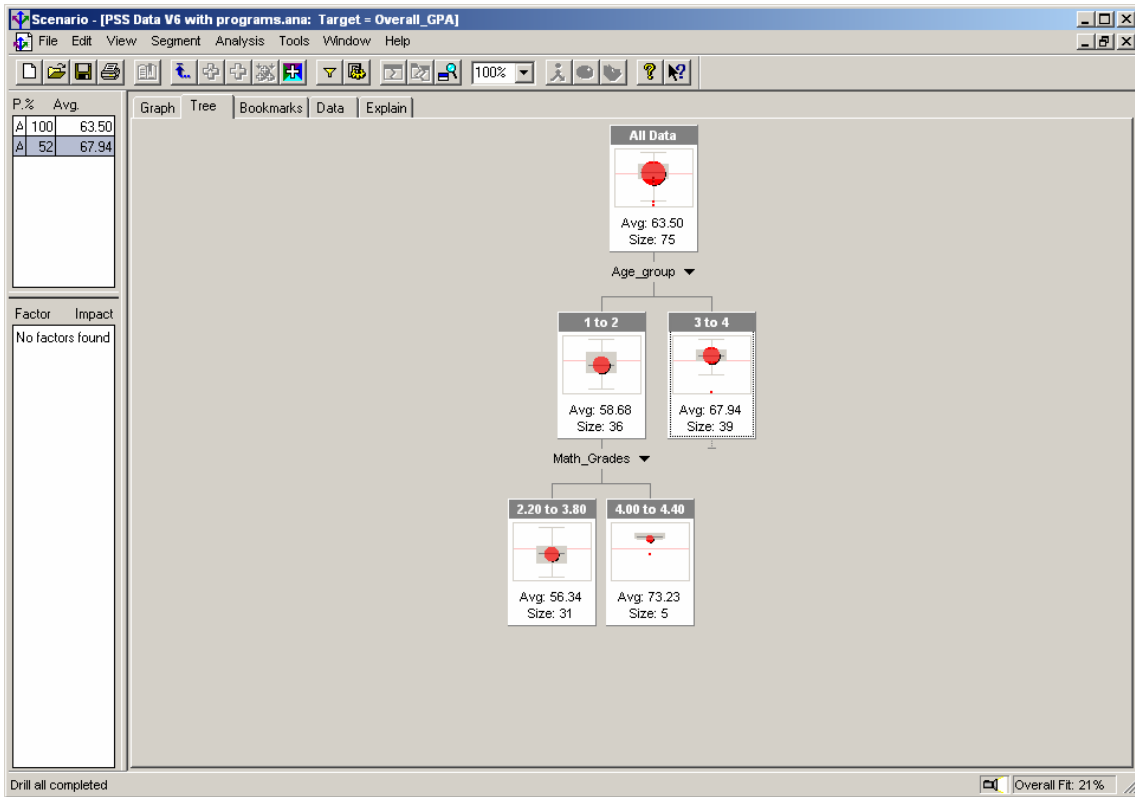
2. Human Resources Management (Overall & Core GPA)



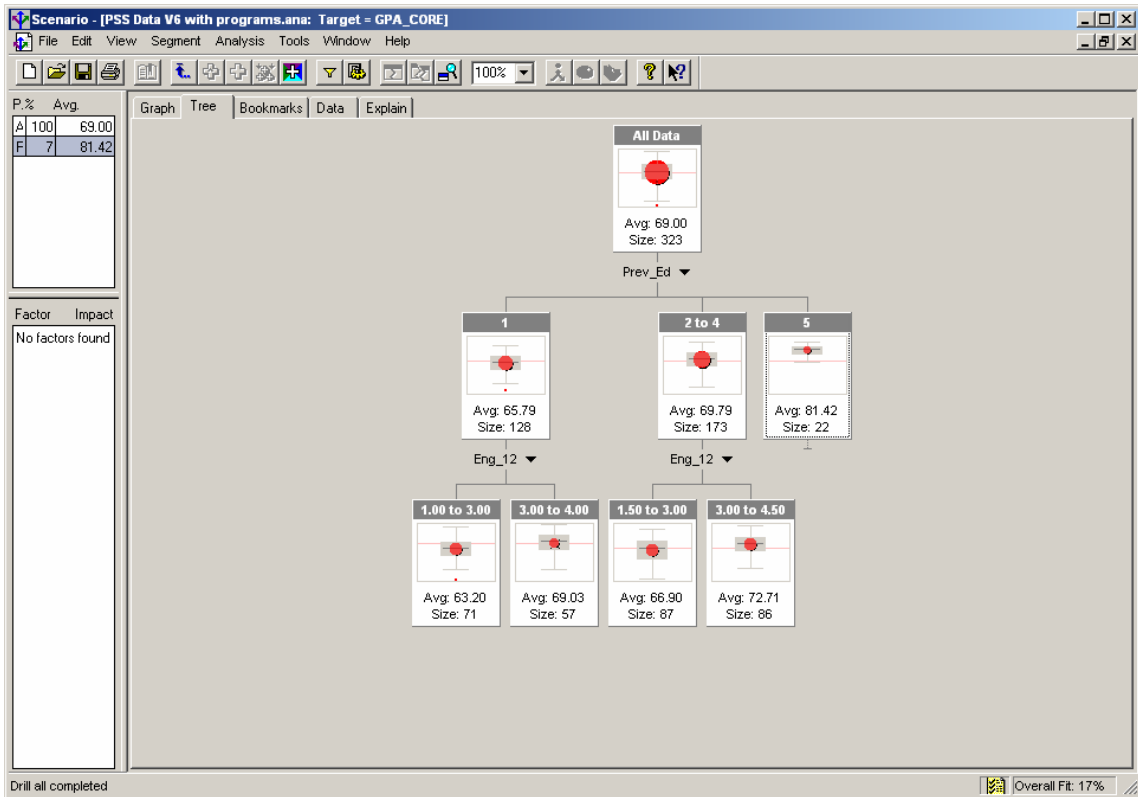
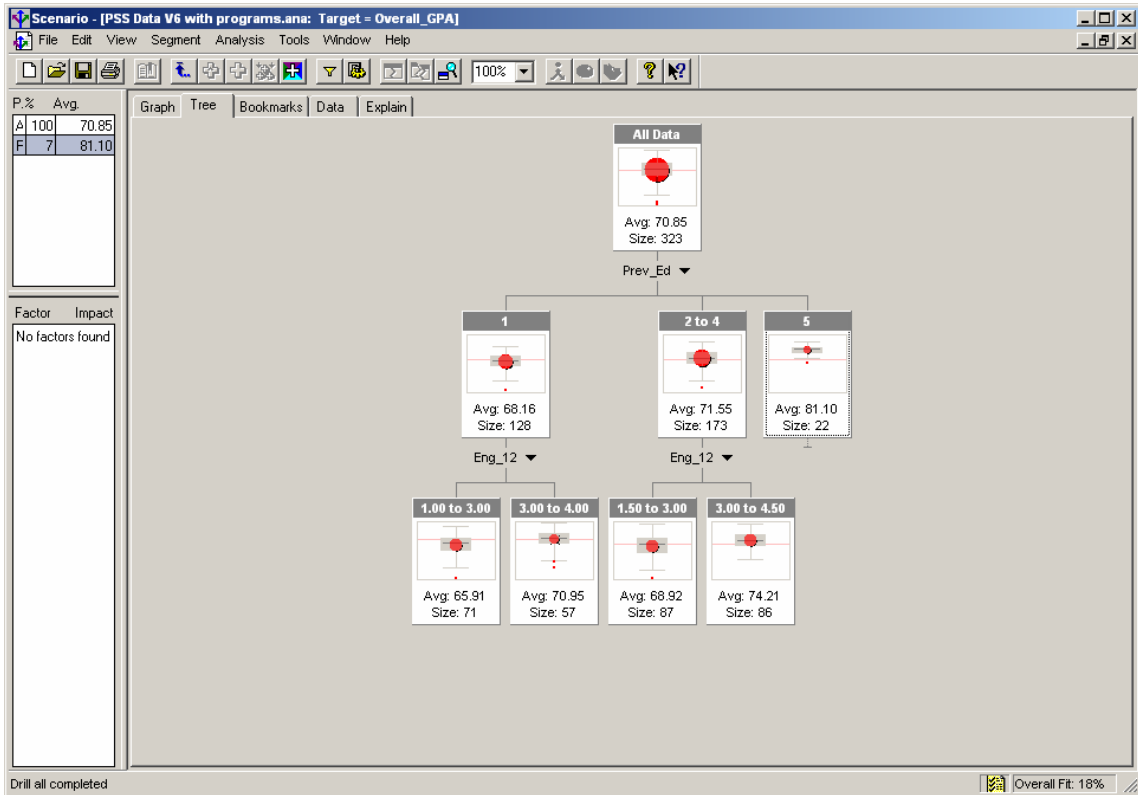
3. Integrated Management/ Information Technology/ Management Systems (Overall & Core GPA)



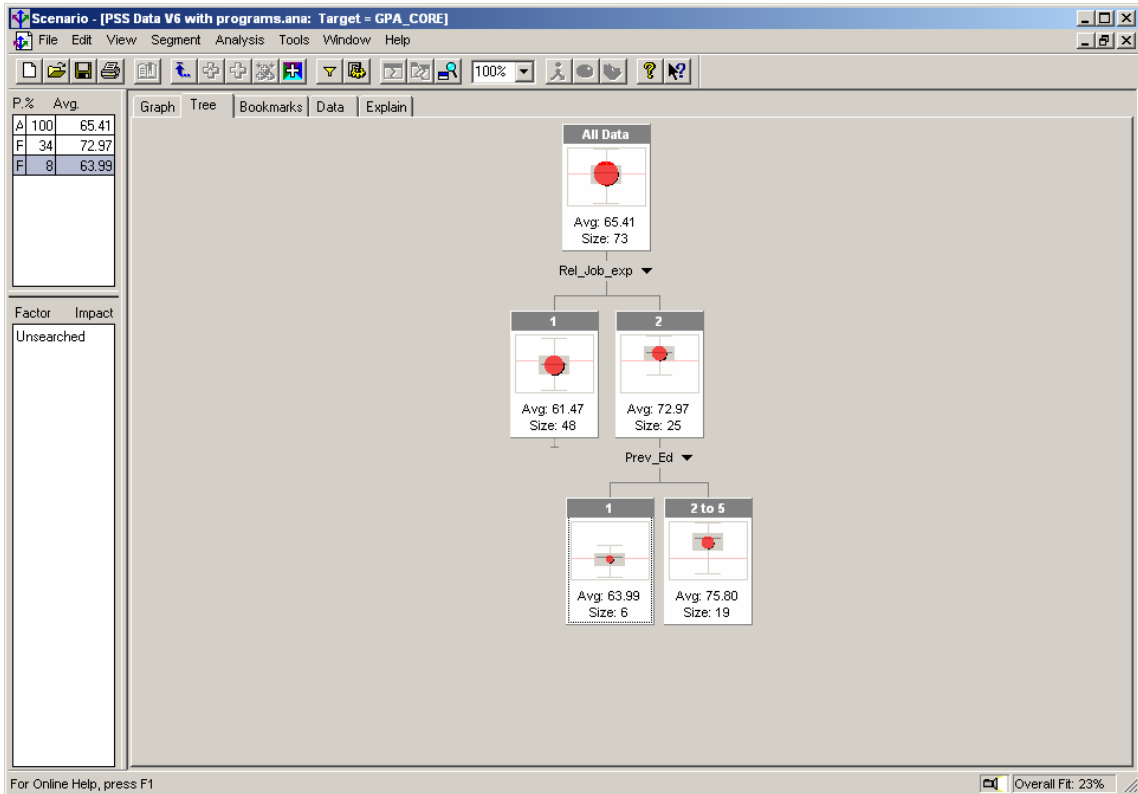
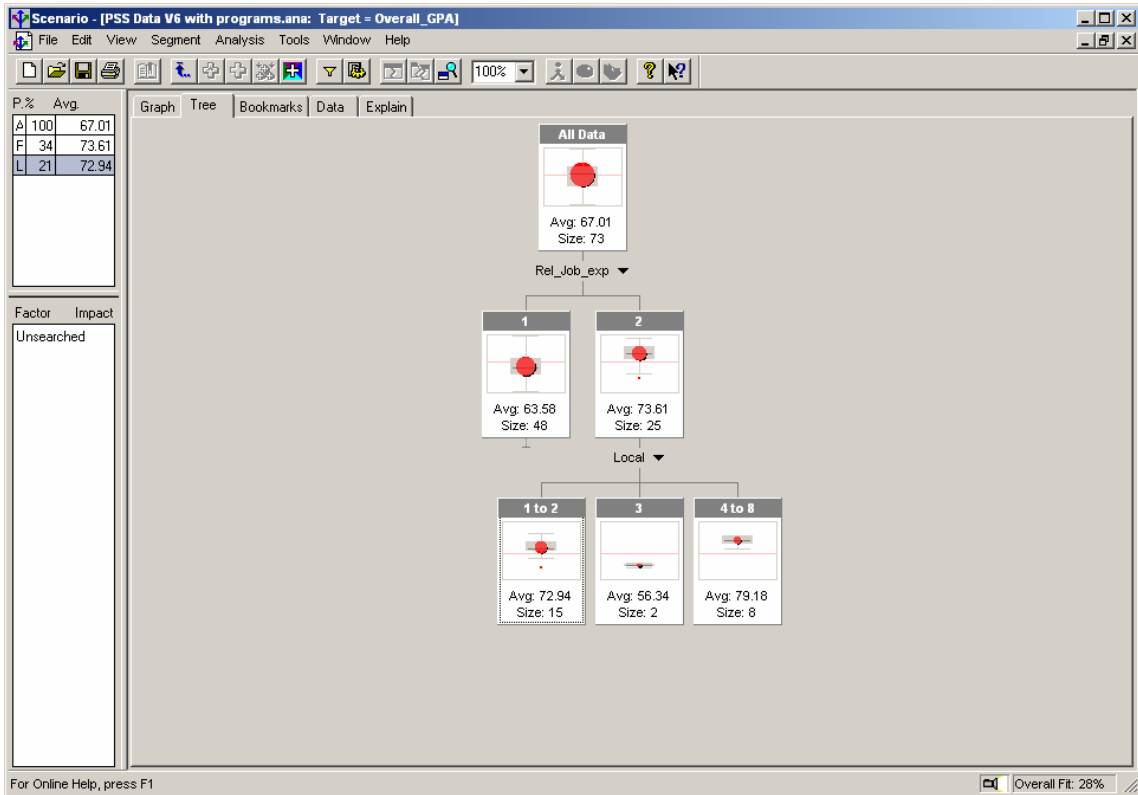
4. International Trade and Technology (Overall & Core GPA)



5. Marketing Management (Overall & Core GPA)



6. Operations Management (Overall & Core GPA)



Appendix 2: Recoding of BCIT's First Term GPA

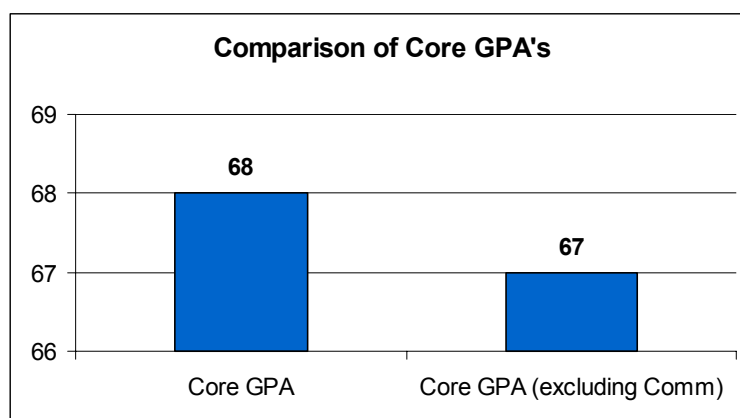
Based on the nature of the project for predicting student success, it was discussed and agreed upon that a new GPA be evaluated to provide a more realistic measure of students course marks. While students in BCIT's School of Business take the same core courses, it was felt that four of the seven courses would provide a more accurate measure of a student's performance due to the content and complexity of the courses. The courses selected for determining the new GPA(1) for this project were:

- COMM 1100
- ECON 2100
- FMGT 1100
- OPMT 1100.

After the calculation of the recoded GPA, it was decided that another GPA(2) score should be determined based on:

- ECON 2100
- FMGT 1100
- OPMT 1100

The reason for including this GPA(2) score was to see if any profound differences occurred between the GPAs. It was later determined that the exclusion of the COMM 1100 mark caused only a slight variation between the student's recoded GPA(1) and GPA(2). A minimal percentage increase, of one to four percent when viewed by program, was noticeable in the student's GPA(2) score compared to the previous recoded GPA (1). The graph below indicates a one percent difference for the entire School of Business when comparing the two scores.



The steps for establishing the new GPA (1) and GPA (2) will be described sequentially below.

Student information was pulled from the Banner Database for the years 1998, 2000, 2001 and 2002. The information was compiled into a large spreadsheet to provide easy access to all information. (As described in the Recoding of Provincial grades.)

Based on the information provided in the spreadsheet from Banner, the necessary information was extracted and placed into a smaller spreadsheet to allow for ease of use for developing calculations, or formulas, to determine the new GPA (1) and GPA (2). The information extracted from the larger spreadsheet was:

- Student Identification Number (SID)
- Program
- Status Code
- COMM 1100
- ECON 2100
- FMGT 1100
- OPMT 1100.

Before a formula could be used to calculate the new GPA (1) and GPA (2), all alpha information contained in the cells had to be converted into a numeric value. The reason this had to be done was that if the formula encountered alpha information it would return a reference error for that student. Some of the alpha information contained in the cells relating to the students courses were withdrawal (W), late withdrawal (LW), earned credit (ECR) and V. The values selected to replace the withdrawals and (quits??) were entered as negatives, so this would have no effect on the formula calculation. The only values the formula would recognize would be positive, creating accuracy for the final GPA (1) and GPA (2) final scores. It was also determined that the students who earned a credit for a course would receive a percentage value of 75%. This was based on the assumption that it would be an average mark students would receive for taking the course prior to the semester. The table below was extracted from the spreadsheet used to calculate the new GPAs to show the values applied to the alpha information:

Alpha Information	Replaced Value
W	-1
V	-2
LW	-3
ECR	75%

The step required before the building of the formula, was determining the credits a student would receive for completing the courses. The credits for each course are the same, regardless of the program a student is taking within the School of Business. The course credit breakdown was:

Course Name	Total Course Credit
COMM 1100	3
ECON 2100	3
FMGT 1100	4
OPMT 1100	4

Formulas were then build that could be applied to each SID to return a new value for the GPA (1) and GPA (2) and the total number of credits the student received for each the calculated GPAs.

The formula for GPA (1) was:

$$=(\text{IF}(\text{B3}>0,\text{B3}*\text{\$B\$1,0})+(\text{IF}(\text{C3}>0,\text{C3}*\text{\$C\$1,0})+(\text{IF}(\text{D3}>0,\text{D3}*\text{\$D\$1,0})+(\text{IF}(\text{E3}>0,\text{E3}*\text{\$E\$1,0}))))/(\text{IF}(\text{B3}>0,\text{\$B\$1,0})+(\text{IF}(\text{C3}>0,\text{\$C\$1,0})+\text{IF}(\text{D3}>0,\text{\$D\$1,0})+\text{IF}(\text{E3}>0,\text{\$E\$1,0})))$$

The type of formula used was an if statement, =IF, which indicated what aspects of the cells based on the entered requirements the formula was to return. This aspect of the formula =IF(B3>0,B3*\$B\$1,0) stated that if B3 was greater than zero, then it should return that value in B3 multiplied by the number of course credits (\$B\$1), if the value was less than zero, the formula was to return a zero. This applied to all four courses selected for GPA (1), cells relating to B3, C3, D3 and E3 and the corresponding course credit amounts, B1, C1, D1, and E1. The second part of the formula was the values in which the first portion, course totals, was to be divided by credit totals. The second aspect of the formula was /(IF(B3>0,\$B\$1,0), which stated that if the value in B3 was greater than zero, than return the course credit amount, B1, but if the value was less than zero, return a 0. This complex formula in turn supplied the new GPA (1) desired for this project.

The same formula was applied to GPA (2) but did not include the course information for COMM 1100. The formula was:

$$=(\text{IF}(\text{C3}>0,\text{C3}*\text{\$C\$1,0})+(\text{IF}(\text{D3}>0,\text{D3}*\text{\$D\$1,0})+(\text{IF}(\text{E3}>0,\text{E3}*\text{\$E\$1,0}))))/(\text{IF}(\text{C3}>0,\text{\$C\$1,0})+\text{IF}(\text{D3}>0,\text{\$D\$1,0})+\text{IF}(\text{E3}>0,\text{\$E\$1,0}))$$
. The same explanation used above applies to this formula for only the three courses:

- ECON 2100
- FMGT 1100
- OPMT 1100.

The formula developed for determining the total number of credits a student received for GPA (1) was:

$$=\text{SUM}(\text{IF}(\text{B3}>0,\text{\$B\$1,0}),\text{IF}(\text{C3}>0,\text{\$C\$1,0}),\text{IF}(\text{D3}>0,\text{\$D\$1,0}),\text{IF}(\text{E3}>0,\text{\$E\$1,0}))$$
. This time the formula was to sum the all values returned based on the values supplied from if statements. The if statements stated that if the values in B3, C3, D3 or E3 were greater than 0, then return the values in B1, C1, D1 or E1, the course credits, but if the values in B3, C3, D3 or E3 were less then zero, than return a 0.

The same formula was applied to GPA (2) but did not include the course information for COMM 1100. The formula was:

`=SUM(IF(C3>0,C1,0),IF(D3>0,D1,0),IF(E3>0,E1,0))`. The same explanation used above applies to this formula but only for the three courses:

- ECON 2100
- FMGT 1100
- OPMT 1100.

GPA Distribution Overview

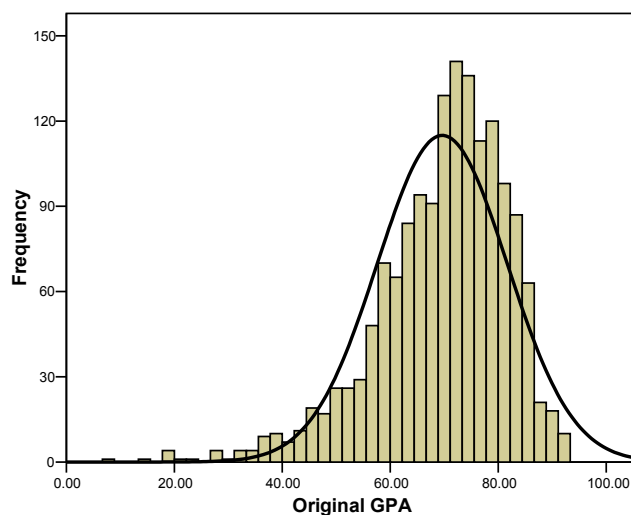
This section presents further details regarding the GPA distribution for the original GPA data and the core GPA data (R1 and R2 respectively are adjusted GPA's for core courses only with the additional exclusion of Communication respectively.) The table below shows summary statistics for these data. Interestingly, while the central tendency of each of the distributions is similar, the variances are measurably different, indicating that as more courses are excluded from the GPA calculation the variance increases.

Overall GPA Statistics

		Orig_GPA	GPA_R1	GPA_R2
N	Valid	1568	1565	1546
	Missing	39	42	61
Mean		69.6100	67.7582	67.2483
Median		71.4016	70.0714	70.0000
Std. Deviation		12.04861	13.69524	16.11957
Variance		145.169	187.560	259.840
Skewness		-1.027	-.727	-.716
Std. Error of Skewness		.062	.062	.062
Kurtosis		1.770	.351	.205
Std. Error of Kurtosis		.124	.124	.124
Minimum		7.30	12.57	8.71
Maximum		93.17	94.14	98.00

Note that negative skewness (a measure of the symmetry of the distribution) indicates a left-handed skewness and can be seen in the fact that the average for these distributions is less than the median. Ideally, skewness should be zero. The kurtosis is a measure of the “peakness” of the data and a normally distributed data-set should have a value of kurtosis of zero. The positive value does indicate some degree of “peakness” in the data.

Overall Original GPA (with est. normal curve)

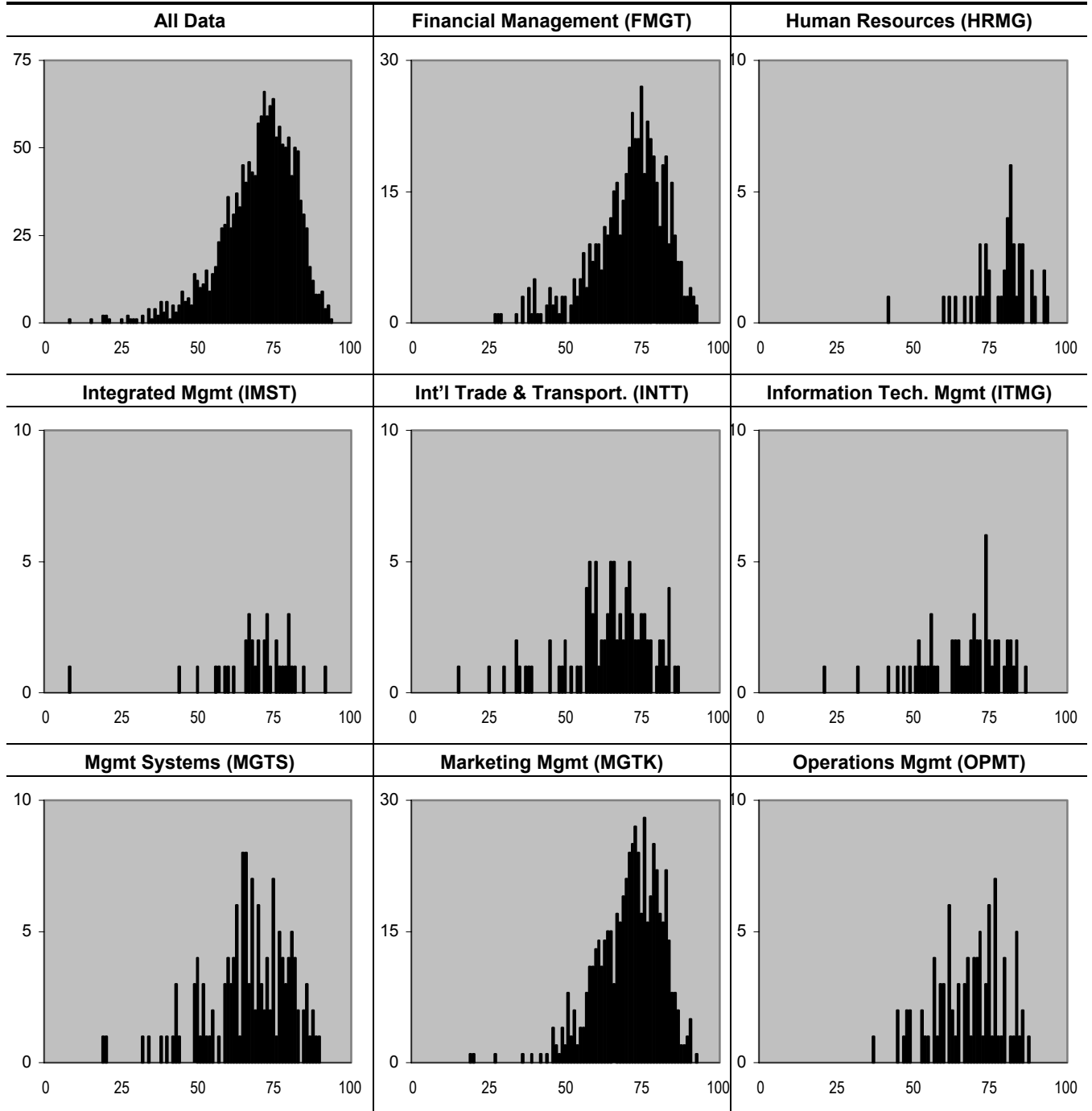


More formally though using a statistical testing procedure for data normality, the test does indicate that the data, statistically speaking, are not normally distributed.

Further outliers exist at the lower end of the distribution (marks below 50), thereby making it more difficult to conduct complex statistical routines. Indeed, the Cognos data mining tool used in this report automatically excluded low value outliers.

GPA Data Frequencies by Program / Option

As seen previously in the report, there are differences in GPA (examining averages) among programs. The data below show the frequency distributions of the original GPA data for each of the options to better give the reader a sense of the overall picture. Note, that the scales are different reflecting the varying sizes of the programs:



(note for all figures: the Y axis is the number of observations and x axis is GPA)

Appendix 3: Recoding Provincial English and Math Marks

For the purpose of running analysis on student information, it was determined that a consistent structure was to be established for students' previous provincial high school marks of English and Math. The reason for developing a consistent structure was that provincial grades were coded either as a percentage or a letter grade. Based on the needs of the project, it was decided that a four-point grade point average (GPA) scale be established to ensure that all provincial grades were consistent with one another. The four-point GPA scale, used for the provincial marks were:

Percentage	4.0 Scale	Letter Grade	4.0 Scale
0.00	0.00	A	4.00
0.50	1.00	A-	3.67
0.55	1.67	A+	4.33
0.60	2.00	B	3.00
0.65	2.33	B-	2.67
0.70	2.67	B+	3.33
0.75	3.00	C	2.00
0.80	3.33	C-	1.67
0.85	3.67	C+	2.33
0.90	4.00	D	1.00
0.95	4.33	F	0.00

Due to the nature of information provided for English and Math, each set of marks had to be coded with different steps. Math and English will be broken down into separate sections to provide an accurate account of the coding. The steps for establishing the four-point scale will be described sequentially under the appropriate headings.

Provincial English Marks

Student information was pulled from the Banner Database for the years 1998, 2000, 2001 and 2002. The information was compiled into a large spreadsheet to provide easy access to all information.

Based on the information provided from the Banner Database, the student identification number (SID), English marks and Expr 1 (comments relating to the students grades) were extracted and placed into a spreadsheet. This was to ensure ease of manipulating data to develop appropriate formulas to calculate the four-point scale.

The text to column function had to be used to separate out any comments from the grades, as the transfer of information between the databases did not allow for the separation of marks and comments. The comments were then deleted from the columns to allow for only relevant information to be left for coding.

Due to the transfer of information from the various databases and the different methods of data entry, the grades data needed to be cleaned up before the use of formulas could be established. The cleanup consisted of removing various styles of brackets and quotations.

Once the grades were cleaned up, a formula was developed for running each of the student's grades through the above-mentioned percentage and letter grade four-point GPA scales. It was important that the grades remained consistent with the SID to ensure accuracy throughout the project. The formula developed for determining the four point GPA mark was: `=VLOOKUP(GH3,'look up table'!B3:C13,2)`. `Vlookup` related to the type of formula excel was to use. `GH3` was the cell that contained the student's grade. `'Look up table'` related to above displayed tables, `B3:C13` referred to the column locations in the table, and `2` referred to the column in which excel was to return the appropriate information, the four-point GPA scale number. The formula was applied to all the applicable SID's and a four-point GPA was determined based on the student's percentage or letter grade.

Some of students from the Banner Database had English marks that related to toefl. Toefl uses a more complex marking scheme, as the marks vary on the type of English test taken by the student. The types of tests are toefl, twe, tse and comp. There were no testing marks relating to tse. A grading scheme was developed for the toefl marks to compare the marks against a letter grade. The scores were established based on the guidelines provided by the Admissions Department. The tables are as follows:

Toefl Score	Letter Grade
640	A+
620	A
600	B+
580	B
560	C+
540	C

Comp Score	Letter Grade
270	A+
260	A
250	B+
240	B
230	C+
220	C

Twe Score	Letter Grade
6.5	A+
6.0	A
5.5	B+
5.0	B
4.5	C+
4.0	C

All of the student's toefl scores consist of two of three above listed tests. Combining the total for each test and then dividing by two determined the final scores. These scores were then run through the four-point scale GPA to determine the final value to maintain accuracy with the other SID's established four-point GPA scores.

Once all the student's English and toefl scores were converted into a four-point scale, a code was established to determine the level (or class) of English in which the score related to. This was developed to see if any trends could be established when running the data through Cognos Scenario. The coding of the English class could also allow for quick manipulation of the data by filtering out any specific course. The class codes are provided in the table below:

English Class Code	
1	E12+
2	EN 12
3	E12C
4	E12B
5	Communications Course
6	BCIT Comp Test
7	Toefl
8	University
9	Province
10	College

Potential Other Explanatory Data Sources

Missing from our data but seen in many other studies is some measure of the student's involvement in activities outside class. Another related measure used is some intrinsic measure of a student's overall strength – such as a ranking on their letters of recommendation if such data exists. These more intrinsically derived variables are difficult to construct if at all possible given the nature of the data. This type of data could be collected from the Entry Student Survey but would be self-reported. The key here is that student involvement is seen as a contributing factor to student success.

BC Secondary Exam and Transcript Information

For those students in this study who reported living in BC at the time of entry into their program (approx. 94%), most of these would have come through the BC Secondary School system. However, we were not able to match most of these students back to these transcript tapes to obtain Post-secondary marks and Provincial Marks. We suspect that a more updated study will obtain a better match rates and therefore this could open up an extremely rich data source on previous secondary school marks. Besides marks, there are other sources of potentially useful information such as Physics or Other Non Core marks that are contained on the Ministry of Education supplied information.

Banner Data

Banner data contains extensive information on students' previous institutions attended. In our testing of the data it also seems complete so could be used in lieu of the previous higher education information which is self-reported on the Entry Student Survey. Of course, the Banner information would be more difficult to code as this information consists of the codes of the institutions the student previously attended. The code table would need to be therefore somehow categorized into broad areas. Another source of information is past credential information as captured in a separate table. This data is to a large extent is not known but is listed here for future reference. Again, the ESS Survey contains self-reported educational level information and was used in this study.

One of the ways in which this does affect pre-ranking of students is to somehow enter a score on the strength of a student's application. These scores are kept somewhere and students are ranked not only on the basis of academic measures but other items such as interview "scores". If that ranking data were available for analysis, past research has shown it can make a difference in this type modeling exercise.