## ACADEMIC STREAMING FOR SECONDARY SCHOOLS IN MALAYSIA USING A TWOSTEP CLUSTERING TECHNIQUE

# ELY SALWANA BINTI MAT SURIN

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA KUALA LUMPUR

2015

## ACADEMIC STREAMING FOR SECONDARY SCHOOLS IN MALAYSIA USING A TWOSTEP CLUSTERING TECHNIQUE

## ELY SALWANA BINTI MAT SURIN

## THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

## FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY UNIVERSITY OF MALAYA KUALA LUMPUR

2015

### ABSTRACT

Academic streaming, for students, is used to identify a learning pathway based on academic performance. This is important in assisting them to reach decisions related to their future careers; by distinguishing their academic strengths, weaknesses, interests, and abilities. There are two types of main academic streaming for secondary schools in Malaysia; the science stream and the arts stream. It is crucial to assign students to the correct stream, to avoid a distribution imbalance of students in both streams; as this will affect the needs potential for human capital in Malaysia. To achieve this, in 2011, the Ministry of Education (MOE) set human capital needs ratios of 40% for the art stream and 60% for the science stream. Based on the literature review, problems in determining streaming include: the academic performance of students in schools does not link substantially to the academic streaming process and the academic streaming cannot be determined systematically, due to a lack of understanding of the important factors that influence the streaming process; hence it is therefore difficult to identify the real potential of individual students. Despite several clustering techniques currently used in the educational environment, there is still a lack of appropriate techniques to be used with educational data which relate to academic streaming and academic performance. Based on these problems, this study aims to i) investigate current practices applied in academic streaming for secondary school in Malaysia and the student's academic performance system; ii) identify factors contributing to the academic streaming process and propose an academic streaming framework based on identified factors; and iii) identify and apply a suitable clustering model to distribute students based on their academic performance. This study uses a mixed method research design; whereby the activities of data collection are performed using quantitative and qualitative methods for two different levels of respondents, namely teachers and secondary school students.

Data was obtained from interview sessions with 17 teachers, surveys, and documents analysis for 465 secondary school students. The findings of this research have highlighted issues in the academic streaming process; mainly on the poor linkage between academic performance and the school's streaming process since both of these processes are still connected manually. This research has also managed to identify important factors that relate to the academic streaming and clustering techniques that consists of nine categories which are education institution, peers, family, historical trends, workplace, globalization, employment market, community groups and socio economic status. Based on these findings, in order to stream students systematically, three frequently used technique for clustering which are TwoStep, K-means and Kohonen are analyzed and compared with each other using IBM SPSS Modeler. Based on the analysis, a TwoStep clustering technique is proposed to allocate students to appropriate streams based on their academic performance data. Through this study, current practices applied in academic streaming can be understood more clearly and the linkage between academic streaming and academic performance is identified. Further, a conceptual academic streaming framework for secondary schools is proposed to help teachers and students to understand all relevant streaming determination factors, and provide balance streaming for secondary schools in Malaysia based on academic performance using a TwoStep clustering technique.

### ABSTRAK

Aliran akademik untuk pelajar digunakan untuk mengenal pasti laluan pembelajaran berdasarkan prestasi akademik. Ini adalah penting dalam membantu mereka untuk mencapai keputusan yang berkaitan dengan kerjaya masa depan mereka; dengan membezakan kekuatan, kelemahan, minat dan kebolehan akademik mereka. Di sekolahsekolah menengah di Malaysia, terdapat dua jenis aliran akademik utama; aliran sains dan aliran sastera. Ia adalah penting untuk meletakkan pelajar ke aliran yang betul, untuk mengelakkan ketidakseimbangan pengagihan pelajar di kedua-dua aliran; kerana ini akan memberi kesan kepada keperluan potensi modal insan di Malaysia. Untuk mencapai matlamat ini, pada tahun 2011, Kementerian Pendidikan (MOE) telah manusia keperluan modal nisbah 40% untuk aliran sastera dan 60% untuk aliran sains. Berdasarkan kajian literatur, terdapat beberapa masalah dalam menentukan aliran, iaitu; prestasi akademik pelajar-pelajar di sekolah-sekolah tidak dikaitkan dengan ketara kepada proses menentukan aliran akademik dan iainya tidak dapat ditentukan secara sistematik, kerana kekurangan pemahaman tentang faktor-faktor penting yang mempengaruhi proses streaming ini; Ia adalah sukar untuk mengenal pasti potensi sebenar pelajar secara individu. Tambahan pula, walaupun beberapa teknik pengelompokan digunakan dalam persekitaran pendidikan, terdapat kekurangan teknik yang sesuai untuk digunakan dengan data pendidikan yang berkaitan dengan aliran akademik dan prestasi akademik. Berdasarkan masalah ini, kajian ini bertujuan untuk i) menyiasat amalan semasa yang digunakan dalam aliran akademik untuk sekolah menengah di Malaysia dan sistem prestasi akademik pelajar; ii) mengenal pasti faktorfaktor yang menyumbang kepada proses streaming akademik dan mencadangkan satu rangka kerja aliran akademik berdasarkan kepada faktor-faktor yang dikenal pasti; dan iii) mengenal pasti dan menggunakan model kelompok yang sesuai untuk mengagihkan pelajar berdasarkan prestasi akademik mereka.

Kajian ini menggunakan reka bentuk penyelidikan kaedah bercampur; di mana aktivitiaktiviti pengumpulan data dilakukan dengan menggunakan kaedah kuantitatif dan kualitatif untuk dua tahap yang berbeza daripada responden, iaitu guru-guru dan pelajar sekolah menengah. Data diperolehi daripada sesi temu ramah dengan 17 guru, kaji selidik, dan analisis dokumen untuk 465 pelajar sekolah menengah. Hasil kajian ini telah mengetengahkan isu-isu dalam proses menentukan aliran akademik; terutamanya kepada hubungan yang lemah di antara prestasi akademik dan proses menentukan aliran akademik kerana kedua-dua proses ini masih dihubungkan secara manual, dan faktorfaktor penting vang berkaitan dengan aliran akademik dan faktor vang mempengaruhinya yang terdiri daripada sembilan kategori jaitu institusi pendidikan, rakan-rakan, keluarga, sejarah, tempat kerja, globalisasi, pasaran pekerjaan, kumpulan masyarakat dan status sosio ekonomi. Berdasarkan penemuan ini, untuk menentukan aliran pelajar secara sistematik, tiga teknik yang sering digunakan untuk pengelompokan iaitu TwoStep, K-means dan Kohonen dianalisis dan dibandingkan diantara satu sama lain menggunakan IBM SPSS Modeler. Berdasarkan analisis, teknik pengelompokan TwoStep adalah dicadangkan untuk menentukan aliran pelajar berdasarkan data prestasi akademik mereka. Melalui kajian ini, amalan semasa yang digunakan dalam aliran akademik dapat difahami dengan lebih jelas mengenai kaitan antara proses ini dengan pencapaian akademik; rangka kerja bagi aliran akademik untuk sekolah menengah adalah dicadangkan untuk membantu guru dan pelajar untuk memahami semua faktor yang menentukan aliran akademik pelajar dan menyediakan keperluan yang seimbang untuk sekolah-sekolah menengah di Malaysia berdasarkan prestasi akademik menggunakan teknik pengelompokan TwoStep.

### UNIVERSITI MALAYA

### **ORIGINAL LITERARY WORK DECLARATION**

### Name of Candidate: ELY SALWANA BINTI MAT SURIN (I.C/Passport No: 810426115388)

Registration/Matric No: WHA100002

Name of Degree: DOCTOR OF PHILOSOPHY

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

### ACADEMIC STREAMING FOR SECONDARY SCHOOLS IN MALAYSIA USING A TWOSTEP CLUSTERING TECHNIQUE

### Field of Study: INFORMATION SYSTEM

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature

Subscribed and solemnly declared before,

Witness's Signature

Name: Designation: Date

Date

#### ACKNOWLEDGEMENTS

Alhamdulillah, I thank Allah for His bless, guidance and giving me a strength throughout the journey. I have to acknowledge that the pursuit of my PhD is not a solo journey. There are definitely many people whom this thesis is owed to. My sincere appreciation, gratitude and heartfelt thanks go to Dr. Suraya Hamid and Dr. Norizan Mohd Yasin, my supervisors cum role model, for their endless support throughout the completion of this research, thank you for your patience and guidance.

I would also like to thank all my colleagues; Ida, As, Yani and Ika, for the words of courage and support in various possible ways; Faculty of Computer Science and Information Technology, for the help and support to its postgraduate students; University of Malaya as my employer and Ministry of Higher Education (MOHE), for providing resources and funding to carry out this research; and not forgotten all the respondents of this study, for their kind cooperation.

Last but definitely not least, my deepest appreciation goes to my husband, Mohd Fairuz Tamby Chik, thank you for your endless support, love, patience, motivation and understanding. To my adorable sons, Muhammad Emir Fa'ez, Muhammad Emran Faarish and Muhammad El-Fateh, who has constantly reminded me of what matters most in life, Mommy always loves you. My father, Mat Surin Mamat and my mother Sulosteri Othman for their do'a and support all the time. All my siblings, my mother inlaw, my in-laws and families.

## TABLE OF CONTENTS

ABSTRACT	iii	
ABSTRAK		
ACKNOWLEDGEMENTS		
TABLE OF CONTENTS	ix	
LIST OF FIGURES	xiii	
LIST OF TABLES	XV	
LIST OF ABBREVIATIONS	xvii	
CHAPTER 1	18	
INTRODUCTION	18	
1.1 INTRODUCTION	18	
1.2 BACKGROUND	21	
1.3 PROBLEMS STATEMENT	27	
1.4 RESEARCH OBJECTIVES	28	
1.5 RESEARCH QUESTIONS	29	
1.6 RESEARCH SCOPE	30	
1.7 RESEARCH DESIGN	31	
1.8 RESEARCH CONTRIBUTION	32	
1.9 THESIS OVERVIEW	34	
CHAPTER 2	36	
LITERATURE REVIEW	36	
2.1 INTRODUCTION	36	
2.2 RELATED ISSUES	36	
2.2.1 Human Capital Development	37	
2.2.2 Planning in Education	42	
2.2.3 Information Technology (IT) in Educational Planning	43	
2.2.4 Information Technology (IT) and Student's Academic Performance	44	
2.3 ACADEMIC PERFORMANCE SYSTEM AND ACADEMIC STREAMING	45	
2.4 STUDENT'S ACADEMIC PERFORMANCE SYSTEM IN MALAYSIA	49	
2.5 FACTORS THAT INFLUENCE THE ACADEMIC STREAMING	54	

ix

2.6 ACADEMIC STREAMING APPROACHES	56
2.6.1 Parental Demand	56
2.6.2 Heterogeneous vs. Homogeneous	57
2.6.3 Ability	58
2.6.4 Cluster	59
2.6.5 Summary of Student's Academic Streaming Approaches	60
2.7 ACADEMIC STREAMING APPROACHES APPLIED IN SCHOOL	62
2.8 DATA MINING TECHNIQUES IN EDUCATION	68
2.8.1 Summarization	69
2.8.2 Classification	70
2.8.3 Association	71
2.8.4 Clustering	71
2.8.5 Summary of Data Mining Techniques in Education	72
2.9 CLUSTERING TECHNIQUE FOR EDUCATIONAL ENVIRONMENT	73
2.10 RELEVANT EDUCATIONAL MODELS	79
2.9.1 John Carroll's Model	79
2.9.2 Proctor's Model	81
2.9.3 Cruickshank's Model	83
2.9.4 Biddle Model	84
2.9.5 Gage and Berliner's Model	85
2.9.6 Huitt Model	86
2.9.7 Huitt Revised Model	89
2.9.8 System Theory Model	92
2.9.9 Summary of the Discussed Models	94
2.11 CONCEPTUAL FRAMEWORK	96
2.12 DISCUSSION	101
CHAPTER 3	104
RESEARCH METHODOLOGY	104
3.1 INTRODUCTION	104
3.2 RESEARCH PHILOSOPHY	105
3.3 RESEARCH PROCESS	106
3.4 RESEARCH APPROACH	111
3.5 DATA COLLECTION INSTRUMENTS	118
3.5.1 Semi Structured Interview	118 •
	Δ

3.5.2	Data Analysis Approach For Interview (Thematic Analysis)	119	
3.5.3	Surveys	120	
3.5.4	Data Analysis Approach for Surveys (Partial Least Square Approach;		
	Structural Equation Modelling, PLS-SEM)	124	
3.5.5	Document Analysis – Content Analysis	124	
3.5.6	Data Analysis Approach for Content Analysis	125	
3.6 RES	EARCH TRUSTWORTHINESS	126	
3.7 VAI	LIDITY AND RELIABILITY OF INSTRUMENTS	128	
3.7.1	Validity	128	
3.7.2	Reliability	129	
3.8 SUN	IMARY	129	
CHAPTI	ER 4	131	
DATA C	OLLECTION	131	
4.1 INT	RODUCTION	131	
4.2 DAT	TA COLLECTION 1 – INTERVIEW	133	
4.3 DAT	TA COLLECTION 2 – SURVEYS	137	
4.4 DAT	TA COLLECTION 3 – CONTENT ANALYSIS	138	
4.5 SUN	IMARY	142	
CHAPTI	ER 5	144	
DATA ANALYSIS AND FINDING		144	
5.1 INT	5.1 INTRODUCTION		
5.2 DAT	TA ANALYSIS FOR INTERVIEW	145	
5.3 FIN	DING FOR INTERVIEW	152	
5.4 DAT	TA ANALYSIS FOR SURVEYS	161	
5.5 FIN	DING FOR SURVEYS	165	
5.6 DAT	TA ANALYSIS FOR CONTENT ANALYSIS	172	
5.6.1	TwoStep	174	
5.6.2	K-means	179	
5.6.3	Kohonen	182	
5.7 FIN	DING FOR CONTENT ANALYSIS	185	
5.8 SUN	IMARY OF FINDINGS	186	
5.9 SUN	<b>IMARY</b>	190	

CHAPTER 6		
DISCUSSION AND CONCLUSION		
6.1 INTRODUCTION	192	
6.2 OVERVIEW OF THE RESEARCH	192	
6.2.1 Restatement of the Problem	196	
6.2.2 Revisited the Conceptual Framework	203	
6.2.3 Contributions	205	
6.3 IMPLICATIONS	207	
6.3.1 Theoretical Implication	207	
6.3.2 Practical Implications	209	
6.4 LIMITATION	211	
6.5 FUTURE RESEARCH	212	
6.6 SUMMARY		
6.7 CONCLUSION		
REFERENCES 21		
APPENDICES		
Appendix A: Interview Questions		
Appendix B: A Letter of Permission For Data Collection (Interview)		
Appendix C: Sample Calculation Table by Krejcie and Morgan (1970)	241	
Appendix D: Surveys Questions		
Appendix E: Data For Content Analysis		
Appendix F: Interview – Generated 53 Categories		
Appendix G: Key Quotations Development		
Appendix H: Key Ideas (Individual)		

## LIST OF PUBLICATIONS

251

## LIST OF FIGURES

Figure 1.1 : Research Design (Lewis, 1998)	31
Figure 2.1 : Framework of Integrated Human Capital Development for Malaysia	41
Figure 2.2: Proctor's Model (Proctor, 1984)	81
Figure 2.3 : The Cruickshank's Model (Cruickshank, 1986)	83
Figure 2.4 : The Biddle Model (B. Biddle & Ellena, 1964)	84
Figure 2.5 : Gage And Berliner's Model (N. L. Gage & Berliner, 1998)	85
Figure 2.6 : The Teaching / Learning Theory (Tlt) In Huitt Model	86
Figure 2.7 : The Systems Theory Framework (Stf)	93
Figure 2.8 : Conceptual Framework	98
Figure 3.1 : Research Design	110
Figure 3.2 : Embedded Concurrent Designs	114
Figure 3.3 : Data Collection Steps	117
Figure 4.7 : Current Student's Distribution	142
Figure 5.1 : Important Factors Involved in Student's Streaming In School	155
Figure 5.2 : Factors Involved in The Research	163
Figure 5.3 : Factor Loadings	164
Figure 5.4 : Difficulties in Choosing Suitable Academic Streaming	167
Figure 5.5 : Knowledge About The Requirement for The Particular Academic	
Streaming	168
Figure 5.6 : Cluster Size and The Cluster Quality Using Twostep Technique	178
Figure 5.7 : The Cluster Size Produced by Twostep Technique	178
Figure 5.8 : The Predictor Importance for The Twostep	179
Figure 5.9 : Cluster Size And The Cluster Quality Using K-Means Technique	181
Figure 5.10 : The Cluster Size Produced by K-Means Technique	181

Figure 5.11 : The Predictor Importance for The K-Means Technique	182
Figure 5.12 : Cluster Size And The Cluster Quality Using Kohonen Technique	183
Figure 5.13 : The Cluster Size Produced by Kohonen Technique	184
Figure 5.14 : The Predictor Importance for The Kohonen Technique	184
Figure 5.15 : Distribution of Related Quotations Across Respondents	190
Figure 6.1 : Academic Streaming Framework	204

Figure 6.1 : Academic Streaming Framework

## LIST OF TABLES

Table 2.1 : Features of Academic Performance System in School	48
Table 2.2 : Student's Performance Systems Used in Malaysia	50
Table 2.3: Factors That Influence The Academic Streaming	55
Table 2.4 : Summary of Student's Academic Streaming Approaches	61
Table 2.5 : List of "A Very High Index" of Human Capital Development Countries	63
Table 2.6 : Education System Policy, Education System, Educational Strategy and	
Academic Streaming Strategy in Norway, Japan, Canada And Iceland	67
Table 2.7 : Data Mining Techniques in Education and Appropriateness of Its Use Fo	r
Academic Streaming	72
Table 2.8 : Clustering Techniques That Used in Educational Data	78
Table 2.9: Models Related to The Study	96
Table 2.10: Factors Influences The Academic Streaming	99
Table 3.1 : Mixed-Method Strategies	113
Table 3.2 : The Details Of Questionnaire	122
Table 3.3 : The Four Trustworthiness Criteria (Lincoln & Guba, 1985)	127
Table 4.1 : Data Collection Phases Image: Collection Phases	133
Table 4.2 : Pre-Interview Preparation	134
Table 4.3 : Demographic Information of Teachers For Data Collection (Interviews)	136
Table 4.4 : The Level / Grade Of School In Malaysia	139
Table 4.5: Student's Academic Performance Data Used in The Clustering Analysis	140
Table 4.6 : Band Using in Penilaian Berasaskan Sekolah (Pbs)	141
Table 5.1 : Samples of The Generated Themes at The Early Stage of Data Analysis	147
Table 5.2 : Samples of Key Points Development for Each Category From Responder	nt 1
(T1)	149

Table 5.3 : Sample of Key Quotations for Category of "Student's Academic	
Performance"	150
Table 5.4 : Refined Theme With Details of Themes, Sub-Themes, Example of Quote	es
and Research Questions / Related Literature Review	151
Table 5.5 : A Summary Of Findings From The Interviews Validated By The	
Respondents	152
Table 5.6 : Linkage Between Academic Streaming and Student's Academic	
Performance in The Existing System	153
Table 5.7 : Education Institution Factors Involved in Academic Streaming in School	156
Table 5.8 : Peers Factor Involved in Student's Academic Streaming in School	157
Table 5.9 : Family Factor Involved in Academic Streaming in School	158
Table 5.10 : Historical Trends Factor Involved in Academic Streaming in School	159
Table 5.11 : Employment Market Factor Involved in Academic Streaming in School	160
Table 5.12 : Label For Factor Loadings	162
Table 5.13 : Demographic Information of Students For Data Collection 2 (Surveys)	166
Table 5.14 : Model Measurement For Data Collection 2 (Surveys)	169
Table 5.15: Discriminant Validity	171
Table 5.16 : Summary For The Three Models Used in The Clustering Analysis	185
Table 5.17 : Numbers of Quotation Linked to Important Factors	187
Table 5.18 : Finding in Surveys and Interview (Quotations)	189

### LIST OF ABBREVIATIONS

APMS	Academic Projection and Monitoring System
HC	Human Capital
HCA	Hierarchical Cluster Analysis
HDR	Human Development Report
ISIS	Integrated Students Information System
JPN	Jabatan Pelajaran Negeri
MOE	Ministry of Education Malaysia
NEM	New Economic Model
PBS	Sistem Pentaksiran Berasaskan Sekolah
PMR	Penilaian Menengah Rendah
RMK	Rancangan Malaysia Ke
RMK10	Rancangan Malaysia ke 10
RMK9	Rancangan Malaysia ke 9
SAP	Sistem Analisis Peperiksaan
SAPR	Sistem Analisis Peperiksaan Sekolah Rendah
SAPs	Sistem Analisa Peperiksaan Sekolah
SBP	Sekolah Berasrama Penuh (boarding school)
SPP	Sistem Pengurusan Peperiksaan
SPPBS	Sistem Pengurusan Pentaksiran Berasaskan Sekolah
SEM	Structural Equation Modelling
SisPA	Sistem Pengurusan Akademik
SISPOL	Sistem Peperiksaan Online
ST	System Theory
TLT	Teaching / Learning Theory
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
UPSR	Ujian Penilaian Sekolah Rendah

### **CHAPTER 1**

### **INTRODUCTION**

### 1.1 INTRODUCTION

It is generally held that education contributes substantially to economic growth via the development of expertise and knowledge. This is because economic growth hinges on the human capital that is the result of the education itself, and these are the people who participate in the economic development (Hanushek & Wößmann, 2007; Olaniyan & Okemakinde, 2008; Rodríguez & Nussbaum, 2010; UNESCO, 2003). A well-organized education system that has the capacity to transform students in order to meet labour demands will produce a highly-skilled human capital. For the transformation of students into valuable human capital it is necessary to focus on the students' performance as it is closely linked to the production of human capital (Odden & Kelly, 2008; Pil & Leana, 2009). A student's performance, which is a skill that can be taught, together with the student's soft skills, may be necessary in a specific context, such as in an examination. A student's performance in school is gauged by the number assessments that are conducted periodically (Ajith, Sai & Tejaswi, 2013; Maki, 2002; Osmanbegović & Suljić, 2012). The student's academic performance should be used for the development of the students (Wong, 2003) into valuable human capital.

In order to ensure that the human capital that is produced fulfils the demands of the domestic market, it is necessary to critically shape and plan the formation of the human capital itself from the early stages of the formal learning process, namely from the school level (Côté & Healy, 2001; Gilbert, 2012; Gilman, Kawachi, Fitzmaurice, & Buka, 2003), because it is necessary to invest early and to constantly monitor the

formation of human capital needs if valuable human capital is to be developed (Childs et al. 2008). One channel is through the streaming of students, which is the practice of placing students, who have been observed to have the same levels of achievement, in the same classroom (Forgasz, 2010; Gamoran, 2002; Zevenbergen, 2003). The reason for doing this is to provide students with a thorough knowledge and fundamental introduction and comprehension of the particular field (Liu, Wang & Parkins, 2005; Tai, Liu, Maltese & Fan, 2006) that is closely connected to the student's proficiency, thus developing their skills and helping them to become valuable human capital.

For a long time schools have been using various methods to determine the streaming of students such as parental demands, homogenous and heterogeneous skills, and cluster grouping. Recently, clustering methods have been used extensively in education for the distribution of students according to their performance (Bian, 2010; Gentry, 1999; Trivedi, Pardos, Sárközy & Heffernan, 2011; Zimmermann & Raedt, 2009). Information technology (IT) can be employed as a tool for the streaming of students and has had a direct and positive impact on planning and education (Delahunty, 2000). The aim of the users and managers of any IT system is to be able to use the resources or data productively. A vital aspect of IT is that it uses a combination of data and analysis to recommend suitable streams for the students. The streaming of students based on their performance will aid in future development plans, particularly for mapping out a student's learning pathway and for facilitating in the decision making process according to the clustering technique analysis. The IT techniques that can be used for the streaming of students include clustering and data mining (Agarwal, Pandey & Tiwari, 2012; Romero & Ventura, 2013; Shovon, Islam & Haque, 2012). When large data is required for better decision making, then an appropriate method, such as data mining, can be employed to extricate the information from repositories. The purpose of data mining is to locate valuable information among large clusters of data (Baradwaj, 2011) by concentrating on the use of a few methods and processes to determine the data trends (Shirwaikar & Rajadhyax, 2012). In a clustering method in data mining, the student data is divided into natural groups and a useful summary is provided on the students' learning progress (Hämäläinen, Kumpulainen & Mozgovoy, 2013).

The clustering method in data mining helps in the streaming of students, since potential students in a particular grade are placed together in one classroom (Gentry, 1999). This method is proposed to help in the streaming of students by placing the students in several groups by means of natural clustering according to their academic performance for the whole year. Moreover, this method supports the policy of the Ministry of Education (MOE), where the distribution of the overall student population into the Science and Arts streams should be in the ratio of 60:40 (Kementerian Pendidikan Malaysia, 2013; Utusan Melayu, 2009). This is in line with the economic development plan of the country to develop a human capital that can actually fulfil the needs of economic growth (Ministry of Education Malaysia, 2012; Multimedia Development Corporation, 2005). This concept is also consistent with the mission of the MOE to develop a world-class quality education system in order to exploit the full potential of the individual and to satisfy the aspirations of Malaysians; to support a national economic model (NEM) that can provide human capital to meet the economic needs of the country (National Economic Advisory Council, 2010). Human capital planning that is not in keeping with economic needs will result in a variety of problems, such as a lack of experts in specific fields and an excess of experts in other fields (Aziz, 2006; Economic Planning Unit, Malaysia, 2010; Haslinda, 2009), thus leading to an imbalanced channelling of manpower for economic development.

#### **1.2 BACKGROUND**

The rapid economic development that takes place in a particular country is due to the support from the population. A government's most valuable asset is the people - its human capital - which should be comprised of people who are highly skilled and knowledgeable in their particular fields of occupation (Becker, 1993; Fang & Bei, 2009; Olaniyan & Okemakinde, 2008; Schütt, 2003). This is because human capital plays a vital role in the economic development of the country through the utilization of the skills and knowledge of the people. The demand for skilled labour has increased with the growth of the industrial sector to support economic development. A first-rate education system is necessary in order to produce a highly-skilled human capital. A few researchers (Adawo, 2011; Dasgupta & Weale, 1992; Kyriacon, 1980; Lan & Jamison, 1991; Lank, 1997; Natoli, 2008; Ndiyo, 2002) share the opinion that education is an important component of economic growth. However, in order to ensure that economic needs are fulfilled it is necessary to focus on an adequate and balanced supply of human capital. Therefore, this study is carried out with the objective of planning for human capital needs from the secondary school level onwards, i.e. for secondary three students around the age of 15 years, by identifying the academic performance of the students and ascertaining the correct streams for them through the use of a clustering technique.

This study explores how students are being placed in academic streams at present, and how IT can be employed to aid in the task of academic streaming. In this study, the academic streaming of students is described as the practice of placing students, who have been observed to have similar levels of performance in certain subjects, in the same classroom (Zevenbergen, 2003). The secondary school students are chosen because a study by Sang (2002) indicates that the inclination to learn depends on the age and maturity of the students in the range of 12.73-18.42 years, which is also the age that is regarded as appropriate for making career decisions and for handling career development tasks (Creed & Patton, 2003; Savickas, 1999). In addition, this is the eligible age for them to select and decide on the direction of their learning in school (Ministry of Education, Malaysia, 2013).

According to previous studies regarding the academic streaming of students, either concerning the rules governing the academic streaming process or the outcome of several studies which evaluated the methods of academic streaming, it is obvious that the academic performance of a student is the primary criterion for the streaming (Bellin, Dunge, & Gunzenhauser, 2006; Burns & Mason, 2002; Burris & Allison, 2013; Jennifer Stepanek, 1999). This study also employs the same trend, which is based on the student's performance. In this study, it is clear that the marks of four major subjects, namely Bahasa Melayu (BM), English (BI), Mathematics and Science for a whole year play an important role in determining the academic streaming (Kementerian Pendidikan Malaysia, 2012). A few methods are discussed in this thesis with regard to the academic streaming of students, including heterogeneous, homogeneous, cluster, ability and parental demand methods (Duru-Bellat & Mingat, 1989; Ekstrom, 1961; Esposito, 1973; Gentry, 1999). In this research, the cluster technique is employed for academic streaming. By means of this technique, students who show similar potential in a grade level are clustered together in one classroom. The clusters comprise groups of students who demonstrate similar levels of performance throughout the whole year (Bian, 2010). By placing students in groups, a good summary can be provided on the learning progress of students, which helps in setting the targets for teaching and tutoring (Hämäläinen et al., 2013). These clusters can help to identify the main activities that differentiate the performance of the students. However, this method is normally conducted manually by schools, whereby teachers will go through a list of the students' performance in the final exam, and then divide the students appropriately into classes. In this study, a data clustering method in data mining is proposed.

A clustering method is made up of unsupervised and statistical data analysis, and is used to compartmentalize the sample data into groups of similar data (Halkidi, Batistakis, & Vazirgiannis, 2001; Rokach & Maimon, 2008; Shovon et al., 2012). The clustering method has to do with the detection of significant trends in unlabelled data by grouping together similar data (Trivedi et al., 2011). This method is employed for large datasets to locate hidden trends and to aid in swift and efficient decision making. In other words, data clustering is used to break up a large set of data into subsets known as clusters, with each cluster being a collection of data objects, where those with similar data are placed in the same cluster, and their objects are dissimilar to those in other clusters (Shovon et al., 2012). This task can be linked to the academic streaming of students in the education field.

The academic streaming of students is used to map out the learning path of the students according to their academic performance. One of the biggest challenges in education presently is charting the learning paths of students (Bhullar, Iaeng & Kaur, 2012) so as to direct students into the appropriate streams based on their academic performance and their proficiencies. Hence, it is important for schools to determine in which classes students should be enrolled, and to identify those students who will be requiring extra help in certain subjects. Moreover, the school management will need more information concerning the students, such as their results, in order to measure the success of each streaming or class. However, the current systems do not appear to have the capacity to methodically link the academic performance of the students to their academic

streaming. No system, tool or even standard guideline has been adopted to ensure the organised academic streaming of students (Hallam, Rogers & Ireson, 2008; Rivkin, Hanushek & Kain, 2005; Sukhnandan & Lee, 1998).

In addition, due to an inadequate academic streaming system, teachers are not fully aware of the learning progress of their students (Hawa, 2009). In other words, the personal performance of students in schools cannot be predicted on a regular basis (Gao et al., 2010) because their performance is not being managed and identified properly. It stands to reason that if the members of an organization are not playing their rightful role, then the organization will not be able to function effectively and efficiently. Thus, it is highly important that the method used for the grouping or distribution of students in schools be improved. If the management of each individual student is conducted properly, then their potential can be identified. So, considerable studies are still required in order to implement future plans for students (Lederer et. al, 1988) such as, charting the student's performance towards an appropriate learning path, and recognizing their potential so as to further enhance their skills.

Based on the detailed explanations given above, this research is significant in having managed to identify a number of problems: (1) It is not easy to identify the actual potential of individual students since their performance in school is not linked to the academic streaming process; (2) The academic streaming of students cannot be methodically ascertained as there is no standard guideline for the procedure; and (3) Although several clustering algorithms are available, there are no general guidelines as to which method is appropriate for the educational data in relation to the academic streaming of students and the student's performance.

The clustering method is the solution for all these problems (Bhullar et al., 2012) because it helps schools to make accurate decisions and is a better tool for forecasting the grouping of students according to their performance. However, no general guidelines are available in a range of clustering algorithms with regard to which method should be selected for educational data (Hämäläinen et al., 2013). It is not easy to select a suitable method for a specific task, such as student's academic streaming. In actual fact, researchers frequently just select the most popular clustering method, which is the k-means method, without giving any consideration whatsoever as to whether its basic assumptions match the data (Hämäläinen et al., 2013). What this means in practical terms is that there might finally be an artificial compartmentalization of data instead of the location of natural clusters.

In view of the current system of implementation and the advantages of a methodical academic streaming of students based on their academic performance, this study will explore the existing practice on student's academic streaming and academic performance in Malaysia. This is done by interviewing experts who are those teachers that are tasked with the academic streaming of students, and also by conducting a close-ended surveys among students to determine factors that have an impact on academic streaming. This method involves a sample of teachers and students (secondary schools) at a particular point in time. Several factors that may have an effect on the academic streaming of students in school are identified, leading to the development of an academic streaming model for students. This is followed by a proposal for the best clustering model based on the analysis.

This research is directed at the academic streaming of secondary school students using information concerning their academic performance, whereby the cluster analysis method in data mining is exploited in order to help identify the appropriate streams for the students. This study explores the use of clustering analysis in data mining in the field of education by using data on the performance of Secondary three students in 2013. The data was gathered from five schools in Malaysia. The study describes the processing of the data, the application of the clustering analysis on the data, and finally the benefits gained from the revealed knowledge.

This study is carried out because the academic streaming of students is vital in ascertaining a student's learning pathway. It is not easy to develop the necessary human capital without information with regard to a student's performance and achievements. Industries are currently facing a shortage of experts with the appropriate skills and knowledge in certain fields, especially in the scientific field (Berita Harian, 2012; National Economic Advisory Council, 2010; Utusan Melayu, 2009). Educational advancement in Malaysia does not appear to correspond with economic growth (NEM, 2010). One reason for this is because the human capital that is demanded by industries is not being generated accordingly. A report in the Utusan Melayu (2009) on the Indicator of Science and Technology Malaysia 2004 revealed that science and technology graduates from the universities meet only 32.4 percent of the 60 percent required in the science field. This is inline with the results of a survey on Public Awareness towards Science and Technology 2004, which showed that 42.3 percent of the population in Malaysia are agree with the opinion that Science is a difficult subject.

Therefore, this study will be helpful to schools with regard to decision making, mapping learning paths and enhancing the performance of students. In addition, all the essential data in relation to the performance of students can be employed to provide information to several associated parties, such as parents and administrators (schools, districts and states), and can also be used to chart the plans for human development in Malaysia.

### **1.3 PROBLEMS STATEMENT**

According to the background of the study, the problems which led to this research proposal were identified when the research statement was being drawn up. A summary of the problems is given below:

- i. The academic performance of the students in schools is not significantly related to the academic streaming system, making it difficult to recognize the true potential of individual students;
- ii. The academic streaming cannot be ascertained methodically as there is a lack of understanding with regard to the crucial factors that have an impact on the streaming process; and
- iii. Although several clustering methods are available for use in an educational setting, there is no specific method that can be used with educational data in relation to academic streaming and academic performance.

The problem with the methods being used presently to stream students according to their performance is that the performance of the individual student is not linked to the student's streaming in a single similar system, making it difficult to determine the actual potential of each student. Existing systems for students' performance do not support academic streaming, so the task is executed according to erroneous rules without focusing on the individual students, and thus the academic streaming of the students cannot be determined methodically. This indicates that the student's learning path is not charted to meet the demands of industries. In future, this will result in low-skilled workers, whereby the skills of workers will be unable to satisfy job requirements. Although the cluster method is being used extensively for the academic streaming of students, it is being applied manually in schools. This method is closely related to the clustering method in data mining. Nevertheless, the data mining itself makes use of a range of clustering models which have to be tested each time a suitable method is to be selected. There is no specific model or method that can be employed with educational data in relation to a student's performance.

These issues should be resolved in a proper manner to ensure that in the long run there will be students with the appropriate skills who can be directed into the human capital plan. Thus, the lack of skilled workers for specific fields will not occur in the Malaysian industrial sector and the national production growth rate will not be affected. All the data with regard to the performance of students via their development must necessarily be put to proper use so that all the important data can be applied for the performance of vital tasks for future development and decision making. It is actually necessary to enhance the academic streaming task itself in order to improve the quality of the workers and to produce individuals who can face the challenges of development and promote knowledge based on market needs.

### **1.4 RESEARCH OBJECTIVES**

The main purpose of this research is to examine the methods used for academic streaming in schools and to identify the problems encountered by students in the selection of the correct streams, and then to propose to the schools an appropriate method that could be employed to assist in the academic streaming of students, while taking into account the academic streaming guidelines set by the Ministry of Education, Malaysia. The following are the research objectives (RO) of this study based on the current problems mentioned in the previous section, and in order to achieve the above mentioned goals:

- RO1. To investigate current practices applied in the academic streaming and student academic performance system in Malaysia.
- RO2. To identify the factors involved in the academic streaming process and to recommend an academic streaming framework according to the identified factors; and
- RO3. To identify and use an appropriate clustering model for the distribution of students according to their academic performance.

### 1.5 RESEARCH QUESTIONS

The primary research question in this study is with regard to how academic streaming can be carried out by means of a clustering technique that is appropriate for the data that is provided on the academic performance of the students. The research questions (RQ) for each research objectives (RO) are outlined in more specific terms below:

*RO1:* To investigate current practices applied in the academic streaming and student academic performance system in Malaysia.

RQ1. How are the academic performance and academic streaming of students being conducted in schools?

*RO2:* To identify the factors involved in the academic streaming process and to recommend an academic streaming framework according to the identified factors RQ2. What are the factors that influence the students in their choice of appropriate streams?

RO3: To identify and use an appropriate clustering model for the distribution of students according to their academic performance.

- RQ3. What clustering models can be employed for the distribution of students into suitable streams?
- RQ4. What particular clustering model is appropriate for an academic streaming process that makes use of data on the academic performance of students?

#### **1.6 RESEARCH SCOPE**

In order to achieve the objectives of this study and to enhance the process of academic streaming for secondary school students in Malaysia, this study will make use of the performance data of students from several schools in Malaysia. All the data will be collected from schools under the Ministry of Education, Malaysia (MOE).

Besides that, the scope of the study only covers data concerning the performance of students within the context of the development of valuable human capital. The data regarding the student's performance will concentrate on the four main subjects that are taken into consideration for the purposes of the academic streaming of students, which is *Bahasa Melayu (BM)*, English (BI), Mathematics and Sciences. All the data collected was with regard to secondary three students.

The data will be analysed according to the recommended conceptual framework in Section 2.10, by means of interviews and surveys. This model will be employed to aid in the academic streaming process by identifying important factors that will have an impact on the process.

### 1.7 RESEARCH DESIGN

A research design is defined as a methodical scheme to examine all the issues that have been identified. This research has been divided into four phases - the development, design, analysis, and evaluation phases – so as to answer the research questions and attain the research objectives. The research design, which was developed according to a method proposed by Lewis (1998), is outlined in Figure 1.1, while the details are discussed in Chapter 3. The development phase is theoretical in nature, and involves a review of earlier research literature concerning academic streaming, methods related to academic streaming, technology that can be used to aid in academic streaming, and related topics. The literature is comprised of previous studies that have been carried out and documents obtained from secondary sources such as books, international scholarly journals, proceedings, online documentations, online journals, online proceedings, published and unpublished theses, related articles, and reports from governments/organizations. This phase involves the development of the preliminary factors with regard to academic streaming and the proposal for a conceptual framework.



Figure 1.1 : Research Design (Lewis, 1998)

In the design phase, the tools for data collection are formulated according to the recommended conceptual model. The third phase involves the analytical study carried out to answer the research questions. This phase is aimed at finding the main factors that influence academic streaming and how these factors are connected to each other, as well as developing a model for academic streaming based on the literature review, the semi-structured interview and surveys that were conducted, and analysis of the documents.

Finally, the fourth phase is aimed at verifying and validating the results of the research with regard to the identification of the factors involved in academic streaming and the development of a model for academic streaming in the Malaysian context.

#### **1.8 RESEARCH CONTRIBUTION**

The results of this study are vital in understanding the existing methods that are being used in the academic streaming process for secondary schools in Malaysia, and in discovering whether there is a connection between the academic streaming and academic performance of students. The findings can help in the management of education through improved planning and decision making so as to map out the learning pathway for students. The research investigates the important factors that should be considered by researchers, teachers, and administrators in the MOE at the district and state levels, as well as policymakers in their attempts to understand the use of the clustering method for academic streaming in secondary schools. The data can be used by the MOE to provide students with additional development programs to match their skills with market needs. Several factors in particular that may have an effect on the academic streaming of students based on data regarding their academic performance

will be identified, and a clustering method that is appropriate for the data will be recommended.

From a theoretical point of view, this research employs the factors from Huitt's (1995) Teaching & Learning Theory (TLT), which is based on Bertalanffy's (1969) System Theory (ST) for understanding the experiences of teachers and students with regard to academic streaming activities. Furthermore, all the factors under the nine categories of Education Institution, Peers, Family, Historical Trends, Workplace, Globalization, Employment Market, Community Groups and Socio Economic Status have to be understood. When all the factors in the TLT are applied to the academic streaming concept, they provide some useful insights into the current practice of academic streaming. In this way the research will be able to contribute to the enrichment of the TLT, which was previously used only for teaching and learning purposes, in understanding the academic streaming process for secondary schools.

Besides the development of an academic streaming model and the proposed clustering method, the same data from this study can be used to provide valuable information for the future development of students and schools. It will also help students to gain a better understanding about suitable career fields for their future. Furthermore, it can enhance the teaching and learning process because teachers will be informed about the academic performance of individual students and their learning pathway, and this information will help them to put every effort into developing these students. A well-developed cluster grouping can place students in the appropriate academic streams, and help teachers to better fulfil the needs of students because by placing high achievers in one classroom, their chances of having their needs met will increase, while other students will have the opportunity to develop in the other classrooms. This is vital for generating skills in specific fields to achieve the government's plan for meeting the human capital needs of the country.

### **1.9 THESIS OVERVIEW**

This thesis is generally organized into six chapters:

Chapter 1 discusses the basis of the research. In this chapter, the research background is outlined and it gives a brief introduction to the research. The chapter also highlights the background concerning the academic streaming in secondary schools based on the academic performance of students. It describes how a student's learning pathway can be mapped out by grouping the students according to their academic performance by means of a suitable clustering method.

Chapter 2 reviews the literature in relation to the student's academic performance and academic streaming, and how these two tasks are connected to each other. The chapter also examines the methods used for the academic streaming of students in the field of education, particularly in schools, the clustering method in data mining that is employed for the distribution of students, and the theories in relation to the student's academic streaming process.

Chapter 3 explains the research methodology used in this study and the way in which the research was carried out. This chapter encompasses the research setting, research sample, research tools, and the data analysis methods. Chapter 4 describes the methods employed for the collection of data. This chapter also gives a summary of the data analysis methods used, and explains the implementation of the analysis.

Chapter 5 presents the significant results of the data analysis as discussed in detail in Chapter 4. It also explains the clustering analysis that is carried out on the student's academic performance data in order to select a clustering model that is appropriate for the data provided.

Finally, Chapter 6 presents the main discussion and conclusion for this research. The chapter describes the major findings, and shows how the concepts and theories discovered in the literature have been adapted. It also presents the research contributions, limitations and implications of the study, as well as suggestions for future research work.

### **CHAPTER 2**

### LITERATURE REVIEW

### 2.1 INTRODUCTION

This chapter aims to review of existing research on academic streaming. The chapter deeply focuses on some ideas on implementation of academic streaming and student's academic performance, academic streaming approach in school, clustering techniques that can be applied and relevant theories that support the academic streaming, as well as the related issues. This chapter focuses on student's academic performance data as an important item to be considered for academic streaming purposes. Though, it is not only focuses on the procedures but also includes the theories and its factors that related to academic streaming. A few previous studies and related issues are reviewed for deeper understanding on the academic streaming issues. The findings are used as a fundamental to identifying the important factors involve, at the same time attempt to propose a conceptual framework for academic streaming, as well as proposed a solution that can assist the academic streaming process. The chapter also presenting an analysis and synthesis of the literature on suitable approach for academic streaming and related technique can be used for the approach.

### 2.2 RELATED ISSUES

In this section, issues related to academic streaming and students' academic performance are explained. These issues are taken into consideration, as they are the basic awareness that contributed to the idea of this study.
#### 2.2.1 Human Capital Development

The term of human capital (HC) does not represent the people or employees in the organization instead; it is a recognition of individual efforts that contributes to the best achievement of the organization. There are two components in managing skills and knowledge effectively. These components can be defined as to identify and manage the work-oriented skills and knowledge, and to identify and manage the worker skills and knowledge.

According to Marimuthu (2009), his study validates the fact that impact on economic performance is highly positive through the consideration of human capitals. There is rationally strong evidence to prove that the blend of human capital development in organizations promotes innovativeness and greater firm performance (Marimuthu 2009). Becker (1993) said there is a relationship between the amount of investment in human capital to improve the HC performance and the quality of the workforce, which at the end of the day will contribute towards the economic evolution within a country.

However, the implementations of the right HC planning that in line with available human resources are always neglected. While the importance of early monitoring of potential HC development is vast, proper planning from the beginning of the learning process requires an attention. Therefore, planning that started early from school will promise the identified potential of this HC during their time at the university and work after graduated.

HC formation exists as early as the learning process began through systematic planning and effective learning. This HC development is strongly supported by the learning plan at the university and school. Thus, the learning plan process should be consistent with the national economic plan as the country's economy relies heavily on HC possessed by particular countries.

The concept of HC can be explained in many ways. However, it can be summarized as the relationship between human and his/her potential, knowledge, skills and enthusiasms to persuade human productivity. It also involves in acquiring knowledge and uses it in various areas of common activity that contributes to grow productivity.

Additionally, the HC is viewed as the key factor of production and investment in the economic development (Deutsche Bank Research Marketing, 2005). In the process of economic growth, the HC involves through a collection of capabilities, knowledge developed through the learning process and familiarity, as well as individuals' internal factor that required completing tasks and creating profitable value (Tadic 2010). On the broader perspective, Leeuwen (2004) viewed HC consisting not only education but also the rising costs. This includes variables such as education, ability and on the job training to reflect the HC as these factors control the earnings.

HC is an important asset to the country, which performs labor and produce economic value. The significant attributes involved in the HC gain by the individual through education and experienced. Over time, HC needs to be enhanced to positively transform the organization to compete in a new era of economic growth. This includes recruiting, investing and supporting people, utilizing various tools among other things, training, coaching, mentoring, internships, business development and human resource management.

The HC development in Malaysia has several groups of training and learning that can be explained as follows. First, all Malaysians are typically undergo 12 years of fundamental education in school. This is followed by the extension of the formal education that can be obtained from the universities, colleges and other similar institutions, or formal training from the technical and vocational institutions such as polytechnics all over the country. Finally, for working individuals, the informal training and development of various programs are occasionally offered by the workplace. Furthermore, various funds for the informal training and development are available to all manufacturing and selected service industries regardless of their sizes. With this financial assistance, the companies are responsible to provide the workers with the knowledge and skills needed by the industries (Abdullah et al. 2007).

In the 10<sup>th</sup> Malaysia Plan report (2011-2015), the Prime Minister stated that the Republic of Korea redeveloped their country by focusing on human capital. They successfully recovered from a country that was damaged in a war to a country with the economy of a superpower nation. Half a century ago, Republic of Korea has many low cost workforces coming out with the ideas that education and development of human capital is an important way to increase productivity in terms of development and attain the status of a developed nation. They realize that to compete globally, they need to create an education system that could ensure all students are supported and succeeded without ignoring the students with low performance. Therefore, they have done a massive investment in education tertiary. To achieve this objective, Republic of Korea gives a full attention to strengthen their education at all levels, from primary school to the university level and skills training. Initiatives taken included; compulsory in attending the secondary school, strengthening technical education and training. The close cooperation with *chaebol* and big companies in the Republic of Korea is

established such as set up a corporate university as lifelong learning institution. Currently, the number of enrollments in the public and private higher learning institution is estimated as many as 3.6 million students. At the same time, the Republic of Korea is planning the entire education system so that the provision of human capital needed are equivalent with the industry demands.

In Malaysia, Prime Minister Datuk Seri Najib Tun Razak emphasized that 'The development of human capital is one of the central strategies in the formulation of the New Economic Model for the country which launched at the end of March 2010'. One of the 10 major ideas in the 10th Malaysia Plan (RMK10) is to nurture, attract and maintain exceptional human capital. He also stressed the importance of creating a perfect seamless continuum in the development of innovative human capital through the national education system. To achieve this objective, Malaysia creates a framework of Integrated Human Capital Development (Refer Figure 2.1) to support the strategic plan and prepare the human capital according to the market needs. This framework will increase the knowledge and individual skill continuously throughout the early education, basic education, tertiary education until post-tertiary education and working. For the time being, workforces in Malaysia are relatively unskilled. A total of 77% of the workforce is having only basic education for 11 years, which is Sijil Pelajaran Malaysia (SPM) or ranks with it, and only 28% of employment is highly skilled occupation group (10<sup>th</sup> Malaysia Plan report).



Figure 2.1 : Framework of Integrated Human Capital Development for Malaysia

After SPM examination, students will pursue pre-university or matriculation program as entry requirement for local university admission. The first Matriculation Program in Malaysia was introduced on 1 September 1998 by the Ministry of Education Malaysia for the 1999/2000 session. Formerly, Matriculation Programs were administered by the local Higher Learning Institutes. Then, the unification of all the Matriculation Programs was established under Matriculation Division by the Ministry of Education Malaysia (MOE).

Matriculation is a preliminary program (pre-university) for Malaysian students with 'Bumiputera' and 'Non-Bumiputera' status under a quota system to enroll in a Degree Program in the fields of Science and Technology offered by local universities. This program is run for two (2) semesters and students are fully sponsored by the MOE. The curriculum focuses on the academic and co-curriculum aspects as students are prepared for academic excellence, leadership and other outstanding qualities. All students from any Matriculation Program in the country are entitled to the same syllabus and undergo the same evaluation methods. The students are assessed by a major exam at the end of each semester along with assignments and lab work. Once the students have passed their Matriculation Program, the selection for admission to university program will begin.

Nevertheless, the Matriculation Program fails to enable the students to fully achieve their potential due to "watered-down" syllabus and the examination structure. As a result, the majority of these students fail to fully cope with the subsequent university education and in certain cases, the knowledge they gain from the matriculation program doesn't apply directly during their studies at the university level.

#### 2.2.2 Planning in Education

Planning is a very critical task in education and without it, achieving the expected results and ultimate goals may be difficult and contributed to the unexpected problems (Chang & Radi, 2001; Chang, 2008; W. Smith, 2011). There are a lot of resources available in schools, including the students themselves. However, more resources do not automatically set for a better outcome, but it depends on the way the resources are utilized.

In management operation, planning involves four essential phases such as analysis, planning, implementation and evaluation. On the other hand, planning in the education sector consists of system analysis, policy formulation and action planning (Chang, 2008). Planning for learning pathway or academic streaming can be categorized as action planning. This plan is intended to identify the appropriate area or pathway that suitable with the student to unlock their potential for future career.

#### 2.2.3 Information Technology (IT) in Educational Planning

Information technology (IT) can be defined as a set of elements or related components, which are collect (input), process, stores and spread (output) the data and information. It provides a feedback mechanism to achieve specific objectives and support the decision making and control in an organization (Turban et al., 2004). Mainly, IT consists of hardware combination, software and telecommunication network that built and used by people, to create, raise and distribute useful data through an electronic network with certain procedures (Mcleod, 1998).

IT and its infrastructure can be utilized as supporting tools in the planning process. It can help individuals or a group of people who may have diverse perception and main concern to interact and organize their activities (Rathwell, 1985). In decision making, coordination and control, IT could also help in analyzing problems, figure out the complex situation and generate new ideas based on the analysis.

IT in educational planning also involves the definition of database and systems to support any related applications. Then, this information is extracted from the database to provide knowledge of particular cases. Additionally, it includes the selection of applications that would best fit the existing operation and needs of the organizations (Lederer et al. 1988). Furthermore, IT in educational planning is much like tactical development in an organization. They consist of several parts that need to be formalized before the plan, such as goals, preferences and authorization is set up. The plan must be detailed to facilitate understanding of important roles for the application.

Basically, educational planning started by identifying the needs of the organization and its whole contents. Then, follows by the process of recognizing potential computer application that the institutions should implement based on their objectives (Lederer et al. 1988). Any computer-based application that is invested and developed should fulfill requirements at all levels of operational (Pant et al. 1995).

A valuable and careful plan employing IT and its infrastructure can assist the institutions to reach a particular goal (Hartog and Herbert, 1986). However, the measurement of the effect of a successful educational planning using IT remains a significant problem. Nevertheless, such combination of education and IT, not limited for teaching and learning purposes has permitted important growth in the operation (State of Tennessee: Information Systems Planning Process 2007).

#### 2.2.4 Information Technology (IT) and Student's Academic Performance

IT is widely implemented in various fields including education especially in teaching and learning in schools. Apart from that, IT is also used to manage student performance, as they are an important asset to the country. Student academic performance is vital to determine the direction of the student and school in effort to develop human capital needed by the country.

Generally, an academic performance system requires a collection of information over a specific period of time for data analysis. This will result in an estimation of the

condition or rate of change in the subsystem or system associated with the institutional actions. The system is based on three main components, which are, 1) regular collection of information, 2) assessment of information 3) evaluation of the results obtained in an institutional action (Kiesler & Sproull, 1982).

Malaysia education system has begun to adopt the technology in their students' academic performance system in the late 90s. During this time, most of the schools in the country were using office software packages such as MS. Word and MS. Excel to record the students' performance. Then, in 2000 the standalone program that operates independently without connecting to any electric transmission and distribution network was introduced. Six years later, web-based technology for students' performance system such as ISIS was established in boarding school. The use of such system was fully expanded to all schools in Malaysia in August 2011 due to the advancement of the network facility throughout the country.

### 2.3 ACADEMIC PERFORMANCE SYSTEM AND ACADEMIC STREAMING

Student Learning Pathway in education refers to academic streaming for a particular student (Cooper, Coll, Bartko, Davis, & Chatman, 2006; Taylor, 2007). It can be described as direction or pathway of the student through their development in school that will support and prepare them to meet and exceed their expectations (Mittendorff, Jochems, Meijers, & den Brok, 2008). Determination of academic streaming is done based on the students' academic performance for a specific subject.

Over the last two decades, there has been an increasing amount of attention paid to student's academic performance system (e.g., Dimmock et al, 1997; Abdullah 1999; Guskey & Bailey, 2001; Smith et al, 2001; *Kalz, 2008;*). Generally, the system keeps a record of student activities and their performances. With this information, they can identify their strengths and weakness and plan for the improvement (Andrade, 2009). In general, Academic performance system is focusing on teaching and learning assessment (Dimmock & Wildya, 1995). It is designed to specifically handle daily routine operations in managing student's marks rather than use available resources of student's academic performance data for students' learning pathway (Petrides & Ufsd, 2002; C. M. M. Smith & Sutherland, 2003).

Most of the academic performance system implemented in the school is focused on school performance rather than plan for students' learning pathway or academic streaming. Only a fraction of the available system is dedicated to academic streaming. However, these systems only consider the year-end performance instead the whole year performance (Rivkin et al., 2005). Moreover, the streaming process is performed manually by the teachers by using independent sources of data that are not related to the plan for students' pathway (A. Z. Abdullah, 2006; Budhwar & Sparrow, 2002). The input data are often incompatible with the academic streaming that focused on a group of students such as class and school not the individual performance. The features of the academic performance system that are currently used in school are listed in Table 2.1.

It is a well-known fact that the learning pathway or academic streaming is depending on the students' academic performance. For that reason, the academic performance data are frequently utilized for academic streaming (Deno & Reschly, 2009; Harlen & Malcolm, 1997; Mandeville, 1988; Wynne, 2011) (Deno & Reschly, 2009; Petrides & Ufsd, 2002; Rivkin et al., 2005; C. M. M. Smith & Sutherland, 2003; Willms, 2000; Wynne, 2011) and day-to-day decisions making by school, district and state administrator pertaining to the distribution of resources (Willms, 2000).

	Teaching and Learning Assessment	A Record of Student's Performance	Student's Strengths	Student's Weakness	Need For Support for Students' Learning Pathway	Not Included Academic Streaming/ Students' Learning Pathway Plan	Use Independent Sources of Data	Focus on Group of Students, Not Individual	Day-To-Day Decisions
Dimmock et al, 1997					-				
Abdullah 1999				V					
Guskey & Bailey, 2001					·				
Smith et al, 2001;									
Kalz, 2008									
Andrade, 2009				$\checkmark$					
Dimmock & Wildya, 1995									
Petrides & Ufsd, 2002;									
Smith & Sutherland, 2003									
Rivkin et al., 2005									
(Abdullah, 2006;									
Budhwar & Sparrow, 2002									
Deno & Reschly, 2009									
Wynne, 2011		•							
Mandeville, 1988									
Harlen & Malcolm, 1997									
Willms, 2000									

# Table 2.1 : Features of Academic Performance System in School Use

MOE has a vision that "Ideal School Generates Glorious Generation" with a mission that intend to develop a world-class quality education system that realize the full potential of individuals and fulfill the aspiration of the nation (Kementerian Pendidikan malaysia, 2012). These vision and mission will be accomplished if the progress performance systems are implemented that involve students as an important asset in developing an excellent human capital to achieve the national aspiration (Abdullah, 2006).

Education in Malaysia is on-going efforts towards further developing the potential of individuals in a holistic and integrated manner. Such efforts will produce intellectual, spiritual, emotionally and physically balanced and harmonious individuals based on a firm belief and devotion to God (Hamid & Zaman, 2009). Additionally, these efforts are expected to generate citizens who are knowledgeable, competent and responsible with high moral standards. A high level of personal well-being will contribute to the harmony and betterment of the society and nation. The existing systems that adopted to manage student's performance are summarized in the Table 2.2 (Kementerian Pendidikan malaysia; Khan, 2005; Lembaga Peperiksaan, 2012).

System	Year	Implementation	Support for Student's Academic streaming in The System?
Record Book	Before 1957	All schools under the Britich colonial	No
Ms. Office (Words, Excel)	Late 90's		No
Sistem Pengurusan Peperiksaan (SPP)	2000	Primary and secondary schools	No
Sistem analisa peperiksaan (SAPR16)	2001	Primary Schools	No
Sistem analisa peperiksaan (SAP123)	2001	Secondary schools (form $1-3$ )	No
Sistem Analisa peperiksaan (SAP45)	2001	Secondary schools (form 4-5)	No
e-Pantau	2001	Terengganu, Kedah, Johor	No
Integrated Students Information System Version 1 (ISIS - 1)	2002	Boarding schools	No
Academic Projection and Monitoring System (APMS)	2002	Boarding schools	No
Integrated Students Information System Version 2 (ISIS - 2)	2006	Boarding schools	No
Sistem Pengurusan Akademik (SisPA)	2008	All schools in Perak	No
Sistem analisa peperiksaan (SAPR16HC) (Mohd Badli Rasli,2009)	2009	Primary Schools	No
Sistem analisa peperiksaan (SAP123HC)	2009	Secondary schools (form $1-3$ )	No
Sistem Analisis Peperiksaan Sekolah (SAPs)	August 2011	Primary and secondary schools	No
Sistem Pengurusan Pentaksiran Berasaskan Sekolah (SPPBS)	September 2011	Primary schools (standard 1)	No

Table 2.2 : Student's performance systems used in Malaysia

Traditionally, the students' academic performance were documented in a record book to measure the their performance (Abdullah, 2006). This record book also used as a channel to report the student academic performance to their parents. To date, there are some schools still remains using the traditional recording book along with electronic-based academic performance systems.

Since a formal education was first introduced in Malaysia, various systems have been used to manage student performance as presented in Table 2.2. In the late 90s, many schools have been introduced to the office software package to manage the students' academic performance such as Microsoft Office (Abdullah, 2006). This software only allowed a limited recording work and analysis.

In 2000, *Sistem Pengurusan Peperiksaan* (SPP) was introduced in schools to record the student achievement and analyze it based on individual, class and stream performance (Yusof, 2000). This system has functionality to generate a graduation certificate based on the final exam results that used only at school level. However, a scheduled report to the district office will be in the form of hard copy that printed from the system.

Further improvement has been made from the SPP system with a more systematic analysis function. This new version of the SPP was released in 2001 and reintroduced as *Sistem Analisis Peperiksaan* (SAP). The SAP system consisted of several versions that serve the same purpose to manage the student performance, namely SAPR16 for primary school students, SAP123 for lower secondary students (Form 1-3) and SAP45 for upper secondary students (Form 4-5) (Mohamad et al., 2009). Later in 2009, the head-count function was added to the system and renamed as SAPR16HC and SAP123HC.

At the same time as SAP was released, another improves version of SPP known as *epantau* system was used in the states of Terengganu, Kedah and Johor (Yusof, 2000). The main function of *epantau* system is to track the student academic performance through periodic testing and monitoring as well as improving the state-wide academic performance in the *Ujian Penilaian Sekolah Rendah* (UPSR) (Rohani & Khiruddin, 2003; Yusof, 2000). This system was connected on-line and assists in the process of

counting marks. Unfortunately, the count has to perform manually by the teachers to reduce load on the server.

A year later in 2002, Integrated Students Information System Version 1 (ISIS 1) was introduced in boarding schools (SBP) to records curricular and co-curricular student performance (Azlina, 2001). This system only analyzed the curricular performance for scoring system and records the co-curricular performance for activity report. Then, the ISIS1 was transformed from a standalone system to an online platform known as ISIS2 in 2006. In the same year, Academic Projection and Monitoring System (APMS) was developed that works with ISIS. It is a headcount extension for the ISIS that aims to calculate the scores, comparing the marks and generate the graph for comparison purposes (Azlina, 2001). Among the functions of APMS are import and export the data directly to Excel, generate slip headcount and print directly after the slip is transferred to Excel. Furthermore, the system allows the data to be generated as web pages and sent to the management of SBP.

In 2008, Jabatan Pelajaran Negeri (JPN) Perak has initiated a system known as *Sistem Peperiksaan Online* (SISPOL), an extension of the SAP system. This system was first implemented in the district of Batang Padang, Perak before it was introduced throughout the state of Perak as *Sistem Pengurusan Akademik* (SisPA) (Mohamad et al., 2009). SisPA is a fully online system and only limited to the state of Perak.

For standardization of academic performance system nationwide, a *Sistem Analisa Peperiksaan Sekolah* (SAPs) was established in August 2011. The SAPs system is an adaptation of SISPOL and SisPA that used to collect, store and analyze internal data for

every school examination in Malaysia. Analysis conducted through this system is based on the Malaysian examination grade scale and using the headcount analysis system.

Then, in 2010 The Deputy Prime Minister, YAB Tan Sri Muhyiddin Yasin was announce to reduce examination-oriented system in the country with the possibility that UPSR and *Penilaian Menengah Rendah* (PMR) examinations to be eliminated (Berita Harian, 2010). Therefore, the MOE have proposed a method for student achievement assessment in a holistic manner by combining the test scores between examination and Sistem Pengurusan Pentaksiran Berasaskan Sekolah (SPPBS) with a ratio more favorable to the PBS (Lembaga Peperiksaan Malaysia, 2011). This system will undergo a schedule implementation in stages, while a thorough study are conducted.

The SPPBS system was first tested in September 2011 to record the performance for Standard 1 students. The assessment structure consists of formative (40%) and summative tests (60%). The formative test is an informal test such as simple tests, quizzes and homework. This test is conducted continuously in the classroom to identify weaknesses in teaching and learning process. Contrarily, summative test is a formal test that usually conducted at the end of the topic, middle or end of the school term. This test will identify the achievement of students in a particular period of learning.

For all systems, a submission of scheduled report to the district office is required for data collection and evaluation. The reports that are generated from offline systems such as Ms. Office, SPP, SAP and SAPR can be submitted manually to the district office. However, the reports that are generated from ISIS, SisPa, epantau, SAPs and SPPBS

require a direct submission to the district office through the system itself. Occasionally, the district office will hold a meeting for further discussions of the collected data.

Various systems have been introduced, implemented and improved to manage the students' academic performance in Malaysia. Nevertheless, these systems are dedicated to record the assessment marks and the generated report has minimal influenced in academic streaming process. Therefore, in order to enhance the streaming process, factors that contributed to the academic streaming should be identified, recognized and established their relationship.

#### 2.5 FACTORS THAT INFLUENCE THE ACADEMIC STREAMING

The academic streaming can be determined based on several factors that must carefully consider. These factors will influence the students in making decisions of their learning pathway and future career. In reality, the students have limited guidance from school and teachers in choosing an appropriate academic streaming. Consequently, they turn to parents and friends for opinions based on their final exam results (Duru-Bellat & Mingat, 1989; Zevenbergen, 2001; Gulyaev & Stonyer, 2002; Chapman & Mählck, 2004; Bardick, Bernes, Magnusson, & Witko, 2006; Designs, 2005; Kilgour, 2008; Forgasz, 2010).

Additionally, elements such as career options and current industry make it difficult for the student to choose an appropriate academic streaming for them. Without a proper exposure to the career trends and market needs, they are likely making a wrong choice. (Forgasz, 2010; Kilgour, 2008; Designs, 2005; Schneider, Judy, Ebmeye, & Broda, 2014). Besides, students often made an inappropriate decision on their learning pathway due to high expectation on future regardless of their academic performance. This will result in incompetence of adaptation and disappointment to the selected pathway. (Kemple & Snipes, 2000; Gulyaev & Stonyer, 2002; Gamoran, 2002).

From the published literature, nine main factors are identified that affect the students' decision on their academic streaming. There are parental and peer pressure, high expectations, multiple career option, no guide, new industry career trends, final exam result, teacher advice, parent's aspiration and friends as summarized in Table 2.3.

	Parental + peer pressure	High expectations	Multiple Career Option	No guide	New industry career trends	Final exam Result	Teacher advice	Parent's aspiration	Friends
(Duru-Bellat & Mingat, 1989)	/		/	/		/	/	/	
Kemple & Snipes, 2000	/	/		/					/
(Zevenbergen, 2001)	/			/		/			/
(Gulyaev & Stonyer, 2002)	/	1		/		/		/	/
(Gamoran, 2002)	/		/				/	/	/
Chapman & Mählck, 2004			/			/		/	/
(Bardick, Bernes, Magnusson, & Witko, 2006)				/		/		/	
Schneider, Judy, Ebmeye, & Broda, 2014	Τ			/	/		/		/
(Designs, 2005)			/		/	/			
(Kilgour, 2008)	/		/	/	/	/			
(Forgasz, 2010)	/		/	/	/	/			

Table 2.3: Factors That Influence the Academic Streaming

Once the factors are identified, the approaches of students learning pathway need to be explored. Only then, a solution can be proposed. Basically, an academic streaming is based on five approaches, which are parental demand, heterogeneous, homogeneous, ability and clusters. All these approaches are discussed in the following section.

#### 2.6 ACADEMIC STREAMING APPROACHES

Academic streaming approaches are the methods to determine an appropriate academic streaming for the students. There are five academic streaming approaches that discussed in the following sub-section, which are parental demand, heterogeneous, homogeneous, ability and cluster. All these five approaches are used and applied based on studies from 1961 until 2013.

#### 2.6.1 Parental Demand

Parental demand is the most classic approach as they are entitled full rights on their child's academic streaming. Frequently, parents outlined the decision and disregard the performance of their children, even the opinions from the student and teachers (Duru-Bellat & Mingat, 1989). Unfortunately, this approach is widely implemented and only prioritized the views and needs of the parents. However, Reay, David, & Ball (2005) recognized that there is a pattern of dissimilarity on the ways decision are made between parents from different socio-economic status. This dissimilarity are more obvious between middle-class and working-class or immigrant parents.

The middle-class parents made a better decision based on the important concerns regardless of their differences in education (Glascoe, 2000). A majority of these parents understood their children (Glascoe, 2000) and prepared to sacrifice for their children's education (Lam, Ho, & Wong, 2012). Furthermore, they understand a specific detachment of responsibility and work, between school and parents (Lam et al., 2012).

They are also involved in the student's learning process and has a significant impact on the student's development (Okagaki & Sternberg, 1993).

#### 2.6.2 Heterogeneous vs. Homogeneous

Heterogeneous grouping is a type of distribution of certain grade students in a school. The students are group according to specific characteristics and ability (Zeng, Xu, Shi, Wang, & Wu, 2007) and taught together in order to achieve academic success for entire class (Shields, 2002). For example, students with high performance are scattered throughout the multiple classroom equally. However, there is a risk that high-performance students may not be comfortable to spend too much time teaching colleagues or low-performance students that need one to one discussion. Therefore, as multiple groups of characteristics and ability in a class, variety of approaches in teaching and learning are required (Slavin, 1992)(Oakes, 2005).

In contrary, homogeneous grouping placed the students with similar abilities in a classroom. This group has minimal range of differences of abilities if compared to heterogeneous classroom (Ekstrom, 1961). All gifted students within the same grade level will be in the same classroom. Several studies has supported that the homogenous grouping are the better way to distribute students between classroom compared to heterogeneous grouping (Gamoran & Berends, 1987; Kerckhoff, 1986; C. Kulik & Kulik, 1982a; C. L. Kulik & Kulik, 1982b). Homogeneous grouping accordance to the ability will motivate the students with a high ability to achieve their potential but the students with low performance will have less chances for improvement (Esposito,

1973). Moreover, this type of grouping will cause polarization in the classroom distribution.

#### 2.6.3 Ability

Ability grouping is the placement of children in a classroom into groups based on their ability. The classroom may contain children with a wide range of ability (Preckel, Götz, & Frenzel, 2010). It is a small, informal groups formed within a single classroom and short-term (never lasting longer than one school year), and varies by subject (Gamoran, 2002). Tomlinson (2001) recommends that a distinguish instructions with innovative study material can be given to high achiever students.

For more effective grouping, students can move between groups as needed for different assignments. For example, a student may excel in mathematics, but having a problem in reading. This student can be arranged in advanced group in a mathematic lesson to perform more challenging assignments and at the same time he/she will be placed in lower group in reading lesson to attend his/her problems with reading. This flexibility of in grouping students will allows more attention given to their needs.

In ability group, it is assumed that all students do not learn in the same approach and at the same speed. Instructions must be altered based on their ability to improve the performance for all students in the group (Siegle, 1992). Therefore, teachers are responsible to discover teaching methods or material that is suitable to the students' ability (Pare, 2004).

#### 2.6.4 Cluster

Cluster grouping has been used in schools to encounter the academic essentials of potential students (Zimmermann & Raedt, 2009). Students with the same grade level are grouped together in one classroom (Gentry, 1999). Consequently, the students who are mathematically gifted will be placed in a classroom while the verbally gifted in another classroom.

Carol C. Burris & Allison (2013) in their study for National Education Policy Centre (NEPC) explored the effect of sorting on overall student achievement or cluster grouping and found a significant learning gains. This grouping technique involved a similar type of materials and assignments, teaching plan based on student needs, and interest-based instructional among others (D. W. Caldwell, 2012). To sort the students, teachers and schools must use various sources of data to recognize their skills effectively. However, there are many unidentified students due to lack of resources (Jennifer Stepanek, 1999).

Overall, cluster grouping can positively affect the performance of all students. A welldeveloped cluster grouping can offer a special program for high achieving students and attend a special need for those that are under performed. However, this grouping should be done dynamically based upon development needs and detailed learning condition Bellin, Dunge, & Gunzenhauser (2006), Benson (2002) and Burns & Mason (2002)

#### 2.6.5 Summary of Student's Academic Streaming Approaches

Generally, the student's academic streaming approaches can be summarized as in Table 2.4. These approaches are widely used to help determine the student's academic streaming. One of the basic approaches is parental demand and it is incompatible with the main purpose of streaming, which is to place the students based on their performance or talents rather than the perspective and needs of their parents.

Another approach is heterogeneous grouping that place the students from different abilities and educational needs in a classroom and learn together. They will be in the same age, but having different levels of achievement. Contrarily, the homogeneous approach placed the students of similar achievement in a classroom. Although a range of achievement exists in this classroom, it is smaller than the range found in the heterogeneous classroom.

The students also can be grouped based on their ability and skills. In ability group, it is assumed that students with different ability have different capability of learning. These students required different ways of teaching and learning process based on their ability and skills.

On the other hand, cluster approach placed all ranges of students in a classroom and later grouped the students in the classroom based on the similar achievements. This approach allows the flexibility of replacing the student according to their achievement on a particular subject. In fact, it is a comprehensive approach to address various needs of every student. To apply this approach, the data mining can be used to assist in the process discovering the students' ability based on the data sets and used it to classify the group.

	Heterogeneous	Homogeneous	Cluster	Ability	Parental Demand
Ekstrom (1961)		/			
Esposito (1973)	/	/			
C. L. Kulik & Kulik (1982)		/			
Kerckhoff (1986)		/			$\boldsymbol{\Lambda}$
Gamoran & Berends (1987)		/			
Duru-Bellat & Mingat (1989)					/
Slavin (1992)		/			
(Okagaki & Sternberg, 1993)			0		/
Siegle (1992)				/	
Gentry (1999)			1		
Tomlinson (2001)				/	
Benson (2002)			/		
Gamoran (2002)				/	
Shields (2002)	/				
Burns & Mason (2002)			/		
Pare (2004)				/	
Reay, David, & Ball (2005)					/
(Oakes, 2005)		/			
Bellin, Dunge, & Gunzenhauser (2006)	9		/		
Burris, Heubert, & Levin (2006)	/				
Zeng, Xu, Shi, Wang, & Wu (2007)	/				
Zimmermann & Raedt (2009)			/		
Preckel, Götz, & Frenzel (2010)				/	
Bian (2010)			/		
Trivedi, Pardos,			/		
(2011)					
Lam, Ho, & Wong, (2012).					/
Caldwell (2012)			/		
Carrol Corbett Burris & Allison (2013)			/		

 Table 2.4 : Summary of Student's Academic Streaming Approaches

#### 2.7 ACADEMIC STREAMING APPROACHES APPLIED IN SCHOOL

Academic streaming is applied in a few countries all over the world. The main objective is to determine an appropriate academic streaming for the students. As per discussed in Section 2.6, there are five approaches used to stream students; parental demand, heterogeneous, homogenous, ability and cluster. In order study more deeply on the implementation of the academic streaming, this research used United Nations (UN) reports on Human Development Index (HDI). Based on the report, countries with a very high HDI ranks are identified, and the way they stream their student also be reviewed.

The HDI is a tool developed by the UN to measure the status of social and economic development in particular countries based on four standards: per capita gross national income, the mean years of schooling, expected years of schooling and life expectancy at birth. The HDI ranking is used to identify changes in the level of development from time to time and to compare the level of development in different countries (United Nations, 2013). The Table 2.5 shows 35 countries that were categorised as having a very high index of human capital development (United Nations, 2011).

	Advance Economic Country	Human Development Index (HDI)		Advance Economic Country	Human Development Index (HDI)
¥	Australia	2		Luxembourg	25
	Austria	19	÷	Malta	26
	Belgium	18		Netherlands	3
÷	Canada	6	÷÷ ¥	New Zealand	5
¥	Cyprus	31		Norway	1
	Czech Republic	27	۲	Portugal	41
	Denmark	16	ø	San Marino	-
	Estonia	34	0	Singapore	26
Ŧ	Finland	22		Slovakia	35
	France	20	-	Slovenia	21
	Germany	9		South Korea	15
Ë	Greece	29	ė	Spain	23
\$	Hong Kong	13	╇	Sweden	10
+-	Iceland	14	÷	Switzerland	11
	Ireland	7	B	Taiwan	-
٥	Israel	17	NK ZK	United Kingdom	28
	Italy	24		United States	4
•	Japan	12			

Table 2.5 : List of "A very high index" of Human Capital Development Countries

Among the 35 countries analyzed, four countries were found to have the top ten HDI ranking over the past 10 years. The countries were Japan (2004-2008), Canada (2003-2013), Norway (2003-2013) and Iceland (2003–2007). These high HDI countries were implementing an organised and well-planned education system. They also have the best education systems in the world (The Organisation for Economic Co-operation and Development, 2015).

In Norway, their education system is an individual-based system, whereby each student is identified with a unique ability. The students are organized according to their ability after year 12 or 13 (Braathe, 2009). Furthermore, their academic streaming strategy enforces 'education for all' as a part of human rights in Norway. This means that every single human being has a right to an education in accordance with his or her needs and interests, abilities and aptitudes. The education policy in Norway focuses on meeting both the short-term and long-term qualification requirements of the labour market, the society and the individual (Norwegian Ministry of Education and Research, 2007). The policy emphasises on a general and broad initial education, leaving specialization to later stages and further training at work. This means that the education process continues and will not finish until students get into a real working environment. This is due to the fact that there is no end in the learning process itself. However, emphasis on such requirements should be seriously examined from the very beginning and improvements must be prepared in-line with the economic and technological development.

The same concept took place in Japan. Japan's education policy emphasises on individual performance and transition to a lifelong learning society (Kashihara & Institut, 2008). It was coping with numerous changes that included internationalization in various sectors of society, and the spread of modern information media consistent with current technology. The desire for a child's academic abilities to use information was enhanced using digital technologies. The effects were objectively measured through the implementation of its educational strategy. The education system in Japan focuses on four aspects (Kashihara & Institut, 2008):

- 1. Transmitting cognitive knowledge;
- 2. Socializing and acculturating;
- 3. Selecting and differentiating young people;
- 4. Legitimizing what they teach.

Japan uses the cognitive learning theory that highlights a few aspects such as awareness, perception, reasoning and judgment through understanding, thinking and intuition.

Students are streamed based on ability. They are placed among similar abilities in a classroom. This group has minimal range of differences of abilities if compared to heterogeneous classroom (Kariya & Rosenbaum, 2014; Kashihara & Institut, 2008). All gifted students within the same grade level will be in the same classroom. Learning methodology gives priority to individual performance; whereby, each individual will have the same experience but different observations. They also emphasise knowledge sharing to ensure that every individual with the same skills can meet the standards and requirements as an expertise who can contribute to economic development (Kariya & Rosenbaum, 2014; Westney, 1999). In addition, the syllabus in universities and other educational institutions are being expanded and improved in terms of information, digital infrastructure, remote education, etc. Furthermore, stable and on-going structures will be established to prevent the occurrence of mismatches with respect to highly-skilled human resources.

Next, look at the education system in Canada. Canada aims to provide job opportunities to their people. So, the academic streaming of high school students done through different sequences of core courses (e.g., English, science, and mathematics) or ability has been practiced in Canada. Students with different abilities and aptitudes should have the opportunity to take different sequences of courses while completing high school. Those most academically inclined should have the opportunity to take the most advanced courses, which will challenge them (Edgerton, Roberts, & Peter, 2013; Klassen et al., 2008). Secondary schools may specialise in academic or vocational streams or the arts; all include some kind of 'streaming' system that's designed to prepare students for a vocational or community college or university. Their system uses

four pillars of innovation (Pollock, 2008) as shown in Figure 1. The four pillars are: leadership, culture & climates, structures & process and people (HR capacity).

Each pillar is interconnected and relies on the other, which Watt (2002) considered as an innovative progress emerging in schools. Canadians believe the need for fresh innovation to stay ahead so as to meet the challenges that support rapid economic growth. Transformation of the economic atmosphere, as well as the rapid development and extension in information technology (IT), have made it increasingly obligatory for Canadian schools to go for innovation as they prepare their students for an up-to-date world (Gibson, 1999; Government of Canada, 1995). This is because Canada wants to provide suitable skills for their people; "The skills that graduates need to enter, stay in and progress in the world of work" (Gibson, 1999). Moreover, they applied the School Strategic Plan to the Canadian Lead Primary School Grampians region 2012-2015, which aims to physically and mentally build up each individual student in a supportive and optimistic environment that focuses on individual needs and student-cantered learning approaches (Department of Education and Early Childhood Development, 2012). It also focuses on student learning, student engagement and wellbeing and student pathways and transitions.

Lastly, education in Iceland has traditionally been organised within the public sector, and there are very few private institutions in the school system (Ministry of Education Science and Culture, 2002; Review & Education, 2005). Almost all private schools receive public funding. A basic theory of Iceland education policy aims to provide individuals with the knowledge, skills and competences needed to participate effectively in society and in the economy. This means that everyone should have equal chances based on their ability to get an education, without considering gender, economic status, religion, residential position and family background (Mikiewicz, Jonasson, Gudmundsson, Blondal, & Korczewska, 2011; Review & Education, 2005). Moreover, they intend to make all levels of education as comprehensive as possible through their education system so that pupils' paths through the system are transparent and direct as possible; although, there is always an emphasis on flexibility. Furthermore, the main idea behind Iceland's educational strategy, which stresses on compulsory schooling, is to train students for life, work in a constantly developing country by individually advising them appropriate chances to examine and develop and train them for employment through specialised studies.

Based on the review above, academic streaming in school for these four countries (Norway, Japan, Canada and Ireland) reviewed in Table 2.6.

 Table 2.6 : Education System Policy, Education System, Educational Strategy and

 Academic Streaming Strategy in Norway, Japan, Canada and Iceland

Country	Education System Policy	Education System	Education Strategy	Academic Streaming Strategy
Norway	Long-term and short- term qualification requirements of the labor market, society and individual	An individual-based system	Education for all	Ability
Japan	Emphasis on individual	Focuses on cognitive knowledge	Using digital technologies	Ability
Canada	Aims to provide job opportunities	Four pillars of innovation	School Strategic Planfor Canadian LeadPrimarySchoolGrampiansRegion2012-2015	Ability
Iceland	Individual knowledge	Student's path through the system is flexible	Meet the minimum requirements of the job market	Ability

The review shows that all four countries established their academic streaming based on ability or performance, considering market needs as the basis for human capital development in order to support economic growth. It shows that the countries are preparing their students to meet the minimum requirements of the job market. In addition, their education systems have a plan that is in-line with the needs of the country. For example, Norway, Canada and Iceland adopted an academic streaming that puts the needs of the industry as an important factor, to be taken into account, in developing the structure of their plan. With these efforts, adequate human resources are produced and human capital requirements, which contribute to economic development, are met.

In order to increase the human capital values, education functions as a deliberate effort in contributing to create a balanced human capital in order to accelerate the country's progress, as well as, the technology used. Therefore, education needs a systematic academic streaming for students to plan the student's direction; especially the progression of students' performance. This will ensure that planning can be done in terms of creating human capital that is in-line with market requirements. The system will use the ability or performance of an individual, thus students' academic pathway can be planned properly to meet the skills demanded and fulfilling the needs of the economy.

#### 2.8 DATA MINING TECHNIQUES IN EDUCATION

Data mining is a systematic data analysis tool to ascertain the unidentified, valid patterns and connections in large data sets. This tool includes mathematical algorithms,

statistical models and machine learning methods (Cristobal Romero & Ventura, 2013; Seifert, 2004). The data mining contains a set of techniques that used to extract a significant and interesting knowledge from educational These techniques can be categorized into summarization, classification, clustering and association (Wu, Zhu, Wu, & Ding, 2014).

#### 2.8.1 Summarization

Summarization is a process of identifying and extracting valuable information from the data to present the key ideas (Crangle, 2002). It can reduce a specified set of transactions into a smaller set of patterns with the same information (Chen & Guo, 2015). Moreover, it is suitable for unstructured text data in to discover unpredicted information that is crucial to an organization (Crangle, 2002).

Examples of summarization task are text mining and tabulating. Text mining is a method for the unstructured domain which group documents that share similar features (Alfred, 2008; Crangle, 2002) while tabulating is mean and standard deviations that are often applied for data analysis, data visualization and computerized report (Chen & Guo, 2015).

Furthermore, summarization technique can be applied to visualization. It will reduce the size and complexity of large multidimensional datasets to more manageable proportions. The summarization can highlight the relevant aspects of the data clearly, leading to coherent visualizations and facilitate an accurate visual analysis (Heer, Shneiderman, & Park, 2012; Kehrer & Hauser, 2013; Kocherlakota & Healey, 2005).

Summarization is implemented using several techniques, which are data classification, association rule mining, data aggregation and outlier detection (Alfred, 2008; Chandola & Kumar, 2007; Crangle, 2002). However, these techniques are may not be incompatible to deal with high dimensionality data (Cordeiro, Faloutsos, & Traina Júnior, 2013; Kocherlakota & Healey, 2005; Zimek, Schubert, & Kriegel, 2012).

#### 2.8.2 Classification

Classification technique is one of the supervised learning techniques that classify the data into a predefined class (Al-Radaideh & Nagi, 2012). Classification technique is commonly used to build models that are used to predict future data trends.

There are several algorithms of data classification such as decision tree and Naïve Bayes classifiers. Classification technique has been applied to extract rules and predict certain activities in education such as new model for improving the efficiency and effectiveness on higher educational process (Delavari, Shirazi, & Beikzadeh, 2004; Chopra & Lal, 2013), predicting the academic performance of undergraduate and postgraduate students (Nghe, Janecek, & Haddawy, 2007), develop a more meaning of data (Zengin, Esgi, Erginer, & Aksoy, 2011) analyze system using knowledge discovery in higher educational databases (Guruler, Istanbullu, & Karahasan, 2010) and predict university students' performance (Al-Radaideh, Al-Shawakfa, & Al-Najjar, 2006; A. Wahbeh & Al-Radaideh, 2011; A. H. Wahbeh, Al-radaideh, Al-kabi, & Al-shawakfa, 2010).

#### 2.8.3 Association

The association rule mining technique is dedicated to discover patterns and dependencies in datasets (Kocherlakota & Healey, 2005). Association rule can be described in the form of, IF *antecedent* THEN (likely) *consequent*. The antecedent and consequent are denoted as the item sets (Han & Kamber, 2001; Tomida, 2015).

A classic application of association rules is the market basket analysis (Han & Kamber, 2001). It analyzes and associates the items purchased by customers. Other applications of the association rule mining education are to categorize students' failure patterns (Chandra & Nandhini, 2010) and to predict their final achievement status upon graduation based continuous assessment and examination scores (Ogor, 2007).

However, one of the problems of mining association rules is to determine all association rules that have support and confidence greater than or equal to the user identified minimum support (*minsup*) and minimum confidence (*minconf*).

#### 2.8.4 Clustering

Clustering technique is used to group data of the same features and a popular technique for data mining (Bachtiar, Eric, & Kamei, 2011; Lopez & Luna, 2012; Osmanbegović & Suljić, 2012; Oyelade, Oladipupo, & Obagbuwa, 2010; Shovon et al., 2012; Tair & El-halees, 2012; Zimmermann & Raedt, 2009).

Using clustering technique, data can be divided into natural groups and summarized the students' learning pattern and support for teaching and learning target (Hämäläinen et al., 2013). Many of clustering algorithms have been proposed in data mining (Jain, 2010) and the most common algorithm used by researchers is K-mean clustering algorithm (Fathian & Amiri, 2007; Firouzi, Sadeghi, & Niknam, 2010; Kanungo et al., 2002; Kao, Zahara, & Kao, 2008; Krishna & Murty, 1999; Laszlo & Mukherjee, 2007; Maulik & Bandyopadhyay, 2000; Niknam & Amiri, 2010; M. Yin, Hu, Yang, Li, & Gu, 2011).

#### 2.8.5 Summary of Data Mining Techniques in Education

The data mining techniques are widely used in education to analyze the student data. It helps to ensure that the knowledge that cannot be seen can be obtained from a large data. Table 2.7 summarized the purpose of data mining techniques in the education. From the previous studies discussed, it appears that the appropriate technique used to streaming students based on their performance is clustering technique. This is because the clustering technique discovers natural groupings of data sets or objects that have same features. This technique created several groups from the data that have the same characteristics.

## Table 2.7 : Data Mining Techniques in Education and Appropriateness of Its Use for Academic Streaming

Data mining techniques	Objective	Suitable For Academic Streaming Purposes?
Summarization	identifying and extracting valuable information from	No
	data repositories to present the key ideas	
Classification	classify data item into predefined class label	No
Association	discovery of patterns and dependencies in datasets	No
Clustering	discover natural groupings of objects that have same	Yes
_	features	
The clustering technique can be used for grouping students based on their same performance. In this study, the clustering technique is expected to divide the students into two groups, which are mainstream of science and art based on their academic performance for the whole one year.

## 2.9 CLUSTERING TECHNIQUE FOR EDUCATIONAL ENVIRONMENT

In the educational environment, K-means algorithm is one of the clustering technique that most frequently used for data mining (Shovon et al., 2012). Oyelade, Oladipupo, & Obagbuwa (2010) used K-means algorithm to monitor the progression and predict students' academic performance. This unsupervised algorithm analyzed the students' results and assigned their scores into groups according to their performance using Euclidean distance. Differently, Bian (2010) applied K-means to cluster the students' learning activity. In this study, the K-mean searches for compact clusters embedded within subsets of features and effective with high dimensional datasets. Since not all students experience similar difficulty in all activities, subspace clustering appears to be beneficial.

Neural Networks (NN) is an important data mining tool used for clustering. The NN consists of several layers of neurons, namely Input Layer, Hidden Layer and Output Layer. The neuron can be trained in many ways by adjusting the weight of each neuron according to the presented input and error of the network. One of the most commonly used algorithms for training these neurons is Back Propagation (BP) algorithm. BP adjusted the weight of the neurons by matching the input and output data. Bachtiar, Eric, & Kamei (2011) groups the student based their English abilities using NN. They

demonstrated fairly low errors for general grouping and ability grouping for Listening, Reading, Speaking, and Reading.

In the study performed by Lopez & Luna (2012), the Expectation-maximization (EM) clustering algorithm is used to predict the final marks. A total of six techniques in clustering are compared, which are EM, Farthest First, Hierarchical clustering, sequential information bottleneck (sIB), Simple K-Means, and X-means for two cases. Among the six techniques, EM algorithm obtained the highest accuracy rate in both cases (A and B) and the best overall accuracy (89.4%). The other clustering algorithms obtained a lower accuracy rate between 50% and 70%.

In Dominguez, Yacef, & Curran (2010), k-means clustering, association rules and numerical analysis are used to generate hints for students who are completing programming exercises.

The study involves several steps:

- Group the students based on their abilities
- Clustered the questions with two distinct aims: find the questions that were similar to each other and group the questions by difficulty.
- Determined the association rules to indicate the arrangement of topics that should be followed
- Apply a simple numerical analysis to determine the frequencies and averages for certain aspects of the data (for example, the total number of submissions made by the students who failed the question)

Another study associated to the clustering technique in education is performed by Hämäläinen et al., (2013) to evaluate a variety of clustering methods from the perspective of clustering student data. They analyzed the main approaches to clustering and investigated the models if the assumptions fit typical student data. The study suggests that the clustering methods performed differently according to the situations and type of data.

Nghe, Janecek, & Haddawy (2007) presented a study to predict the academic performance for undergraduate and postgraduate students at two different academic institutes. This study compared the accuracy between Decision Tree and Bayesian Network algorithms on the same academic performance data. The finding shows that the Decision Tree was consistently 3-12% more accurate than the Bayesian Network.

Sembiring, Zarlis, Hartama, & Wani (2011) has proposed the combination of Kernel kmeans clustering techniques and Smooth Support Vector Machine (SSVM) classification to group the students according to their behavior and performance. Through this study, the Data Mining Techniques (DMT) can provide an effective improving tool for student performance particularly in higher education to predict the final performance of the students.

The graduate students' performance can be improved and overcome the problem of low grades using four techniques, which are Association, Classification (Rule Induction and Naïve Bayesian classifier), Clustering (k-means) and Outlier detection rules (Distance-based Approach and Density-Based Approach) as suggested by Tair & El-halees

(2012). Each one of these tasks can be used to improve the performance of graduate student:

- Association rules searches for interesting relationships among items in a given data set.
- Classification predict the Grade of the graduate student (Excellent, Very good, Good, and Average).
- Clustering find high-quality clusters.
- Outlier analysis detect the outliers in the student data (students with learning problems).

At the end of this study, the knowledge can be extracted and described their importance in the educational domain for each tasks. Furthermore, data mining are useful and can be used in higher education particularly to improve graduate students' performance.

Generally, clustering in educational environment can be categorized into few objectives as follows:

- Performance prediction: Predict the student's performance based on their previous achievement
- Grouping : group students based on certain features
- Generate hints: generate hints based on the activities
- Method evaluation: evaluate the methods that have been used for certain purposes in the educational environment
- Performance improvement: improve the student's performance based on their achievement

All the related studies to the clustering techniques in educational environment are summarized in Table 2.6. Mainly these studies are for prediction or plan for improvement to produce excellent students.

No	Author	Idea	Algorithm	Technique	Objective
1	(Zhang, Ramakrishnan, & Livny, 1996)	To produce the best quality clustering with the available resources	Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH)	Demonstrates that BIRCH suitable for very large databases	Methods evaluation
2	(Nghe et al., 2007)	Predicting the academic performance of undergraduate and postgraduate students	Decision Tree and Bayesian Network	Compares the accuracy of Decision Tree and Bayesian Network algorithms	Prediction
3	(Oyelade et al., 2010)	Prediction of Students' Academic Performance	K-means	Analyzing students' results based on k- means clustering analysis and uses standard statistical algorithms (Euclidean distance)	Prediction
4	(Bian, 2010)	Clustering Student Learning Activity	K-means (Subspace clustering)	Searches for compact clusters embedded within subsets of feature.	Grouping
5	(Dominguez et al., 2010)	To generate hints for students	k-means clustering, association rules and numerical analysis	Group students based on their abilities and clustered the questions with two distinct aims. Find association rules and apply a simple numerical analysis	Generate hints
6	(Bachtiar et al., 2011)	Grouping rate on students' English abilities for Listening, Reading, Speaking, and Reading.	Neural network (back propagation algorithm)	Back Propagation (BP) based on the principle of gradient decent learning.	Grouping
7	(Sembiring et al., 2011)	To develop the model of student performance predictors.	Smooth Support Vector Machine (SSVM) classification and Kernel k-means clustering techniques.	SSVM - predicate students final grade Kernel k-means - to grouping students according to their behavior and performance	Prediction
8	(Lopez & Luna, 2012)	To predict the final marks	Expectation-maximization (EM) clustering algorithm	Test the accuracy 6 techniques in clustering	Prediction
9	(Tair & El-halees, 2012)	To improve graduate students' performance, and overcome the problem of low grades of graduate students	Association, Classification (Rule Induction and Naïve Bayesian classifier), Clustering (k-means) and Outlier detection rules (Distance-based Approach and Density-Based Approach)	Extract useful knowledge from graduate students and applied all the four data mining techniques.	Improve performance
10	(Hämäläinen et al., 2013)	To evaluate a variety of clustering methods	Introduce the main approaches for clustering	Analyze the main approaches to clustering and see how useful models they produce and how well their underlying assumptions fit typical student data	Methods evaluation

# Table 2.8 : Clustering Techniques That Used in Educational Data

### 2.10 RELEVANT EDUCATIONAL MODELS

Models are developed based on theories and used for analyzing, understanding, describing, explaining, predicting, design and action (Gregor, 2002). In this section, the relevant educational models for the academic streaming are presented.

### 2.9.1 John Carroll's Model

Most existing models that categorize the factors or justifications on educational processes today are originally from Carroll (1963) in a seminal article defining the major factors related to school learning. Klausmeier & Goodwin in 1971 describes that Carroll specialized in language and learning, the study that relating to words and their definition to the cognitive concepts and factors, which they build.

A simple equation for Carroll's model (Carroll, 1989) is:

School Learning = f (time spent/time needed).

In his model, the most important variable in school learning is time.

The model describes that the time spent is the effect of opportunity and perseverance. Opportunity in Carroll's model is defined by the classroom teacher, which is time allocated for learning session that is set up by classroom teachers. Perseverance is the student's participation with academic content during the allocated time. The perseverance can be measured through the percentage of the allocated time that students are involved in the learning process and labeled as engagement rate. The factor proposed in this model acquired by means of allotted or allocated time multiplied by engagement rate. It is a measurement of time spent that known as engaged time or timeon-task.

This model suggested that the time to learn academic content, the capability to understand the teaching content, and the quality of learning is depending upon aptitude. To predict the school learning, detailed of teacher and student actions, and student characteristics are the only variables required. However, this model excludes the influences of family, community, society and the other factors.

The Mastery Learning model proposed by Bloom (1971) adopted the same ideology as Carroll's. Bloom discovered that in the traditional schooling, students' mastery of the basic skills is a better predictor of school learning if time is not set constant for all students. In this case, almost all students can excel in the class if:

- Students are given sufficient time to learn basic information taught in school
- Students are delivered with quality teaching.

Bloom described the quality teaching as the teachers' ability to:

- Manage all subject materials into manageable learning subjects,
- Detail learning objectives for each subject,
- Develop a proper formative and summative assessment measures
- Design and implement cluster teaching plans, with time allocations, exercises, and educative instruction for all students to improve their skill.

### 2.9.2 Proctor's Model

Proctor (1984) expands the studies of good teaching and features that contributed to this quality by adding the important factors of teacher and student behaviors as predictors of student performance as depicted in Figure 2.2. This model redesigned the teacher and classroom based models with additional features that highlight teacher expectations.



Figure 2.2: Proctor's Model (Proctor, 1984)

Based on the study by Rosenthal & Jacobson (1968), the Proctor's model described that the environment of a school can have an effect on the achievement of its students. A transformation in student achievement can be made by relates it with good attitudes, rules, and the values of a school and teacher. The Proctor's model is influenced by Rosenthal & Jacobson (1968) model in a context of social environment but not a teacher/student relationship. Most of the literatures support Proctor's (1984) theory that student self-esteem and characteristics are influenced by teacher efficacy (Colquitta, Pritcharda, & McColluma, 2011; McIlrath & Huitt, 1995; Waterman, 2014; Weinstein, Madison, & Kuklinski, 1995).

The Proctor's (1984) model creates the factor of the School's Social Climate that includes several variables such as the attitudes, norms, beliefs, and prejudices. Further, the school climate is influenced by factors of student characteristics (race, gender, economic level, and past academic performance) and people involved in the schooling practice (administrators, teachers and students). The intermediate outcomes of student learning and student self-expectation will increase if the instructional used, feedback, and communication among students, parents, and teachers are good. If the student selfexpectation is affected by negative attitudes of instructors and administrators, their academic performance will decrease.

The connections among factors in the Proctor's (1984) model consist of school's rule on permitting time or teaching hour for students to learn. This includes quality of teachers or instruction as mentioned in Carroll's (1963) model or teacher behaviors as proposed in Cruickshank (1986) model.

Finally, the student's performance level in Proctor's (1984) model is a result of the combinations of various factors and repeated connection among these factors. In Proctor's model, the academic performances in a classroom during a specific school session are influenced by the social environment of the school and the improvement can be made by modifying the factors.

### 2.9.3 Cruickshank's Model

This model was developed by Cruickshank (1986) and influenced by Biddle, Flanders, and Mitzel models. However, the Cruickshank emphasized the model on the classroom and teacher-based. Figure 2.3 shows the Mitzel model is adopted by the Cruickshank model as referenced in the concept of categorizing factors as "product, process, or presage" (Cruickshank, 1986).



Figure 2.3 : The Cruickshank's Model (Cruickshank, 1986)

The details of the categories are:

- Product: the student, based on their learning part (change in performance or performance possible)
- Process: includes communication between student and teacher.
- Presage: the teacher characteristics such as intellectual, level of knowledge, achievement and other that related.

The variable of "teacher and student classroom talk" is proposed by Flanders (1965) and developed an instrument concentrated on this variable. It allowed the observation of verbal influence by the teacher defined as 'teacher talk' and 'student talk' in a variety of classroom situations. Cruickshank placed them all together and added additional presage variables such as student characteristics, properties (abilities and attitudes) and school, community and classroom climate. Theoretically, presage can influence process, and the process will influence the product.

### 2.9.4 Biddle Model

The Biddle model (B. Biddle & Ellena, 1964) depicted in Figure 2.4 showed a link between a specific learning activities and the impact from the teacher. This model provides the foundation for Cruickshank's (1986) model and suggests seven categories of variables associated with schooling and student performance: school and community contents, formative experiences, classroom situations, teacher properties, teacher behaviors, intermediate effects, and long-term consequences.



Figure 2.4 : The Biddle Model (B. Biddle & Ellena, 1964)

Biddle model is based on the transactional process of the classroom by analyzing the structure and function of the communication process, which is the interaction between teacher and pupil classroom behavior. This model gives a solution to the issue on the amount of attention each student gets from the teacher, or level of teacher understanding on their student's problem (Adams & Biddle, 1970). The model also explained the non-cognitive variables that influenced the emotional domain such as self-concept and self-esteem of the students. Examples of these variables are teacher genuineness, teacher-offered conditions of respect and modification of low self-concept.

### 2.9.5 Gage and Berliner's Model

The Gage and Berliner's model is an instructional process model that stressed on the role of a teacher or a "classroom and teacher based". As shown in Figure 2.5, this model concentrates on variables that must be measured by the teacher, since the teacher is responsible to plan and delivers the instruction to students. Furthermore, it defined the "quality instruction" in detailed and offered five tasks linked to the instruction/learning process.



Figure 2.5 : Gage and Berliner's Model (N. L. Gage & Berliner, 1998)

A role of a teacher begins by outlining the objective and end with an assessment. Instruction links between objective and assessment based on the teacher's understanding of the students' characteristics and how to encourage them. If the assessment showed that the desired results have not been reached as per planned, the teacher will redeliver the material and starts the process from the beginning. The classroom management is incorporated under the rubric of inspiring students. So, it is recommended that the teacher used the study and ideologies from educational psychology to improve the teaching procedures to get the best outcomes.

### 2.9.6 Huitt Model

The teaching and learning theory produced by Huitt (1995) categorized the major groups of variables that affected the school or student achievement to address the issue of the diverse amount of effort are observed between student in learning process. This model incorporates the basic influences such as school, classroom, teacher, student, and additional contextual influence as shown in Figure 2.6.



Figure 2.6 : The Teaching / Learning Theory (TLT) in Huitt Model

The model is a modification from (Squires, Huitt, & Segars, 1983) which exclusively concentrated on school and classroom activities. The expected school performances are measured on regular tests of basic skills and the redefined of the Academic Learning Time is an important addition in this early model.

It had been documented that Carroll's conceptualization of the time consumed, measured the amount of time in academics, but lacking the quality of that time. Fisher, Marliave, & Filby (1979) had added the new idea of achievement as a key factor of quality of time consumed and created the Academic Learning Time (ALT) which they described as "involved in academic learning at a high success rate". Besides that, Brady, Clinton, Sweeney, Peterson, & Poynor (1977) added another quality component called content overlap, the content that covered in the classroom overlaps to content tested. However, Squires et al., (1983) referred to the more comprehensive description of ALT suggested by (Caldwell, Huitt, & Graeber (1982), which is "the amount of time students are successfully engaged on content that will be tested".

Huitt's (1995) adds the factors associated with the context of student and teacher characteristics. Parts of these factors are from the Proctor (1984) model and Cruickshank (1985) model. This collaborating model suggested that important factors must be measured because our society is rapidly changing from an industrial base to a knowledge base.

The Huitt's model presents a relationship among the categories of Context (family, home, school, and community environments), Input (what students and teachers bring to the classroom process), Classroom Processes (what is going on in the classroom)

and Output (measures of learning done outside of the classroom). These categories appear superimposed on the model since it is essentially intertwined in the learning process. The Input and Output representing the start and end of the teaching/learning process.

The model suggested that teachers must first identify the outcome of the learning process for an effective assessment that probably influenced the variety of important predictor or factors. Once the outcome is identified, those variables or factors that caused variability or inconsistency in measurement can be explained. Altering a specific objective may modify the variables of predictor or factors. Thus, identification of the Output is the most important part of the Huitt's model.

The most direct impact on factors of school learning is Classroom Processes. This category includes three subcategories, which are Teacher Behavior and Student Behavior, and Other. The Other subcategory consists of factors such as classroom climate and student leadership roles. The category of Teacher Behavior consists of planning, management, and instruction. Commonly, planning has a little predictable connection to the student performance while management and instruction have a moderate connection to achievement (Gage & Berliner, 1992). The three factors such as teachers providing corrective feedback, teachers' use of reinforcement, and level of student-teacher interaction appear to be the best predictors of student success (Rosenshine & Stevens, 1986).

In Huitt's (1995) model, the activities in a classroom along the teaching/learning process are considered as Input. The subcategory of Teacher Characteristics includes such variables as values and beliefs knowledge of students and the teaching/learning process; thinking, communication and performance skills; and personality. While each of these variables are significant to the classroom environment, teacher efficiency is the best predictors of student success from this subcategory (Ashton, 1984; Proctor, 1984). The second subcategory of Input is Student Characteristics that explained the relationship between student and the teaching/learning process and their outcome. Study Habits; Learning Style; Age; Sex/Gender; Race/Ethnicity; Motivation; and Moral, Socio-emotional, Cognitive, and Character Development all become important in the correlation of classroom behavior and school achievement (Huitt, 1995).

Finally, Context includes subcategories such as School Processes and Characteristics, Family, Community, State and Federal Government, TV/Movies, and the Global Environment. All of these variables are important and influence variables in the other three major categories, probably the two most important are Family and the Global Environment.

### 2.9.7 Huitt Revised Model

After come up with teaching and learning theory (1995), Huitt create new model that considers the development of the whole person from the perspective of developing capacities in nine domains (Huitt, 2011) as per shows in Figure 2.7: (1) temperament, personality. self-views; (2) cognition/thinking; (3) affect/emotion; and (4) physical/bodily-kinesthetic; conation/volition (or self-regulation); (5) (6) social/interpersonal; (7) spiritual/ transpersonal; (8) moral character; and (9)citizenship. It focuses on improving scores on standardized tests of basic skill.

Perhaps the most widely accepted approach to identifying a variety of human capacities work on multiple intelligences. Seven Intelligences identified. Three of the intelligences have been labeled Symbol Analytic in that they involve making a conversion from a symbol to a higher-level mental code (i.e., linguistic-translating letters and words into knowledge and concepts; logical-mathematical-converting numbers to quantitative concepts and to think rationally and/or logically; musical-translating written musical symbols in timbre, pitch, and rhythm). Two of the intelligences are considered Personal intelligences in that they are oriented to the person (intrapersonal-knowledge of one's self and interpersonal — knowledge of others, especially their moods and motivations). Finally, two additional intelligences might be considered as Object-oriented intelligences (spatial-the ability to mentally rotate an object in space and bodilykinesthetic-the ability to control one's body and handle objects skillfully). In the process of investigating the concept of multiple intelligences, eight domains were identified separately through previous research: (1) temperament, personality, and selfviews; (2) cognition/thinking; (3) affect/emotion; (4) conation/volition (or selfregulation); (5) physical/bodily-kinesthetic; (6) social/interpersonal; (7) spiritual/ transpersonal; and (8) moral character. Interestingly, five of these relate to terms used by the ancient Greeks to describe different aspects of a human being: body (bodilykinesthetic), mind (cognition/thinking, affect/emotion, and conation/volition), and soul/spirit. Research also identified human capacities for social interaction, morality, and the creation of self-views. A ninth dimension, citizenship, was added as that is the role within which these capacities will be used.



**Figure 2.7 : Huitt Revised Model** 

The domains might be considered as points on a star with the self and self-views central to the image and the concept of local-global-cosmic citizenship used as the amalgamation of the adult roles in which individuals engage in the modern world (see Figure 2.7). This is a revision of the Becoming a Brilliant Star framework previously developed (Huitt, 2006a). It shows that the domains are developed within a multi-level sociocultural context where family, school, religious organizations, and friends, as well as the connections among them, provide the most direct influence on one's development, but that development is also influenced by larger social and cultural forces.

### 2.9.8 System Theory Model

Systems theory or ST (Bertalanffy, 1969) has been proposed as a potential overarching framework for dealing with many issues in human behavior. Contributors to systems theory have originated from diverse fields, including physics (Capra, 1982), biology, anthropology and psychology (Bateson, 1979). The work on living systems by D. Ford (1987) and M. Ford and D. Ford (1987) has served to develop an integrated framework of human development and has furthered the development and understanding of systems theory. Developmental Systems Theory (DST, D. Ford & Lerner, 1992) and Motivational Systems Theory (MST; M. Ford, 1992) have illustrated the applicability of systems theory principles to human behavior.

The foundation of the Systems Theory Framework (STF; Patton & McMahon, 1998) is based on the System Theory (Bertalanffy, 1969). The theory is construed as an overarching framework, which all concepts related to pathway described in the plethora of career theories can be usefully positioned and utilized in theory and practice.

The STF provides a map for understanding the origins of student learning pathway and the dilemma it is now facing. Student learning pathway is a unique process built on a foundation of student's development. Traditionally, the student learning pathway has tended to focus on specific concepts relevant to individual performance in the final examination only. By only focusing on one aspect relevant to pathway decision-making, others academic performances for the whole one year are inevitable to be undervalued or ignored.



Figure 2.8 : The Systems Theory Framework (STF)

As shown in Figure 2.8, fundamental to the STF is the individual system, which portrayed a range of intrapersonal factors on individual development, such as personality, ability, gender, and sexual orientation. The individual system is linked to the factors that include the individual's social system as well as the broader environmental/social system as they not live in isolation. While the impact of many factors, such as geographic location and political decisions on academic pathway is less well understood within the theoretical literature, their influence is profound to see the importance.

The STF viewed the academic development or pathway as a dynamic process that represented through process of factors, recursiveness, changes over time and chance (see Figure 2.7). Essentially, the STF is the idea that individual system is an open system by focusing on stimulus from the outside. Such contact is called recursiveness in the STF, which in illustrative method is represented by broken lines that symbolized the permeability of the limitations of each system. It is well recognized that influences on an individual may change over time. The final process of factors and chance is illustrated on the STF diagram as lightning blazes, reflecting a better acknowledgement of the part chance plays in academic development or pathway. In all systems, the factors involved are located within the perspective of time - past, present and future - all of which are correlated; past influences by the present and also future influence by the past and the present.

### 2.9.9 Summary of the Discussed Models

All the models discussed in sub-section 2.9.1 to 2.9.8 finds the important factors relevant to the school for teaching and learning process, as well as contributing an important information on factors that influence the academic pathway for the students. These models are summarized in Table 2.9. Over a period of years, beginning with Carroll (1963) until Huitt (2011) (at least as far as this review is concerned), the models have been examined, reviewed, revised and modified to fit into today's modern society. In these models, teachers and school systems, families, communities and entire countries have an influenced on students' academic pathway in school. None of the variables appear to be so influential that we need only pay attention to that particular factor in order to produce the kinds of students we need. For example, an individual teacher can play a role in determining the direction of the student (as seen in Cruickshank's 1985 model), but can predict the institution's performance as well (as seen in Proctor's 1984 model). Or, the school may successfully develop students'

academic skills, but students could still not be successful in life because erroneously in determining the appropriate fields with them (Whetzel, 1992).

This study using Huitt model and System Theory model. These two models have been used for this study because both models have a few elements that are related to student's achievement and student development or planning. The elemets in the models also found in the previous study as discussed in Section 2.5. While the other model is more focused on teaching and learning. The Huitt model is based on Teaching and Learning Theory (TLT) which is also originated from System Theory (ST). These two models categorizes the major groups of factors that have been related to the student's performance and influenced the academic streaming. Understanding all the factors and their relationships among each other that influenced the student academic development and pathway may be more than we can imagine. Researchers may never completely understand the significance of the whole practice, but can make every effort to understand the practice as much as possible, and tried to identify factors that are give an effect to choose a suitable academic pathway. Besides, by referring to the models, important factors within a group or subgroup can be identified and various factors across the models can be took into account.

Model	Idea	Area
John Carroll's Model	Time spent is the effect of	Learning
	opportunity and perseverance.	
Proctor's Model	teacher and student behaviors as	Teaching and learning
	predictors of student	
	performance	
Cruickshank's Model	Categorizing factors as "product,	Teaching and learning
	process, or presage"	
Biddle Model	Seven categories of variables	Teaching and learning
	related to schooling and student	
	performance	
Gage and Berliner's Model	An instructional process model.	Teaching
Huitt Model	Categorizes the major groups of	Teaching and learning, human
	variables that have been related	behaviour
	to school or students	
	achievement	
Huit Revised Model	Student's development in	Teaching and learning
	classroom environment	
System Theory Model	Students development / planning	Human behavior

Table 2.9: Models Related to the Study

The models are useful to better understand the factors that affect the learning process of the students and techniques to determine the direction or pathway of the students. At first glance, the model may only provide more queries, but by further observing the model it can provide a starting point to use the factors in the models for the academic streaming purposes.

# 2.11 CONCEPTUAL FRAMEWORK

It is important to provide a multidimensional approach by combining the performance of students as human capital, IT, economy, social, institutional dimensions to achieve a student's performance data that will support for planning tasks, which is an academic streaming process to identify a suitable area for the students. School is an early stage of formal education in Malaysia has the responsibility to ensure the all the factors that influence the student's academic streaming can be identified and the process can be done with a systematic method to make sure the development of human capital will produce an adequate supply of intellectual assets to support and sustain an ideal, competent, and mobile workforce with relevant knowledge and skills (Ahmed 2008).

A conceptual framework was constructed based on the literature review and the relevant models that were discussed in previous sections in this chapter, to provide a clearer picture of the research. The framework is assumed as a system of variables or key factors and theories that support the research. Figure 2.9 shows the conceptual framework for the research.

The framework identifies the categories of factors that are influence the academic streaming. The framework is related to the factors that influenced the decision made by students to choose the academic streaming and not use in clustering process. The clustering process only using examination results, since in Malaysia, currently the academic streaming only involved exam result. This framework categorized and organized all the factors based on origin model and the literature review done; that might be used to answer the question "What are the factors that cause the students have difficulty in determining suitable streaming?" These categories seem included in the framework since they are principally intertwined in the process of determination the appropriate streaming for the students. Whereas, in the middle of the framework is the output that generated by the factors surrounding the framework.



**Figure 2.9 : Conceptual Framework** 

The previous models by Huitt (1995) and Bertalanffy (1969) are dedicated to the factors for teaching or learning process. The models focused on school and classroom processes that predicted school learning from measurement on standardized tests of basic skills. The important part of these models is the redefinition of academic learning time - the quality of that time engaged in academics. As discussed in Proctor model (1984), the concept of success is an important component of quality of time spent and created the term Academic Learning Time (ALT) which they defined as "involved in academic learning at a high success rate".

The Huitt (1995) model adds factors for context and student and teacher characteristics, some of which were the focus of the models by Proctor (1984) and Cruickshank (1986). Huitt believes that these factors must be considered because our society is rapidly changing to a knowledge base. This model still be used until now even it has a revise model, because the revise model is focusing more on student's development in classroom environment (Huitt, 2011). This model is suitable to be adapted in this research since it is view students as a member of a complex society, which effects and transform the way they learn as well as defines the important knowledge and skills that must be developed to be successful in that society. In the proposed framework, a few factors were identified from the reading that has been done. These factors are divided into groups or categories as in the original model system theory that has been developed by Patton & McMahon (1998).

These categories obtained based on the Huitt Model (1995) that originated from TLT and ST. The categories consist of several factors that will affect the process of determining the appropriate academic streaming for secondary students. All the categories and the factors involved, as per discussed in Section 2.5 are summarized in the Table 2.10.

Category	Factors	
Education Institution	Guide	
	Teacher Advice	
Peers	Friends	
Family	Parental & peers pressure	
	self-concept	
	Aptitudes	
Historical trends	Parent's aspiration	
	High expectation	
Workplace	Multiple career option	
Globalization	New industry career trends	
Employment Market	Interest	
	Skills	
	exam results	
Community groups	Gender	
	Ethnicity / Race	
Socio economic status	Family head occupation	

**Table 2.10: Factors Influences the Academic Streaming** 

This framework shows the outcome in the center of framework for academic streaming process to identify a suitable area or class for the particular students whether they are suited for science or arts classes. The teacher must initially identify or suggest an outcome (N. L. Gage & Berliner, 1998) since the outcome will influence the choice of important factors. Once the outcome is determined, the factors involved that explain the fluctuation or variability in the study can be focused. Thus, the Output or Outcome category is the most important and the focus of the proposed framework. The factors influenced by Education Institution category are most important and directly impact the student's academic streaming or pathway. This category includes two major subcategories, which are Guide from the school itself and Teacher's advice.

In addition, friends in Peer and Family category play an important role in influencing the student's decisions to determine which field or streaming is appropriate for them. The category of Family includes the subcategories of Parental & peer pressure (requests and suggestions from parents and friends), Self-concept (related to ambition, passion and desire built on family influences) and Aptitudes (a natural ability to do something). In general, Parents and friends influenced the decision-making to determine the appropriate streaming for the student. The parents may request changes on the choice of streaming if they do not satisfied with the final decision (Duru-Bellat & Mingat, 1989), whereas, a friend is going to influence their close friends to learn in the same class with them (Zevenbergen, 2001).

The Historical trends category consist of factors that relate to the Parent's aspiration and High expectation, which is a closely linked to the aspirations derived from the student's parent and high expectation for the streaming that they chose (Gulyaev & Stonyer, 2002). In the Workplace category, multiple career option is a factor that influences the students to determine their chosen fields. This multiple choice will lead to confusion in choosing the right career for them (Chapman & Mählck, 2004; Gamoran, 2002). Meanwhile, the employment and industry trend factors in the Globalization category can assist the student to plan for a secure occupation or industry (Designs, 2005; Schneider, Judy, Ebmeye, & Broda, 2014). However, the students are less exposed to the new industry trends and labor market information. Therefore, they lack of accurate information on the latest trends in the existing and the future available career.

Next, the Employment market category in educational decisions is steps taken toward implementing career decisions. As we know, the employment market is defined by the interaction of demand and supply (Katz, Murphy, Bound, Freeman, & Coldin, 1991). Besides a demand for an excellent performance in school, the employment market is determined by the skills and interest in of the student in a particular area (Hager, Holland, & Beckett, 2002; Leuven, 2004).

The last two categories are Community groups (gender and ethnicity/race) and Socio economic status (background and family status or family head occupation) that influenced by demographic factors. These categories were chosen as education is always associated with a person's social life (B Duncan & Duncan, 1968; Beverly Duncan, 2014). Consequently, these two categories are important to determine if the students' decision in selecting for appropriate streaming are influenced by the demographic data.

# 2.12 DISCUSSION

From the reviewed literature, various academic performance systems have been used in Malaysia. The systems used a variety of methods and platforms that involved various factors such as examination performance and co-curricular activity. Among the platforms used are manual, semi-automated which are standalone and fully automated using a web-based technology. Most of these systems are using the same concept and implemented at the school level. Different schools utilized different systems and did not coordinate between schools.

Recently, the uniformity of the system throughout the school in Malaysia has been implemented but, with several weaknesses. Most of these academic performance systems are focused on students' academic performance, rather than identifying the academic streaming. Trends in the student's performance in Malaysia indicated that the academic performance and student's academic streaming doesn't link to identify the student's pathway. In this case, the real potential of the student cannot be identified and contributed to the human capital plan. Basically, the academic performance system should be linked to planning division, to ensure the planning for student's learning pathway is always interconnected with the academic performance.

The academic streaming process is performed manually and without using specific models or tools. Usually, this process is based on parental demand, heterogeneous, homogeneous, ability and cluster technique. The cluster technique is the most referred and used in academic streaming, however it was implemented manually. A systematic clustering technique in education involved the data mining technology for prediction and rarely for planning purposes. The academic performance is analyzed by clustering technique for implementation in academic streaming. Using this analysis, streaming process will be more systematic and the teachers could devote a full attention to their main tasks, which is to instruct students' in their learning process.

For the newly proposed and developed conceptual framework, theory or conjectures, the ideas should explain factors in the academic streaming. In this study, the theory is used to understand the academic streaming environment as the factors that affect the process of academic streaming need to be identified and understand its importance. Therefore, the teaching/learning theory (TLT) and system theory (ST) are used. The theories have been proposed as a potential overarching framework for dealing with many issues in human development. For this research, the theories is used because the potential of the theories in furthering the integration of theory and practice, incorporating aspects of TLT and ST into theoretical formulations, to drawing on conceptual frameworks of academic streaming in secondary school. It is derived from TLT and ST as frameworks within which to further understand important factors of academic streaming.

### **CHAPTER 3**

### **RESEARCH METHODOLOGY**

# 3.1 INTRODUCTION

This chapter outlines the research design for this explanatory study and the manner in which the research was conducted. This chapter will cover topics such as research setting, research sample, research instrument, method of distribution and data analysis techniques.

It is divided into four major parts comprising of four major phases of the study namely: research review, semi-structured interview, surveys and content analysis. Validation procedures by experts are presented in order to undertake a manageable surveys and research instruments were verified and validated before they were applied in the actual field study. The analytical approach consists of reliability and validity of the instruments. The primary objectives of the research are to investigate how academic streaming is applied in school and to discover problems faced by students in choosing the right streaming and then, give suggestion to schools with regards to suitable techniques that could be used to helps student's academic streaming.

This study was conducted using three major processes of data collections; analysis of relevant documentation, interviews, and studies using questionnaires to evaluate factors that are able to influence student's academic streaming process in school. The research adopted a mixed methodology, using embedded strategy method as the overall research approach.

### **3.2 RESEARCH PHILOSOPHY**

A research philosophy is a belief about the way in which data about a phenomenon should be gathered, analyzed and used. Three philosophical approaches are positivism, interpretivism and critical realism. The studies related to the data mining technology which is clustering technique and educational show a trend towards an interpretive approach in investigating and interpreting the research phenomenon (Chang, 2008; Chapman & Mählck, 2004; Klein & Myers, 1999). Besides, application of clustering approach in academic streaming is the most approachable application in educational environment. Based on Yin (1994), an interpretive investigation is suitable to be used if the topic of the study is not yet a well-researched area. Furthermore, according to Schwandt (1994), interpretive provides a deep insight into "the complex world of lived experience from the point of view of those who live it". Its application is suitable in this study because it helped to produce an understanding of the context of clustering technique usage in academic streaming and is influenced by the context.

Interpretivism approach is suitable for use in this research compared to the other two philosophical approaches, because interpretivism research claims that the development, usage and implementation of any technology close to people that involved in a situation, and in an educational environment it is better to study processes involving a community. On the other part, positivist approach highlights the quantification of constructs whereby the use of variables has a central role (Mingers, 2003). Positivism paradigm is not suitable to use in this study as it does not provide the mechanism to explain detailed processes of academic streaming. Besides, application of critical realism pursues

substantial fundamental links among phenomena rather than recognized links (Mingers, 2003).

Interpretivist approach is suitable to be used as it will investigate the research phenomenon by involving the people who are contributing in academic streaming in secondary school. Further, input from teachers and secondary school students would become the primary information for understanding and analyzing the research phenomenon. Therefore, in order to understand current practice of this approach, a qualitative and quantitative investigation which will provide rich data from various data content analysis.

#### **3.3 RESEARCH PROCESS**

Research process describes how the study was conducted from the beginning until the end. Figure 3.1 describes research design or research process. This study adapted research design method from Lewis (1998). The most important part is the problem analysis which is to identify problems in current situation of student's academic streaming based on literature review. It was done to find related studies to increase understanding of current situations and the problems itself. Then, relevant theories or concepts that are suitable in this study were also identified, in order to describe how the situation or problems can be addressed. All these tasks are used to assist the creation of conceptual framework based on possible factors involved. The conceptual framework is then developed to be examined using research instruments (interview and surveys). Then, content analysis is conducted on academic performance data of 456 students to find clustering technique that suited the data. The solutions are produced based on content analysis, and consistent with proposed conceptual framework. With the proposed solution, students can be grouped based on their academic performance and their learning pathway can be planned. Apart from that, the important factors influencing student's academic streaming can also be identified through the conceptual framework. These factors can be used as a guide in determining academic streaming. There is no decision activity in the process because typically the activity joined and blend together through the whole process, with regular looping back to previous stages to understand the problem.

Here are detailed steps of the research design:

### Step 1: Analysis

Research problem or problem recognition is identified by reviewing previous works related to students' learning pathway. It focuses on students' academic streaming; such as how the streaming was done in school, techniques that are commonly used, problems faced by students in choosing the right streaming, and how the concept of student's distribution based on their performance can be adapted using clustering technique. To confirm the findings from literature, data collections are then conducted.

# Step 2: Design

This process consists of two tasks:

- 1. Design the instruments to be used:
  - a. Interview To investigate how to identify real potential of each student and how students' academic streaming is determined, this item is discussed in section 3.5.1.
  - b. Surveys to find important factors influencing the students to determine suitable streaming, this item is discussed in section 3.5.2

- c. Content analysis identify related documents to review and followed by request of permission from schools. This item is discussed in section 3.5.3
- 2. Choose a Clustering technique that will be used to test using real student's performance data for the one whole year. This item is discussed in details in section 4.4.1.

# Step 3: Development

This process consists of two tasks:

- a. Develop and conduct data collection using Interviews with 17 teachers who teaches in schools under Ministry of Education Malaysia (MOE). This item is discussed in section 4.2
- b. Develop and conduct data collection using Surveys. 700 set of surveys for Secondary four students (who are just select their streaming) are distributed to 7 schools under Ministry of Education Malaysia (MOE). This item is discussed in section 4.3
- c. Develop analysis for three most commonly used Clustering technique to group students based on their performance. The analysis is using real student's performance data for the one whole year. This item is discussed in details in Chapter 4.4.

# Step 4: Evaluation

Evaluation of data collection tasks are done by: an analysis and a surveys as below.

a. Thematic analysis – to evaluate interview data. It is for identifying, analyzing, reporting and validating patterns or themes within data. This item is discussed in section 5.2.
- b. Partial Least Square (PLS) to evaluate surveys data partial least square approach - Structural Equation Modelling (PLS-SEM) is applied, which is to tests the degree of importance of all factors in the model. This item is discussed in section 5.4.
- c. IBM Modeler for clustering analysis to discover patterns and trends in structured or unstructured data more easily, using a unique visual interface supported by advanced analytics. It also supports to quickly build predictive models using student's performance data and deploy them to improve decision making on the student's academic streaming task. This item is discussed in details in Chapter 5.6.



Figure 3.1 : Research Design (Lewis, 1998)

#### **3.4 RESEARCH APPROACH**

Researchers usually use one method for their research. But good research usually benefits from the use of a combination of methods. Hansen, Cottle, Negrine, & Newbold (1998) proposed that researchers should not only consider which is the most appropriate method for their study or problem, but also what combination of research methods might produce a better and deeper understanding of it. The aim should always be to choose those methods, or combination of methods, which can elaborate the most angles and dimensions of what are invariably multidimensional and complex processes and phenomena (Hansen et al., 1998).

Keeping in view the nature and requirement of this study, interview, surveys as well as content analysis was adopted to explore and examine the current implementation of academic streaming in regards to students' academic performance in secondary schools.

# Mixed-methods

Mixed methodology involves combination of quantitative and qualitative techniques. It focuses on the same phenomenon, which is the data collected can incorporate strength and support the findings of another (Creswell 2005; Creswell & Vicki 2010). This opinion was also agreed by Onwuegbuzie and Collins (2007). In addition, Abusabha and Woelfel (2003) suggested three main reasons to use mixed-method that includes:

- All data have objective and subjective components.
- A variety of methods allow results of the study to be compared and strengthened.
- The use of the method is to complement the deficiencies of each other.

This method provides an opportunity to use a variety of materials from a variety of approaches to get information in the study (Axinn & Pearce, 2006). Mixed-method used in the research is a combined analysis of quantitative and qualitative data as suggested by (Rogers, 2002) and (Sandelowski, 1995). According to Gay et al. (2006), a mixture or combination of methods, is that the research design approach combines qualitative and quantitative data analysis in a study. This statement is also agreed by Creswell(2009), which they define as a mixture of research and analysis of the process of gathering data and merger information using both qualitative and quantitative approaches.

Creswell (2005) states that the main purpose of mixed-method is to produce a strong qualitative and quantitative research to fully understand the phenomenon of study compared the use of qualitative and quantitative methods on their own. Besides, the technique significance is to strengthening support for an issue, arguments or results obtained from various perspectives. There are several reasons to use mixed-method, including validity, balance, complement, explanation, unexpected results, and the development of instruments, samples, credibility, context, illustrations, utilities and data validation (Lawrence 2006). Mixed-method also used to deal with unexpected results for example; an irrelevant or information different from the expected outcome (Awal & Bakar, 2008).

(Creswell, 2009) highlighted six major strategies for mixed-method evaluation design that might enhance the evaluation, as shown in Table 3.1.

# Table 3.1 : Mixed-method Strategies

Strategy	Steps	Purpose	Strength	Weaknesses
Sequential Exploratory Strategy (2 stages)	<ul> <li>Priority in 1<sup>st</sup> stage</li> <li>Collect &amp; analyzes (qualitative)</li> <li>Collect &amp; analyzes (quantitative)</li> <li>Integrated during interpretation phase</li> <li>May or may not have theoretical perspective</li> <li>Quantitative assists interpretation of qualitative</li> </ul>	Explore a phenomenon (determine the distribution of a phenomenon within a chosen population); grounded theory	Separate stages in design, description, reporting	Length of time b/c of 2 separate stages; can be difficult to build from the qualitative analysis to quantitative data collection
Sequential Explanatory Strategy (separate stages)	<ul> <li>Collect &amp; analyze (quantitative)</li> <li>Collect &amp; analyze (qualitative)</li> <li>Integrated during interpretation phase</li> <li>May or may not have theoretical perspective</li> </ul>	Purpose: use qualitative to assist in explaining and interpreting the quantitative.	Separate stages in design, description, reporting	Weakness: length of time b/c of 2 separate stages (especially if equal emphasis)
Sequential Transformative Strategy (2 stages)	<ul> <li>Either method used first, either priority or equal emphasis</li> <li>Results integrated during interpretation phase</li> <li>Theoretical perspective drives research not just methods</li> </ul>	Employ the methods that will best serve the theoretical perspective	Separate stages in design, description, reporting	Length of time b/c of 2 separate stages (especially if equal emphasis)
Concurrent Triangulation (one stage)	<ul> <li>2 methods in attempt to confirm, cross-validate, or corroborate findings within one study.</li> <li>Methods offset weaknesses of other method</li> <li>Ideally, priority is equal but not always practical</li> </ul>	Integrates results during interpretation phase: convergence strengthens knowledge claims or explains lack of convergence	Familiar; well- validated and substantiated findings; shorter data collection than 2 stage studies	Great effort and expertise to study phenomenon with 2 methods; difficulty comparing the results b/c of different methods; unclear how to resolve discrepancies in findings between methods
Embedded Strategy (one stage)	<ul> <li>The predominant method that guides project (lesser is embedded or nested, which can address a different question or seek information at a different level)</li> <li>Data mixed during analysis phase</li> </ul>	Broader perspective than one method (embedded quantitative can enrich description of the sample participants; embedded qualitative describe aspects of quantitative that can't be quantified);	Shorter data collection; both quantitative; gain multiple perspectives from different types of data or different levels within study	Data must be transformed to be integrated within an analysis phase;
Concurrent Transformative Strategy (one stage)	<ul> <li>Specific theoretical perspective drives research (critical theory; advocacy; participatory research; or a conceptual or theoretical framework)</li> <li>Choice of model (triangulation or nested) facilitates theoretical perspective</li> <li>Equal or unequal priority during single collection stage</li> <li>Integration most often during the analysis phase (but can be during the interpretation phase)</li> </ul>	Purpose: theoretical perspective drives all methodological choices (problem definition, design and data source identifying, analyzing, interpreting, reporting results throughout the process)	Transformative framework; shorter data collection; both quantitative; gain multiple perspectives from different types of data or different levels within study	Data must be transformed to be integrated within an analysis phase;

This study uses mixed methodology using embedded strategy. This strategy can be conducted either as a qualitative or quantitative study that embeds a smaller strand of other method, as an enhancement. A secondary strand can be concurrent or sequential. In this study, activities of data collection and analysis are performed using two methods simultaneously, in which qualitative is a primary method and quantitative is a secondary method that provides a supporting role in the procedures. It is useful to capture the best of both quantitative and qualitative approaches. Furthermore, it is used to generalize findings of a population and develop a detailed view of meaning of the phenomenon or concept for individuals; since the data collections involved two different levels of respondents. Interview involves teachers and surveys involve Secondary four students.



**Figure 3.2 : Embedded Concurrent Designs** 

The concurrent embedded strategy may be used to serve a variety of purposes. Often, this strategy is used to gain broader perspectives as a result of using different methods as opposed to using a single method. This embedded strategy of mixed methods identified by its use in data collection phase, during which both quantitative and qualitative data are collected simultaneously (see Figure 3.2). Unlike the traditional triangulation model, an embedded approach has a primary method that guides the project and a secondary database that provides a supporting role in the procedures. Given less priority, the secondary method; which is surveys and content analysis embedded within the predominant strategy (interviews). This strategy is used to seek information at a different level of analysis. The analogy to hierarchical analysis in quantitative research is helpful in conceptualizing these levels (Tashakkori & Teddlie, 2003). The mixing of data from the two methods is often done to integrate the information and compare one data source with the other, typically accomplished in a discussion section of data analysis indicated in Section 4.5.

In this study, the qualitative data implemented, together with quantitative data. The qualitative was carried out to get detailed information about linkage between academic streaming and academic performance from teacher's perspective, since they are involved in managing academic streaming in secondary school. While, quantitative data is done to identify the factors involved in the academic streaming selection, based on student's perspective. The qualitative data explores processes experienced by individuals in the academic streaming which are the teachers, while the quantitative data addresses experience by groups of secondary students using the same process. The qualitative method using interviews, while the quantitative implemented using surveys. Interviews used in order to explore the problem, and detailed understanding of a phenomenon to support each factors in the conceptual framework. The literature reviews yield little information about the phenomena of study; so we need to learn more from participants through interviews. Literature does not adequately addressing the use of clustering techniques in academic streaming process and involvement of students' academic performance directly in the process. That is why a qualitative research study is considered to explore this kind of situation from the perspective of teachers that currently handle the process. Apart from that, surveys used to collect opinion in quantitative to explain the experienced by groups of secondary students. The third data collection is content analysis which is done to find appropriate clustering technique to be used with the student's academic performance data.

This technique was chosen because; they are 2 levels of respondents that play different roles:

- Teachers handle students' academic performance in school
  - Major importance of respondents is the teachers that are closely involved in academic streaming at schools. In this case, obviously teachers that handle the academic streaming process are an expert which is his/her opinion is more likely precise than non-expert and she/he is totally involved in the process of data collection. Each school only has one or two teachers that handle the process.
  - Involved in face-to-face interviews with 17 teachers to obtain specific voices about the academic streaming process
  - Secondary school students to choose academic streaming based on their academic performance in final year examination.
    - Students are responsible to make a decision and a planning that relates to their learning pathway by choosing their academic streaming
    - content analysis and surveys with a large number of individuals, consists
       of 465 secondary students from 5 schools under MOE
    - Surveys To understand student's perspectives about academic streaming process in secondary schools
- Content analysis Using academic performance data of Secondary three students for year 2013 to test suitable clustering technique to assist students in determining their academic streaming.



The data collection steps that followed is illustrate in Figure 3.3.

**Figure 3.3 : Data Collection Steps** 

Based on the Figure 3.3, collected data mixed during at the end of the analysis for interview and survey. The data is not compared side by side but presented as a joined report that provides an overall composite assessment of all factors involve in academic streaming (Section 5.8). Whereas, content analysis is done to propose a clustering model that can be used based on students' performance data.

This method is used because it is a broader perspective than one method. Embedded quantitative can enrich description of the sample participants while qualitative can describe aspects of quantitative that can't be quantified. The strength of this strategy is; it involves shorter data collection because both quantitative and qualitative data were collected simultaneously; as well as gains multiple perspectives from different types of data or different levels within the study.

# 3.5 DATA COLLECTION INSTRUMENTS

Data collection is a process of gathering and assessing information related to the study. The main purpose of collecting data is to obtain answers for research questions and evaluate outcomes. There are three types of instruments used in this study; i) interview; ii) surveys and iii) content analysis. The next sub-section will describe each instrument in details.

# 3.5.1 Semi Structured Interview

A key feature of a semi-structured interview lays in the partial pre-planning of the questions. Semi-structured interviews still allow for replications of interviews with others, but are being less controlled. Semi-structured interviews may be conducted in various modes: this research applied face-to face interviews. It involves 17 teachers from 14 schools in Malaysia under MOE and was held in November and December 2013. Interview questions can be found in Appendix A. Interview initiated by sending a letter of permission to conduct the interviews to 25 schools throughout Malaysia. Letter to obtain permission can be seen in Appendix B. After two weeks, follow-up is done and 8 schools declared their permission. A week after that, permission from the other 6 schools was obtained; bringing the total number of schools that involved in the interview about 14 schools.

In this research, interview was chosen as the first instrument because:

- The primary advantage of interviews is that they provide much more detailed information than what is available from the other data collection methods, in this instance, surveys.
- Standardization of at least some of the questions increases data reliability.
  - They also able to provide a more relaxed atmosphere in which to collect information—people may feel more comfortable having a face-to-face conversation as opposed to filling out a surveys.

## 3.5.2 Data Analysis Approach For Interview (Thematic Analysis)

Interview data was analyzed manually using thematic data analysis. Thematic analysis is a method for identifying, analyzing, and reporting patterns or themes within data (Braun & Clarke, 2006). It slightly classifies and explains the data set in detail and thoroughly. Anthropologist Morris Opler (1945) saw the identification of themes as a key step in analyzing cultures. In this study, thematic method is suitable, since this analysis was conducted to analyze cultures, or circumstances of a process.

All the interview data were printed, read multiple times and put in the potential themes in a form of margins in order to identify the themes. These were then organized, revised, and studied for relations and redundancies. From time to time, the themes expanded, contrasted and improved as more transcripts were analyzed. To ensure the accuracy of the analysis, the concept of trustworthiness by Lincoln and Guba (1985) was applied and is also discussed in the Section 3.6.

This method is continued by analyzing across and between the data until no more thematic patterns could be recognized and the data is considered had been saturated. Once these patterns were identified, they were grouped into the major themes as presented in Section 5.2 and discussed in Section 5.3. Further, data analysis also indicated that the amount of data collected was sufficient to describe confirmation of the findings that; it is difficult to identify real potential in individual student and students' academic streaming cannot be determined systematically based on any rules, as well as other perceptions from the interviewee to answer the research questions.

#### 3.5.3 Surveys

This surveys uses closed-ended items. Closed-ended questions are used for group of respondents that originated from the same demographic or profile (Graff, 2009). In this study, respondents to this surveys have been known to come from the same profile of students which is Secondary four students who just started to be streamed, and the aim of the surveys is to determine all factors that influencing the selection of streaming by the students.

Respondents were asked to respond to all the statements using the Likert scale ranging from Strongly Disagree (5); Disagree (4); Not Sure (3); Agree (2) and Strongly Agree (1). A likert 5 point agree/disagree scale with middle option was used to understand the extent in which the respondents agree or disagree with the statement that relates to the streaming process. To be more precise, the surveys are conducted to discover the factors that influencing the students in choosing the right streaming. Students were sent surveys that included closed-ended items; but, data for the quantitative and qualitative strands were gathered and analyzed separately since the level of respondents are different.

Various models related to student performance and student's academic streaming techniques, from the literature were considered before designing the questionnaire. The teaching and learning model by Huitt (1995) which is based on system theory model developed by Bertalanffy (1969) was chosen, because its environment and process, as well as related factors was closely related to the needs of this study.

The objective of this surveys is to discover the problems faced by students in choosing the right streaming. Processes involved for this questionnaire are discussed in subsections 3.5.3.1 until 3.5.3.4.

## **3.5.3.1** Population and Sampling

The unit of analysis was an individual student from Secondary four students who just started to be streamed, with a total of 465 students from seven schools all over Malaysia. The sample size used is based on the techniques adopted by Krejcie & Morgan (1970) as per shown in Appendix C . Based on the table developed by Krejcie & Morgan (1970), the recommended sample size for a population more than 1 000 000 it is 384. Hence, a representative sample size of 465 was deemed reasonable to give a satisfactory response rate. For this study, convenience sampling was used. In convenience sampling, a sample is selected from elements of a population that are easily accessible and willing to be studied (Creswell, 2009). In this study, the respondents provide a reasonable representative profile of all students.

#### **3.5.3.2 Instrument Development**

This questionnaire consists of two parts. Part 1 is related to personal details and Part 2 is a four-page Likert 5 points agree/disagree scale w/middle option. The questionnaire was divided into the following two parts, as shown in Table 3.2:

Part 1: Demographic details, such as gender, ethnicity/race and family head occupation and level of difficulty in choosing academic streaming (five questions)

Part 2: Problems respondents experienced during the process to choose a suitable academic streaming (questions divided into 13 categories; as of 13 factors and total number of questions is 47).

Part	art Item on the		Type of scale	Remarks
	Question	naire	C	
1	1,2,3		Categorical	Demographic details
1	4 & 5		Categorical	Difficulty in choosing academic streaming
	А	1-6	Categorical	
	В	1-4	Categorical	
	С	1-4	Categorical	
	D	1-4	Categorical	
	Е	1-4	Categorical	
	F	1-4	Categorical	
2	G	1-3	Categorical	Factors relate to academic streaming selection
	Н	1-3	Categorical	
	Ι	1-4	Categorical	
	J	1-4	Categorical	
	Κ	1-4	Categorical	
	L	1-4	Categorical	
	М	1-4	Categorical	

 Table 3.2 : The Details of Questionnaire

The initial questionnaire was pre-tested using Cooper and Schindler (2011) pretesting method; whereby the questionnaire was reviewed by expertise that related to the area of study and from the same field, in order to ensure that the questions were both valid and accurate. In this study, teachers and principals are involved to validate the questions. The questionnaires were also distributed in a class of 36 students to ensure they understand the requirements of the questions.

## **3.5.3.4 Data Collection**

Data was collected using a hardcopy questionnaire that was distributed to seven schools secondary schools under Ministry of Education Malaysia (MOE), where 100 sets were distributed for each school, with the approval from their Principals. A follow-up phone calls was made and e-mail was sent to the principals and the teacher in-charge two weeks after the questionnaire distribution to remind respondents to complete their questionnaire because the questionnaire will be collected by the researcher. After a month, a second follow-up was made to schools that have respondents who have not yet responded. All feedback was obtained after 3 months. From a total of 700 questionnaire distributed for seven schools, only five schools provide good feedback, with a total of 483 sets of surveys were returned. After data cleaning processes, by considering completed answers, and avoid duplication of data, only 465 sets can be used for analysis.

# 3.5.4 Data Analysis Approach for Surveys (Partial Least Square Approach; Structural Equation Modelling, PLS-SEM)

The information for Part 1 (refer sub-section 3.5.3.2), obtained from the returned questionnaires was coded and transferred into Statistical Package for Social Science (SPSS) version 17 for Windows and data analysis, such as descriptive statistics and Chi square testing, was used to obtain the results for demographic data. Whilst the information for Part 2 (refer sub-section 3.5.3.2), was coded and transferred into the Microsoft Excel and saved as .csv file to be analyzed using smart PLS. The data was analyzed using smartPLS and applied a partial least square approach - Structural Equation Modelling (PLS-SEM) which is to tests the degree to which certain factors in the framework are related. PLS path modelling is discussed as soft–modelling–technique with minimum loads regarding measurement scales, sample sizes and residual distributions (Monecke & Leisch, 2012). The questionnaire is appended in Appendix D. Each respondent is labelled using unique ID as 1, 2 and so on in both types of file.

PLS-SEM is chosen because of its exceptional capability and fewer assumptions than covariance-based structural equation modelling (CB-SEM). The capabilities of PLS-SEM are such it is able to handle small sample sizes, no assumptions of the particular scale and as well as the normality of the data distribution (Fornell & Bookstein, 1982; Hair, Sarstedt, Ringle, & Mena, 2011; Wold, 1985).

#### 3.5.5 Document Analysis – Content Analysis

Document analysis is used as a systematic procedure for reviewing or evaluating printed documents related to student's academic performance. This study used content analysis

of the documents involved since the study seeks to show patterns of regularities in content through repetition. The categories from the data are derived (e.g., name of the subjects - bm, bi, math, sc) in order that it can be compared with other data. Then, the words (gred for academic performance – A, B, C, D, E, F) in the documents are transformed into numbers. The number of times in which a word occurs in the text is taken as an indicator of its significance in the text. This quantification strategy is expected to improve both the reliability and validity of the classified data. The documents are specifically includes academic performance data for Secondary three students during the year 2013 from five schools under Ministry of Education Malaysia (MOE). Subjects involved are four main subjects which is used for the purpose of academic streaming which is *Bahasa Melayu* (BM), English (BI), Mathematics and Science. Data of these documents can be seen in Appendix E.

### 3.5.6 Data Analysis Approach for Content Analysis

Like other analytical methods research, content analysis requires that data to be examined and interpreted in order to elicit meaning, gain understanding, and develop empirical knowledge. In this study, the content analysis is used to identify academic streaming for the students. The documents involved were analyzed using IBM SPSS Modeler® 15.0.

The IBM SPSS Modeler® 15.0 is used in this research because it can discover patterns and trends in structured or unstructured data more easily, using a unique visual interface supported by advanced analytics (IBM Corporation, 2012). It is also able to quickly build predictive models using educational knowledge and deploy them into daily operations to improve decision making (Devi, Rao, Setty, & Rao, 2013).

In addition, SPSS Modeler offer a variety of modelling methods either using data from machine learning, artificial intelligence, or statistics. The methods allow us to derive new information from data and to develop predictive models. In addition, in this study it can be used to prepare data for analysis, find the best clustering model based on hidden patterns in data and quickly produce consistent and accurate results (IBM Corporation, 2011). This is important to achieve one of the objectives of the study, which is to find the best clustering model that suits the educational data used. In this case, this study will be more focused on knowledge discovery rather than on technical tasks like scripting and coding.

# 3.6 RESEARCH TRUSTWORTHINESS

Qualitative researchers discuss validity of data in terms of trustworthiness and credibility. In this research, for the qualitative part (interview), the validity is determined by adhering to the principles of trustworthiness (Graff, 2009). The trustworthiness was a term suggested by Lincoln and Guba (1985) and is frequently discussed to as 'goodness of fit' criteria which matches the term rigor in quantitative research. The trustworthiness also refers to a finding that are "worth paying attention to" (Lincoln & Guba, 1985). To determine the trustworthiness, it is divided into four criteria which is credibility, transferability, dependability and conformability (Lincoln & Guba, 1985).

Table 3.3 summarizes how the four trustworthiness criteria were implemented in this study.

No	Criteria	Description	Tasks
1	Credibility (in preference to internal validity)	Confidence in the 'truth' of the findings	<ul> <li>Literature review of relevant studies and theories (refer Chapter 2).</li> <li>The data is a real data obtained from schools (teachers and students) who have involved in the streaming process (refer Chapter 4).</li> <li>To ensure the current situation in the academic streaming process, the findings were shared and discussed with the research participants (teachers) and supervisors.</li> </ul>
2	Transferability (in preference to external validity/generali zability)	Showing that the findings have applicability in other contexts	<ul> <li>Data collection process was presented in detail with demographic information as in Chapter 4.</li> <li>The number of participants is appropriate and reached data saturation as all information obtained interrelated and no new information was available.</li> <li>Data coded based on emerging themes in the data, the factors of Teaching Learning Theory based on System Theory and from the relevant literature review guided the research questions, as well as assist for data collection. (Refer Chapter 3).</li> <li>Research involving teachers and students from various backgrounds and does not focus on top students only</li> </ul>
3	Dependability (in preference to reliability)	Showing that the findings are consistent and could be repeated	<ul> <li>Developed and used the same interview questions for all research respondents involved to confirm the acceptancy and consistency. (refer Appendix A for questions)</li> <li>Data were analyzed and findings were reviewed to be complemented to the research questions, conceptual framework, research findings, the solution and research conclusions (refer Chapters 5 and 6).</li> </ul>
4	Conformability (in preference to objectivity).	A degree of neutrality or the extent to which the findings of a study are shaped by the respondents and not researcher bias, motivation, or interest.	<ul> <li>The data is a real data obtained from different schools (teachers and students) who have involved in the academic streaming process (refer Chapter 4).</li> <li>A record of data collected was developed (i.e., interviews: Appendix A, transcriptions: Appendix F, interview notes: Appendix G).</li> <li>Findings from the interview were compared and shared among the same respondents for validation and confirmation.</li> </ul>

# Table 3.3 : The Four Trustworthiness Criteria (Lincoln & Guba, 1985)

#### 3.7 VALIDITY AND RELIABILITY OF INSTRUMENTS

The research instruments need to be verified the validity and reliability. Survey questions have to be carefully examined especially when the survey has not been previously verified. In this study, the surveys and interview protocols were field-tested with the help of students and teachers who were not part of the sample population of this study. A reliability and validity test of the instrument was conducted. The validity of the instrument was determined, followed with the reliability of the instruments.

#### 3.7.1 Validity

Reliability alone is not sufficient. For a test to be reliable, it also needs to be valid. Validity refers to the degree in which our instruments is truly measuring what we intended it to measure. For this study, firstly, a thorough check for hard words and clarity in the questions during instrument development is performed to lessen the logical error in the measure. To make sure content validity of the questionnaire, idea was obtained from several sources. Popham (1999) defines content validity as the attempt "made to judge the degree to which a test is consistent with the content, skills, or objectives it is supposed to measure" (p. 123). A thorough check, by two experts in the subject matter and in measurement, addresses the issue of content validity. All research instruments were reviewed to determine the suitability and significance of the content, to the focus of the instrument. Two faculty members also reviewed the instruments to explicitly align the survey questions with the research questions. Researchers from the other universities (International Islamic University Malaysia and Universiti Kebangsaan Malaysia) also reviewed and made few ideas to augment the instrument.

Reliability is the degree to which an assessment tool produces stable and consistent results. To test the reliability of the instrument, the survey process and the questionnaire were field-tested.

A number of changes ensued to simplify the questionnaire and the input process. The initial concern was the ability of the respondents to understand the questions and follow the instructions for conducting the survey. Thirty six Secondary four students who just started to be streamed from a school in Kuala Lumpur, Malaysia, participated in this field-test. The field-test was conducted in their classroom during school sessions, and carried out for approximately 1 hour 30 minutes. After they finish completed the survey, they were asked for the length of time it took to complete the survey. The time it took them to complete the task given ranged from 50 minutes to over one hour. They also asked about changes they would make if questions were either not clear or ambiguous. Their comments were noted, and changes to the questionnaire were made. Several modifications were made to the questionnaire including the elimination of impertinent survey questions that did not apply to the research questions.

# 3.8 SUMMARY

This chapter discussed the research design for the study in detailed and systematic practice. This chapter covers topics such as research setting, research sample, research instruments, method of data collections and data analysis techniques. The validity and reliability of the instruments also covered in this chapter. Details of the data collection process for interview, surveys and content analysis are discussed in Chapter 4. Then, Chapter 5 presents the data analysis and findings from the data collections.

university

#### **CHAPTER 4**

## **DATA COLLECTION**

# 4.1 INTRODUCTION

This chapter describes how the data collections are carried out in details. It explain how the mixed-method discussed in Chapter 3, which is using embedded strategy are applied in this study.

Data collection consist of three phases; i) data collection 1 is using interview, ii) data collection 2 is using surveys and iii) data collection 3 is using content analysis. Data collection 1 (interview) is to explore the problem, and detailed understanding of an academic streaming phenomenon in secondary school to relate with the factors in the conceptual framework; as in the research background the difficulties in identifying real potential in individual student because, students' academic streaming cannot be determined systematically based on any rules. Hence, interview is conducted to find out how the academic streaming is done in schools. Then, the surveys are conducted to collect opinion in quantitative to discover problems faced by secondary students in choosing the right streaming, as well as to support the conceptual framework. Further, content analysis is conducted to analyze student's academic performance data and identify an appropriate clustering technique that can be applied to support academic streaming.

This chapter aims to answer the following research questions:

# Data collection 1

*RQ1* : How student's academic performance and student's academic streaming is applied at school?

Based on this research question, the research respondents (teachers) were asked about how real potential in individual student is identified and how academic streaming is done in school. What is a role of student's academic performance data to assist process of student's academic streaming in school? The question seeks to understand the academic streaming process and the teachers' view regarding the process itself; types of main academic streaming in school, the main factors for choosing students for certain streaming, and whether there are any guidelines or policy to be followed in determining the streaming of students (refer to Appendix A for Interview Questions)

## Data collection 2

*RQ2*: What are the factors that influence the students in determining suitable streaming?

Based on this research question, the surveys was asked to a group of students in what would be the actual problems and challenges that cause the students to have difficulty in determining suitable streaming for them. The surveys is related to the conceptual framework developed in Chapter 2, that mentioned about related factors in identifying suitable streaming (refer Appendix D for the Surveys Questions).

#### Data collection 3

- *RQ3*: What are the clustering models that can be used for the distribution of students into appropriate stream?
- *RQ4*: What is the specific clustering model that is suitable to support academic streaming process using the student's academic performance data?

Based on literature in Chapter 2, a technique that often used for handling a distribution of student is a clustering technique. Therefore, this section aimed at answering RQ3 and RQ4 that are related to find appropriate clustering techniques that can be used to produce a balanced distribution of students.

Phase	Instruments Respondents	
Data Collection 1	Interview	17 teachers from 14 schools
		in Malaysia
Data Collection 2	Surveys	465 Secondary four students
		who just started to be
		streamed from seven schools
Data Collection 3	Content Analysis	500 Secondary three students
		from five schools

**Table 4.1 : Data Collection Phases** 

Table 4.1 shows the phases involved in the process of data collection in this study, instruments used and also the respondents for each phase.

# 4.2 DATA COLLECTION 1 – INTERVIEW

This section covers data collection technique of semi-structured interviews and selection of the research participants.

To support the literature that; *it is difficult to identify real potential in individual student and students' academic streaming cannot be determined systematically based on any*  *rules*, an interview which is a face-to-face interview with the teachers is conducted, consist of teachers and principals.

The interview questions were formed based on three main components: research questions, conceptual framework, and teaching / learning theory (TLT). Following a thorough literature review (refer to Chapter 2), the conceptual framework was developed based on research questions and teaching / learning theory (TLT). Before the interview was conducted, a few preparations for the interview have been done to ensure smooth running of the interview. The Table 4.2 describes the preparation before the actual interview was conducted.

Task	Sub- tasks	Activity
Planning	Identify likely participants to be interviewed	Identify schools to be sent the letter of permission
	Determine an adequate total of participants	A non-probabilistic, purposive sampling approach was used.
	65	Participants: teachers that directly handle the academic streaming process.
		interviewed teachers who met at least three basic criteria: (i) handle streaming process; ii) familiar with student's academic performance; and iii) willing to respond to the finding of the interview
	Prepare official letter for permission to conduct the interview	provide a letter of permission stating the objective of the interview and detail about the research that being conducted
Developing	Prepare what to say to interviewees when setting up the interview	<ul> <li>Prepare what to say to interviewees when beginning the interview in concluding the objective of the interview.</li> <li>Also informed consent and confidentiality of the interviewee confidentiality</li> </ul>
	What to do during the interview	Recording and take notes

 Table 4.2 : Pre-interview preparation

The interview involves 17 teachers from 14 schools in Malaysia under MOE. It was held in November and December 2013. Interview questions can be found in Appendix A. Interview initiated by identifying schools to be sent the letter of permission and total number of participants. This study used a non-probabilistic and purposive sampling approach. The participants for the interview are teachers that directly handle academic streaming process. The teachers should met at least three basic criteria: (i) handle streaming process; ii) familiar with student's academic performance; and iii) willing to respond to the finding of the interview. Then a letter of permission was sent to 25 schools throughout Malaysia. Letter of permission stating the objective of the interview and details about the research that being conducted. Letter of permission interview can be seen in Appendix B. After two weeks, follow-up is done and 8 schools stated their permission. A week after that, permission from the other 6 schools is obtained; bringing the total number of schools involved in the interview about 14 schools. Before the interview is conducted, scripts during the interview session to conclude the objective of the interview are prepared. Consent and confidentiality of the interviewee are also informed.

#### **Research Respondents for Interview**

For data collection 1, a total of 17 interviews with teachers in 14 schools under Ministry of Education Malaysia (MOE) were carried out. The interviews were held in November and December 2013. There are three types of teachers involved in the interview, which are teacher (data), teacher (exam), and school's principal. Teacher (data) indicates teachers that handle student's data related to their details, while teacher (exam) indicates teachers that handle student's academic performance data. Demographic information on the research respondents in data collection 1 is summarized in Table 4.3.

<b>Fable 4.3 : Demographic</b>	Information	of Teachers	for Data	Collection	(Interviews)
--------------------------------	-------------	-------------	----------	------------	--------------

ID	Gender	Age	Position	Urban School / Rural School
T1	F	25 - 35	Teacher (data)	Urban
T2	F	25 - 35	Teacher	Rural
T3	М	25 - 35	Teacher (exam)	Rural
T4	F	36 - 45	Teacher (data)	Urban
T5	F	36 - 45	Teacher	Urban
T6	F	25 - 35	Teacher	Urban
T7	М	36 - 45	Teacher	Urban
T8	F	>45	Principal	Urban
T9	М	36 - 45	Teacher	Rural
T10	М	25 - 35	Teacher (data)	Rural
T11	М	25 - 35	Teacher	Urban
T12	F	25 - 35	Teacher	Rural
T13	М	25 - 35	Teacher	Urban
T14	М	>45	Principal	Rural
T15	М	25 - 35	Teacher (exam)	Rural
T16	F	25 - 35	Teacher (data)	Urban
T17	F	25 - 35	Teacher (exam)	Urban

Legend:

F = Female = 9 Total Respondents = 17

M = Male = 8

This interview aimed to investigate current practices applied in student's academic performance system and students' academic streaming in Malaysia, in order to understand how student's academic performance and student's academic streaming is applied in school. In this case, role of teachers in school and types of school will give more revelatory insights to the research. The role or position of teachers in school is identified to determine teachers' understanding on how academic streaming is done. Hence, the respondents come from various positions, such as Teacher (data), Teacher (exam), Principal or ordinary teachers. In addition, type of school (rural or urban) may influence the students in determining their streaming. Rural school is a school that is located in the remote areas, while urban school is a school that is located in the city. From a total of 17 teachers; four of them are teachers (data); three are teachers (exam), two principals and the rest are ordinary teachers. Furthermore, ten schools are from

urban area, and four schools are from rural area. The next sub-section provides a view of the data analysis process performed for this research.

# 4.3 DATA COLLECTION 2 – SURVEYS

This section covers the data collection 2; technique used here is surveys. This section also covers the selection of research respondents.

Data collection using surveys aimed to identify the factors that influence the students in determining appropriate streaming. In this case, views from the students who are directly involved in the process of academic streaming selection, will give more revelatory insights to the research. The next sub-section provides a view of the data analysis processes performed for this research.

In order to support the findings indicated in literature review: *the factors that influence the students in determining appropriate academic streaming*, a surveys was conducted as a closed-ended surveys for form four students, because they have chosen academic streaming that is assumed appropriate for them. The surveys questions were formed based on three main components: research questions, conceptual framework, and teaching / learning theory (TLT) itself. Following a thorough literature review (refer Chapter 2), a conceptual framework was developed based on the research questions and TLT.

#### **Research Respondents for Survey**

For data collection 2 which is surveys, a total of 700 sets of surveys were distributed to 7 secondary schools under Ministry of Education Malaysia (MOE), where 100 sets were distributed for each school. Circulations of the surveys are carried out by teachers in that school with the principal's permission. From a total of 7 schools, only 5 schools provide good feedback, with a total of 483 sets of surveys were returned. After data cleaning processes, by considering completed answers, avoid duplication of data and ensure students involved the assessment of their academic performance for 1 whole year of 2013, only 465 sets can be used for analysis. The surveys was distributed in January 2014 and about 3 months later, all feedback was obtained.

A total of 465 respondents responded to the questionnaire. 65.8% of respondents are Female. The respondents are categorized based on the occupation of the family's head; employed (57.8%), self-employed (40.1%), a pensioner (1.4%) and not working (0.7%). The majority of respondents are Malay (77.6%), Chinese (18.7%), and Indian (3.7%).

### 4.4 DATA COLLECTION 3 – CONTENT ANALYSIS

This section will explain data collection 3 which is content analysis. The content analysis of the documents would be quantitative analysis. The documents used are student's academic performance data for year 2013. The content of the documents includes the academic performance of secondary three students from five schools under Ministry of Education Malaysia (MOE), for four main subjects that are used for identifying academic streaming.

The finding then discussed and linked to the objectives of the data collection in the Chapter 5. In this data collection 3, the documents are used to run the analysis that is carried out for the purpose of discovering hidden patterns and relationships, as well as to identify the best natural cluster model for the educational data involved. This is because, based on literature, clustering is an important technique to support the process for student's academic streaming or distributing the students based on their performance. For this purpose, method of analysis used is clustering analysis, using IBM SPSS Modeler® 15.0.

The result of the analysis shows three main information from the data, specifically; model summary, cluster size and predictor importance. From this information, an appropriate clustering model can be determined.

Level/Grade	Age
Preschool	
Pre-school playgroup	3-4
Kindergarten	4-6
Primary school	
Primary 1	7
Primary 2	8
Primary 3	9
Primary 4	10
Primary 5	11
Primary 6	12
Secondary school	
Secondary 1	13
Secondary 2	14
Secondary 3	15
Secondary 4	16
Secondary 5	17
Post-secondary education	
Tertiary education (College or University)	Ages vary

Table 4.4 : The Level / Grade of School in Malaysia

In Malaysia, the school year is divided into two semesters. The first begins in the beginning of January and ends in May; the second begins in June and ends in November. Level or grade of school divided into four stages, which are preschool, primary school, secondary school and post-secondary education. Table 4.4 shows level / grade of school in Malaysia.

Data that have been used is the data from secondary three students, because the academic streaming for placing students in science or art classes done after secondary three is completed. The data obtained from five schools throughout Malaysia, which involved 465 students. Table 4.5 describes the variables used in the analysis.

Variable	Description	Possible Values
Id	Student id	Continuous number
Name	Student's name	String
bm (bm, bm2, bm3, bm4, bm5 – 5 times assessment in a year)	Bahasa melayu(BM) subject	$ \begin{array}{c} 1 \text{ (excellent)} \\ 2 \\ 3 \\ 4 \\ 5 \end{array} $
bi (bi, bi2, bi3, bi4, bi5 – 5 times assessment in a year)	English (BI) subject	$ \begin{array}{c} 6 \text{ (poor)} \\ 1 \text{ (excellent)} \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \text{ (poor)} \end{array} $
mt (mt, mt2, mt3, mt4, mt5 – 5 times assessment in a year)	Mathematics (math) subject	1 (excellent) 2 $\land$ 3 4 5 6 (poor)
Sc (sc, sc2, sc3, sc4, sc5 – 5 times assessment in a year)	Science (sc) subject	1 (excellent) 2 $\land$ 3 4 5 6 (poor)
Class	Group for student whether in science class or art class	c1 = science group c2 = art group

 Table 4.5: Student's Academic performance Data Used in the Clustering Analysis

Currently, MOE using *Penilaian Berasaskan Sekolah* (PBS) or school-based assessment. PBS is a holistic form of assessment that assesses cognitive (intellectual), affective (emotional and spiritual) and psychomotor (physical) in accordance with the National Education Philosophy and Curriculum Standard for Primary Schools (KSSR). It using band 1 until 6 to reflect the student achievement as shows in Table 4.6.

Table 4.6 : Band Using in Penilaian Berasaskan Sekolah (PBS)

Band	Standard
6	Know, understand and can do with exemplary ethics
5	Know, understand and can do with civilized admirable
4	Know, understand and can do with politeness
3	Know, understand and can do
2	Know and understand
1	Know

Band 1 : Students know the basics, or can perform basic skills or respond to the basics.

Band 2 : Students show understanding to change the form of communication or translate and explain what they have learned.

Band 3 : students can use the knowledge to perform their skills in a given situation.

Band 4 : Students perform a skill politely, courteously or perform something systematically and follow the procedures.

Band 5 : Students perform a skill politely in a new situation, systematically in accordance with the procedures, consistent and positive attitude.

Band 6 : Students were able to express ideas that are creative and innovative, have the ability to make a decision in order to adapt everyday demand and challenges. They also can publicly speak using their own words as moral and exemplary consistently, to get and disseminate information.





# Figure 4.1: Current Student's Distribution

Based on the academic performance data which was obtained, it found that distribution of students in science and art classes is not as suggested by the MOE. This is because, MOE requests more students in science class, rather that art class. The percentage of students' distribution is shown in the Figure 4.1. The figure shows current student's distribution for a total of 500 students. C1 is for science class and C2 for art class. Total for C1 class is 193 students which are about 38.6% and C2 is 307 or 61.4%. While the requirements set out by the MOE is contrary with the data; which 40% are art students and 60% are science students. These requirements are consistent with the human capital needed by the country. Hence, clustering analysis is used to obtain a more appropriate distribution of students with the requirements set out by the MOE.

# 4.5 SUMMARY

In this chapter, data collection are described on how the tasks are carried out in details, and how the methods that have been used (Refer Chapter 3) is applied in this study. The data collection is done three times; data collection 1 (interview), data collection 2 (surveys) and data collection 3 (content analysis). The interview is analyzed using thematic analysis, the surveys is analyzed using PLS, and the content analysis is analyzed using IBM SPSS Modeler. The details of data analysis and findings for all the data analysis are presented in Chapter 5.

university

# **CHAPTER 5**

## DATA ANALYSIS AND FINDING

### 5.1 INTRODUCTION

This chapter describes the data analysis done and its findings. It is about how the data analysis are carried out for the data collections (data collection 1: interview, data collection 2: surveys and data collection 3: content analysis). Then the finding from the data analysis are presented.

As described in Chapter 3, the data collection 1 (interviews) is analyzed using Thematic analysis, data collection 2 (surveys) is analyzed using the Partial Least Squares Approach - Structural Equation Modelling (PLS-SEM), and data collection 3 (content analysis) is analyzed using IBM SPSS Modeler® 15.0.

The finding is divided into three parts; finding for interview, finding for surveys and finding for content analysis. Section 5.3 describes finding for interview, subsection 5.5 describes finding for surveys, and subsection 5.7 describes finding for content analysis. The findings for data collection 1 and 2 then integrated or combined, and represent in subsection 5.8.
#### 5.2 DATA ANALYSIS FOR INTERVIEW

For data collection 1 (interview), the unit of analysis is the teacher, with a total of 17 teachers from 14 schools all over Malaysia. In order to analyze the interview data, deductive approach is used, which is thematic analysis (Braun & Clarke, 2006). This kind of analysis provides direction to the study, since both RQ1 and RQ2 of the study are looking for a set of information that has been identified earlier based on literature review. Once the interview has been completed, the interview responses were read and patterns or themes among the data are identified. A variety of themes, codes, or even possible categories that will provide the beginnings of analysis, or ideas if there are needed for upcoming interviews were discovered.

A total of 17 written and audio recordings obtained from the interviews, with duration of approximately 1 hour 45 minutes were transcribed. The transcripts were read many times and prudently examined and analyzed to achieve a complete understanding of the interview. Based on the proposed conceptual framework; which combines the concepts from the literature, categories and factors that relate to Teaching / Learning Theory (TLT), as well as System Theory (ST), 53 themes were identified during the first stage of data analysis. These themes include on whether the student's performance were embedded in student's academic streaming, strategies applied in school in order to assist students in choosing their streaming, processes of academic streaming itself, usage of student's performance data in identifying student's academic streaming, problems with current process of academic streaming, and motivation to use technological tools to assist with student's academic streaming and more. All these were then revised, examined for relations and redundancies. Then, the categories were coded and labelled into main themes. Over time, the themes were expanded, compared and altered as more transcripts were analyzed.

In order to deliver specifics and examples of the real coding process, steps in thematic analysis (Braun & Clarke, 2006) are used:

#### **1.** Becoming familiar with the data

In this step, the researcher tries to understand the data in-depth. It involves constantly reading the data repeatedly, and understanding the data in a few view of points, searching for meanings, flows, patterns and so on. Repeated and constant reading of the entire data set before the coding process is started is beneficial to to identify possible patterns from the data. For accurateness purposes, the transcripts are checked against the original audio and written recording to ensure the accuracy of transcription.

In this research, the researcher generated fifty three (53) themes from the first reading of the transcripts with an explanation for each of the categories. Table 5.1 provides examples of the 53 generated themes at the early stage of data analysis. The themes also recognized based on factors and categories identified early based on literature and TLT used.

#### Table 5.1 : Samples of the generated themes at the early stage of data analysis

N	D Themes	Definition	
1	schedule for question maker	Full schedule to provide marks for each student	
2	exam time table	Detailed schedule for students to sit for examination	
3	dateline for marking and keyin	due date for make sure the student's performance (in examination) n	
4	submit marks to PPD (Hcopy)	send the student's performance marks to PPD in hardcopy	
5	submit marks to PPD (Scopy)	send the student's performance marks to PPD in softcopy	
6	performance in final exam	marks for each students in final examination	
7	gred based on particular subject	marks and gred for for student for particular subject	
•	SAD	System Apolico Deperileccon	
0	SAD123	Examination Analysis System (form 1.3)	
10	SAP45	Examination Analysis System (form 4.6)	
11	SAPR16	Examination Analysis System (torn 4-0)	
		Examination Analysis System (standard 1-6)	
12	enantau	monitoring system for stydent's performance	
	Cpuntuu	monitoring system for student's performance	
33	School performance	school performance for the whole year	
34	Target for student's performance	plan to increase student achievement	
35	streaming	class divisions according to student performance	
36	School performance	school performance for the whole year	
37	reward	The award is given to outstanding students	
38	manually	the streaming process done manually and burden the teachers.	
39	Total of marks	calculation of marks for each student and each class is still done by tea	
		teachers are confusing about several system that keep change. differer	
40	confusing students performance year by year		
41	difficulty in planning (target)	Teachers cannot plan for the target for the students and schools perfor	
		the server load can not accommodate large data. server can not handle	
42	the server over load	in a time, since a teacher need to access the system in the same time fi	
43	valuable analysis	Provide valuable analysis to assist streaming	
44	a balance class distribution or streaming	can suggest for balance class distribution or streaming	
40	retrieve data	easy to retrieve data	
40	data redundancy	eliminate data redundancy	
4/		active menore students performance data	
40	easy to manage	easy to manage students performance data	
49	easy to keep track	easy to keep track students performance data	
50	understand	easy to understand	
51	report	easy to report	
52	planning	easy for planning (for teachers load & target for students performance	
53	teachers can focus	teachers can focus more on teaching and learning	

## 2. Generating initial codes

In this stage, notes are taken and ideas are marked for coding and continued with a formal coding process. The name of each code (and a brief description) is written on a separate piece of paper and organized them into theme-piles. The relationship between codes, themes, and different levels of themes (e.g., main overarching themes and sub-themes within them) is started to be considered.

#### **3.** Searching for themes

Process of determining the themes is done manually which are line-by-line analyses has been done; writing notes on the texts that were analyzing by using highlighter pens to identify potential patterns. This is because; this kind of techniques is suited for rich and complex narratives. A technique to discover a theme in qualitative data is important to describe, compare, and explain about the data that we have. Main themes are developed based on some initial codes, however others formed sub-themes and remaining others were thrown out because of redundancy or there were some reason such as not covered in the main process of academic streaming. Once the themes were satisfied, the researcher is ready to conduct Step 3 of the reviewing themes. An inventory of key points for each sub-theme was developed. In this stage, the entire data set need to be understood systematically, giving full attention to each data item and identifying motivating parts in the data items that created the basis of repetitive patterns through the data set. Table 5.2 provides example of key points development for each category from interview transcripts of respondent 1 which is a teacher (T1).

148

# Table 5.2 : Samples of key points development for each category from respondent 1

**(T1)** 

Kespondent	Key Points	Notes
11	Tasks of teacher that handling student performance data are setting dateline	Tasks in student's performance and studer
	for marking student's papers and key-in their marks, then submit all marks	streaming:
	to Ministry of Education (district) in soft copy.	<ul> <li>Setting dateline for marking studen neners and key in their marks</li> </ul>
	Performance in the final exam and grade based on a particular subject are	<ul> <li>Submit all masks to district (in soft</li> </ul>
	considered for students' streaming in students' performance report.	copy).
	The system used	<ul> <li>Important for school performance,</li> </ul>
	- Enantau	set the student's target and best
	- SPPBS	students reward
		Streaming process:
	The student's performance system is not linked to the student's streaming	<ul> <li>Not linked to the student's streaming</li> </ul>
	process.	process
	Streaming is done: menually no guideline, refer to the student's final result	• Done:
	barad on total of teacher in particular streaming, and barad on total of	<ul> <li>Manually no guideline</li> </ul>
	students	<ul> <li>Refer to the student's final re</li> <li>Based on total of teachers for</li> </ul>
		particular streaming
	A student's performance manually be linked to the streaming process	- Based on total of students
	To identify student's streaming, school use final examinesult, student	Strategies:     Einel according
	request and man request and final exam result	- Final exam result
		- Student request
	Straaming Process:	<ul> <li>Map request and final exam</li> </ul>
	1 The student gets final exam result	result
	2 The student will choose the area	Steps:
	3 Taachars will set the class based on final arom result total of	<ul> <li>The student gets final examine</li> </ul>
	alars and total of sequestby the students	<ul> <li>The student will choose the area</li> </ul>
	class and total of requestoy the students.	<ul> <li>Teachers will set the class bas</li> </ul>
	In helping students make decisions: Which class should they choose the	on final exam result and total o
	school provides guide based on current market needs and advice that man	class and total of requestby
	the streaming with their nerformance	students
	the streaming with their performance	<ul> <li>The school provides:</li> </ul>
	There have been cases of the class selection because following their friends	<ul> <li>A guide based on current man needs</li> </ul>
	Some of them select class because of their parents	<ul> <li>Advice that map the streamin with their performance</li> </ul>
	Some of them have nearesed themselves before the final eveningtion and	what alon performance
	borne of menninave prepared manselves before merinar examinatoriand	Factors influencing streaming
	has set a tai get to get into certain classes	<ul> <li>Friends</li> </ul>
	Student's performance capacitat school lovel is were important for school	<ul> <li>Parents</li> </ul>
	performance and to set the student's target	<ul> <li>Exam results</li> </ul>
	Student's nerformance report at MOE district and stata level is important	Problems:
	for school nerformance and for best students reward	<ul> <li>Manually</li> </ul>
		<ul> <li>Burden the teachers</li> </ul>
	School handle the student's streaming process by manually, at the same	<ul> <li>Confusing (different systems used</li> </ul>
	time burden the teachers. Teachers are confusing because different systems	every year)
	are applied for students' performance year by year. Teachers face difficulty	<ul> <li>Difficulty in planning (setting a</li> </ul>
	in planning for setting a student's target.	student's target)
	An electronic/computerized system that provides rules to automatically	Motivation for automatic information
	suggest streaming based on student's performance from the overall total of	system suggest streaming based on stud
	students will make the streaming process is more conducive because it can	performance:
	suggest for balance class distribution or streaming. It also will give less	<ul> <li>Can suggest for balance class</li> </ul>
	additional load for teachers. Besides, it is easy to manage student	distribution or streaming
		· · · · · · · · · · · · · · · · · · ·
	performance data and easy to keep track students' performance data.	<ul> <li>Less additional load for teachers</li> </ul>

#### 4. **Reviewing themes**

In this step, key quotations for each category that have developed earlier are associated using MS-Word. MS-Word is used to keep a record of the key quotations as shown in Table 5.3. The figure shows the association of key quotations taken from interview 149

transcripts to the categories of "student's academic performance" and "student's academic streaming school".

Themes	Sub-themes	Quotations
Student's academic	How it relates to the student's	Usually performance in final exam and subject's grade is considered
performance (RQ1)	streaming	for students to choose streaming. [T2]
		Student's streaming is done manually. There is no guideline to refer. Commonly, we are referring to the student's final exam result. <b>[15]</b>
		Besides, the final exam result, student's streaming is based on total of teachers for particular streaming area and based on total of students itself. Because usually, each school only can provide 1 or 2 classes for science streaming, the balance is put into art streaming. <b>[T6]</b>
		A student's performance is not linked to the student's streaming process. We are doing it manually. <b>[T6]</b>
		In order to identify a student's streaming, the school is using final exam result. Then the student will request for the particular class that they are interested in. Lastly, the request and the final exam result will be mapped to make sure that they qualify for the class. <b>[T7]</b>
		I believe that we need for a tool such as computerized system that can help for balance student's distribution. The tool should can provide rules to automatically suggest streaming based on student's performance from the overall total of students. <b>[T9]</b>
		An electronic/computerized system will make the streaming processis more conducive because it can suggest for balance class distribution or streaming. It also will give less additional load for teachers. At the same time, it is easy to manage student performance data and easy to keep track students' performance data [T10]
		The motivation to use the technology based system that can help to assist for student's streaming are; easy to understand and easy to report. [T11]

# Table 5.3 : Sample of key quotations for category of "student's academic

# performance"

## 5. Defining and naming themes

At this step, the early fifty three categories managed to be reduced into just two main themes by reorganizing and refining the categories into the finalized themes. Participant IDs were included in each category together with the selected quotes. In order to support the themes and the quotations, related literature and research questions were identified as well. Table 5.4 illustrates an example of a refined theme, description of the theme, sample quotations, research participant IDs, and research questions and/or related literature review.

# Table 5.4 : Refined Theme with Details Of Themes, Sub-Themes, Example OfQuotes And Research Questions / Related Literature Review

Themes	Sub-themes	Quotations	RQ / Related LR
Student's streaming	Student's streaming	The student's streaming process involves 3 main steps. Firstly, the student gets final examesult second they will choose the	RQ1 - How student's a cademic performance and student's streaming is applied school?
an occide of	process	area that they prefer. There are a lot of factors that influencing	sacanning is applied serioor.
		their decision. Lastly; the teachers will set their class or streaming based on their final examresult, total of class and request from the student. <b>[T3]</b>	Student's streaming considered about final exam and base on teacher recommendation (Esposito, 1973; Forgasz, 2010)
		In helping students make decisions; which class should they choose, the school provides guide based on current market needs and a dvice that map the streaming with their	Student's streaming is done based on achievement (Slavin, 1990)
		performance. [T4]	Students are grouped in particular streams, based on their interest and ability (Stevens & Vermeersch, 2010)
		There have been cases of the select class because following	
		their friends, some of themselect class because of their parents and a few of them have prepared themselves before the final examination. They have their own target to get into certain classes. [T12]	Student's streaming needed because it prepares students to find employment after they finish secondary education or prepare students for specific, more applied forms of higher education (Stevens & Vermeersch, 2010)
		Because school manually handles the student's streaming process, it was given an additional burden to the teachers.	
		Teachers are confusing because different systems are applied	
		for students' performance year by year, and when comes to streaming, they need to do it traditionally. Besides, the teachers are difficult to plan for a student's target TI 13	

#### 6. **Producing the report**

At the end of this step, after completing data coding and analysis, findings of the study are reported in Section 5.3. The proposed solutions to the problems that arise, discussed from the findings. Then, discussion based on the findings is offered in Chapter 6. As mentioned earlier in Section 4.1, the research questions are linked to the initial finding in the literature in Chapter 2, factors and concepts of TLT and ST. For data analysis, the interpretation of the data was informed by TLT and ST. The aim was to get as broad and rich explanation as possible from the research phenomenon. Therefore, although the theoretical constructs or factors informed the interpretation of the data in conceptual framework, the thematic analysis also be done. The categories and finalized themes generated from the data analysis are also depicted in Appendix F.

Factor influences academic streaming	T2	T3	T7	T11	T13
Guide	$\checkmark$		$\checkmark$	$\checkmark$	
Teacher Advice			$\checkmark$		
Parental pressure					
Self-concept	V	V	V		V
Aptitudes			$\checkmark$		
Parent's aspiration	V				V
Interest	$\checkmark$		$\checkmark$	$\checkmark$	V
Skills					$\checkmark$
Exam results			V	$\checkmark$	V
Friends			V		$\checkmark$

# Table 5.5 : A Summary of Findings from the Interviews Validated by the Respondents

Lastly, a summary of findings from the interviews are emailed back to the same research respondents to get their comment, for validity purposes (refer Table 5.5).

## 5.3 FINDING FOR INTERVIEW

Data collection 1 (interview) is done for answered the RQ 1 about *How student's* academic performance and the student's academic streaming is applied in school? Then, the student's academic performance and the student's academic streaming was viewed in depth in the interview to see the linkage.

Data collection 1 revealed two key findings, which are:

- Finding 1: Linkage between student's performance and student's academic streaming in school
- Finding 2: Factors influenced the student's academic streaming in school

Finding 1: Linkage Between Student's Performance and Student's Academic Streaming In School

The findings answer the research question of "how" student's academic performance and student's academic streaming is relate to each other in school, by looking at - how these two is linked to each other in an existing system. Table 5.6 summarizes the respond on the linkage among student's academic performance and academic streaming in the existing system. The relationships are described and the numbers of respondents are also summarized in the table.

 Table 5.6 : Linkage between Academic Streaming and Student's Academic

Relationship	Explanation	Respondent, N=17
Manual	student's academic performance and student's	T1, T2, T3, T4, T5,
	academic streaming are manually linked in order to	T6, T7, T8, T9,
	identify suitable academic streaming for the	T10, T11, T12,
	students	T13, T14, T15,
		T16, T17
Embedded in	student's academic performance and student's	
existing system	academic streaming are embedded electronically	NONE
	linked in existing system in order to identify	NONE
	suitable streaming for the students	

**Performance in the Existing System** 

The student's academic performance and academic streaming connected manually in the existing system. There is no system that connects both these processes electronically to the existing system. However, it was found that interaction between the student's performance data and student's academic streaming occurred manually throughout the process to make sure that the academic streaming is done correctly. The situation can be seen as the progress from student's academic performance to identify student's academic streaming, and at the same time the interaction occurs throughout the process.

This fact also supports the literatures that claim the most of the academic performance systems were run manually and unconnectedly to other related systems (Abdullah, 2006). The academic streaming process is happened outside the system, which is performed manually by the teachers, and used independent sources of data, not relate to plan for students' pathway (A. Z. Abdullah, 2006; Budhwar & Sparrow, 2002).

#### Finding 2: Factors Influenced the Student's Academic Streaming in School

This findings answer the question on how the student's academic streaming process is done in school and "what" are the important factors that influencing the student's academic streaming process.

There are five main factors which used in student's academic performance are:

- a) Education Institution factor Guide, Teacher Advice
- b) Peers factor Friends
- c) Family factor Parental pressure, self-concept, Aptitudes
- d) Historical trends factor Parent's aspiration, high expectation
- e) Employment market factor Interest, Skills, exam results

The analysis demonstrated a number of factors involved in student's academic streaming in school as shown in Figure 5.1. These factors are classified into several groups: Education Institution factor, Peers factor, Family factor, Historical trends factor and Employment market factor.



Figure 5.1 : Important Factors Involved in Student's Streaming in School

The Education Institution factor includes factors that come from the school that will influence the selection of streaming by the students. The Peers factor viewed the individual that encourages students to change their decision. Peers usually considered as people who have similar interests, age, background, and social status. The peers are likely to influence the person's beliefs and behavior. The Family factor relates to the items that come from the parents and students themselves in order to identify suitable streaming for them. The Historical trends factor relates to something that happened before that will influence students in their decision for academic streaming. Lastly, the Employment market factor relates to aspects that influence students to be marketed in the future. It is contextually relevant opportunities to get involved in certain areas. The following sub-sections discuss the detail of Finding 2:

a) Education Institution factor – Guide, Teacher Advice

The analysis of findings revealed similarities in terms of guidance and advice that is provided by the school in terms of students' academic streaming. Table 5.7 summarizes the response on education institution factor and described each of the important factors provide by the educational institution in order to assist students choose the right streaming. The numbers of respondents are also summarized in the table.

Table 5.7 : Education Institution Factors Involved in Academic Streaming in

School

Factors	Explanation	Respondent, N=17
Guide	Standard guidelines provided by the school to help	T1, T3, T4, T6, T7, T8,
	students in choosing academic streaming	T10, T11,T13
Teacher Advice	Advisory assistance provided by the teachers to	T1, T2, T3, T5, T7, T9,
	help students in choosing academic streaming	T12, T14, T15, T16

*Guide* - A standard guideline provided by the school, which will be given to the students on how their performance will determine the suitable area for them.

*Teacher Advice* - Student advisory assistance provided by the teachers (usually a counseling teacher) to help them in choosing an appropriate academic streaming.

Duru-Bellat & Mingat (1989), Bardick, Bernes, Magnusson, & Witko (2006), Designs (2005), Forgasz (2010), Gamoran (2002), Gulyaev & Stonyer (2002), Kemple & Snipes (2000), Kilgour (2008), Schneider, Judy, Ebmeye, & Broda (2014) and Zevenbergen (2001) believed that students are given less guidance from school and teachers on how to choose an appropriate academic streaming. In summary, most schools provide a basic guidance and advice to the students in helping them to choose a streaming, but the teachers are not given any guidance on how to determine the streaming that is appropriate for the students.

#### b) Peers factor – Friends

The analysis of the findings revealed that all respondents agree that friends play a role in influencing a student's decision to determine the streaming. This statement is based on the past experienced by the teachers. Table 5.8 summarizes the response on the peer factor and described the friends factors in influencing the academic streaming that will chose by the student. The numbers of respondents are also summarized in the table.

 Table 5.8 : Peers Factor Involved in Student's Academic streaming in School

Factors	Explanation	Respondent, N=17
Friends	Friends of the same age. Previously, study in the same	T1, T2, T3, T4, T5, T6,
	class, done many activities inside and outside the	T7, T8, T9, T10, T11,
	classroom together.	T12, T13, T14, T15,
		T16, T17

In summary, close friends were among the important factors that will influence the students to choose a streaming. Usually, the decision on the academic streaming selection will be made by discussion with friends to ensure they can be in the same class.

c) Family factor – Parental pressure, self-concept, Aptitudes

The third groups of factors involved in the student's academic streaming in school were related to the family factors. There are three factors in this category which are parental pressure, self-concept and aptitudes.

Table 5.9 summarizes the response on the family factors and described each of the related factors in influencing the academic streaming that will be chosen by the student. The numbers of respondents are also summarized in the table.

Factors	Explanation	Respondent, N=17
Parental pressure	Directives issued by the parent, and unable	T1, T3, T4, T8, T9,
	to be avoided.	T10, T11, T12, T13,
		T14, T15, T16
self-concept	The image of ourselves. This self-image	T1, T2, T3, T4, T5, T6,
	form and change over time. This image	T12, T13, T14, T15,
	particularly influenced by interactions	T16, T17
	with important people in our lives. Future	
	image of ourselves will influence the	
	decisions made; in the context of selecting	
	appropriate streaming.	
Aptitudes	A component of competence to perform	T2, T3, T4, T6, T7, T9,
	certain types of work, which can also be	T10, T11, T12, T14,
	regarded as a talent - innate or acquired	T15, T17
	capacity for something. Relates to	
	capability and ability	

 Table 5.9 : Family Factor Involved in Academic Streaming in School

The analysis of findings revealed a balance agreement between the respondents involved, in terms of factors that contributed by the family in influencing the academic streaming chosen by the student. Thirteen respondents agreed that the decision of the academic streaming selection is caused by force from their parents. This is because the parents have been designing the future of their children. Logically, the suggestion from parents is indisputable, because they have a strong influence on their children. Whereas, twelve respondents agreed that self-concept and aptitudes also plays a role in determining the students' academic streaming.

In summary, within the family factors, the participating respondents were agreed that student's academic streaming selection was influenced by parental pressure, self-concept and aptitudes. Study shows that parents' decision are precise and they are equally able to raise important concerns regardless of differences in education (Glascoe, 2000). Parents' concerns can be stimulated quickly and 92% of parents can react to the questions about their children (Glascoe, 2000). Furthermore, parents have a high sense of concern in their children's education and they are prepared to sacrifice for their children's education (Lam et al., 2012).

d) Historical trends factor - Parent's aspiration

Another group of factors involved in the student's academic streaming in school was related to the historical trends factor, which is relates to parent's aspiration. The parent's aspiration means, "How the performance of the parents in their life would have an influence on their children's decision". This including their achievement in life, places and perceptions of parents towards the educational attitudes. Table 5.10 below summarizes the response on the historical trends factor and described parent's aspiration in influencing the academic streaming that chosen by the student. The numbers of respondents are also summarized in the table.

 Table 5.10 : Historical Trends Factor Involved in Academic Streaming in

Factors	Explanation	Respondent, N=17
Parent's aspiration	The influence of parents on their children's	T4, T6, T7, T8, T9,
	decision, in terms of their achievement in	T11, T12, T13, T14,
	live, places and perceptions of parents on	T16, T17
	educational attitudes.	

Eleven out of seventeen respondents have agreed that students choose a particular academic streaming as influenced by their parents' lives now. Most of them make a success of their parents as a guide to choosing a career in the future. In literature, previous studies from Duru-Bellat & Mingat (1989), Bardick, Bernes, Magnusson, & Witko (2006), Designs (2005), Forgasz (2010), Gamoran (2002), Gulyaev & Stonyer (2002), Kemple & Snipes (2000), Kilgour (2008), Schneider, Judy,

Ebmeye, & Broda (2014) and Zevenbergen (2001) also concluded that in determining the future of students in academic, among factor involved are parents' achievement.

e) Employment market factor – Interest, Skills, exam results

For the employment market factor, the analysis of findings revealed a similarity provided by most respondents in the use of interest, skills and exam results in determination of students' academic streaming. Table 5.11 below summarizes the respond on the Employment Market factors and described each of the related factors in influencing the academic streaming that will be choose by the student. The numbers of respondents are also summarized in the table.

 Table 5.11 : Employment Market Factor Involved in Academic Streaming in

Factors	Explanation	Respondent, N=17
Interest	A condition of inquisitive or relates	T2, T3, T4, T5, T8, T9,
	to attention to something	T10, T11, T12, T13, T14,
	D	T15, T16, T17
Skills	is the knowledgeable capability to	T1, T2, T3, T4, T5, T6,
	convey a task with pre-determined	T7, T8, T10, T11, T13,
	outcomes	T14, T15, T16, T17
exam results	Relates to student's performance,	T1, T2, T3, T4, T5, T6,
	the marks achieved for the	T7, T8, T9, T10, T11,
	particular examination	T12, T13, T14, T15, T16,
		T17

School

The analysis of findings revealed almost a same agreement between the respondents in the employment market factors that influenced the academic streaming selection. Fourteen respondents agreed that the decision of the suitable streaming is caused by the student's interest in a particular area. Fifteen of them agreed that the student's skills are decisive in specifying the appropriate academic streaming for students. Whereas, all the respondents agreed that the performance of the students in academic which is their exam result plays a most important a role in determining the suitable academic streaming for the students.

In summary, within the employment market factors, the participating respondents were agreed that suitable student's academic streaming were influenced by the student's interest, skills and exam results.

Based on these two findings, it can be concluded that currently, the student's academic streaming process in not linking to the student's performance in school. These two processes are done separately and conducted not systematically. Besides, there are five main factors identified that influence the student's academic streaming in school, which are Guide, Teacher Advice, Friends, Parental pressure, self-concept, Aptitudes, Parent's aspiration, Interest, Skills, and exam results. Since all these factors have been identified, hopefully problems involving this factor can be handled thoroughly.

#### 5.4 DATA ANALYSIS FOR SURVEYS

For data collection 2 (surveys), the unit of analysis is the student, with a total of 465 students from 5 schools all over Malaysia. In order to analyze the surveys data, PLS is used to find the importance of each factors identified earlier in literature. This kind of analysis provides the factors that can be used in the proposed framework.

Once the feedbacks of the surveys are obtained, the data were key-in into a Microsoft Excel to create .cvs file which will be used in the smartPLS. SmartPLS is software used to run the PLS-SEM analysis. Based on the proposed conceptual framework; which combines the concepts from the literature, Teaching / Learning Theory (TLT) and as

well as System Theory (ST), 13 factors were identified. All the factors have their own factor loadings that labelled A-M as shows in Table 5.12. The labelling is based on the surveys questions (Appendix D).

Label	Factors	<b>Factor Loading</b>
А	Parental and Peer Pressure	A1 – A6
В	High expectations	B1-B4
С	Multiple Career Options	C2-C4
D	No Guide	D1, D3, D4
E	New Industry Trends	E1, E2, E4
F	Exam result	F1-F4
G	Teacher Advice	G1-G3
Н	Parent's Aspiration	H1-H3
Ι	Friends	I1-I4
J	Self-concept	J1-J4
K	Aptitudes	K1,K2,K4
L	Interest	L2-L4
М	Skills	M1, M3, M4

Table 5.12 : Label for Factor Loadings



**Figure 5.2 : Factors Involved in the Research** 

For validity purpose, the study needs to identify factors that should be measured. Figure 5.2 shows all the factors that are considered in this study based on literature, ST and TLT. In this research, there are 13 factors from 5 different group of important factors need to be measured. Each factor as shown in Figure 5.2 considered is able to influence student's academic streaming process. Each factor has its own loading which are used to evaluate the factor. These loadings are obtained based on the surveys question.

After the PLS algorithm implemented, the below diagram shown in Figure 5.3 is produced in order to represent the importance of each factor and the factors loading.



**Figure 5.3: Factor Loadings** 

This surveys is necessary to ensure reliability or the extent of its usability. Reliability is aligned with consistency of our measurement. This is to ensure questions used in a surveys produce the same type of feedback under identical situations. Reliability is also interrelated to internal uniformity, which discusses to the level of dissimilar questions measure the same items. In this study, reliability is obtained based on the data used; factors used are reliable to be applied in this study as per proposed in conceptual framework, as they are meeting a good indicator scale:

Loading,  $r^2$  is between 0.5 and 0.7 AVE(R) > 0.5CR > 0.7

Based on the indicator scale, it found that all the factors involved are influence the academic streaming process. However, there are some factor loadings should be eliminated because they measuring a similar thing with other factor loadings. Therefore, they are not taken into account, but this does not affect the overall analysis.

#### 5.5 FINDING FOR SURVEYS

The findings start with the description of the significant relationship between the demographic profile and the problems faced by students in choosing the suitable academic streaming. Then, a description of the data obtained for the study followed by analysis using PLS to identify the importance of each factor.

Table 5.13 shows the demographic information of students for surveys. The demographic information are divided into three parts: Gender, family's head occupation and Race. This is to see whether the demographic information are influenced the academic streaming selection.

<b>Demographic Profile (n=46</b>	Responses N (%)					
Gender	Male	160 (34.4%)				
	Female	305 (65.6%)				
Family's head occupation	Employed,	269 (58%)				
	Self-employed	186 (40%)				
	A pensioner	7 (1.5%)				
	Not working	3 (0.5%)				
Race	Malay	(77.6%)				
	Chinese	(18.7%)				
	Indian	(3.7%)				

Table 5.13 : Demographic Information of Students for Data Collection 2 (Surveys)

Chi-square tests were performed to examine the relation between demographic profiles and problems faced by students in choosing the right academic streaming.

# • Relation between problems and gender

There is no significant relationship between problems and gender,  $\chi^2(4, N = 465) =$ 

5.32, p = 0.26.

# • Relation between problems and family's head occupation

There is no significant relationship between problems and family's head occupation,

 $\chi^2$  (12, N = 465) = 9.24, p = 0.68.

#### • Relation between problems and race

There is no significant relationship between problems and race,  $\chi^2$  (12, N= 465) = 4.18, p = 0.38.

Based on the results, it is found that there is no significant relationship between demographic profiles and the problems faced by students in choosing the suitable academic streaming.

In the first part of the surveys questions, after the demographic profile, there is the question regarding the level of difficulty to make decisions in determining the suitable academic streaming (refer Appendix D). The question is intended to determine whether respondents can realize their problem while choosing the suitable academic streaming. The findings revealed that the majority (60%) of the respondents had *strongly agreed* that it is difficult for them to choose the suitable academic streaming based on their potential (Figure 5.4). Besides, 25.2% are agreed, 10.3% are natural and 3% are disagreed with the statement.



**Figure 5.4 : Difficulties in choosing suitable academic streaming** 

There were only 7 respondents which is represent 1.5% of the respondent had *strongly disagreed* that it is difficult for them to choose the suitable academic streaming based on their potential. Most of the respondents said that it is difficult for them to choose the suitable academic streaming based on their potential because they do not know the requirement for the particular academic streaming, because the requirement is always changing. Besides, they are also not get detail exposure about the requirement for the particular academic streaming. This causes them to make decisions based on other external factors. This is demonstrated in Figure 5.5. The figure shows that most of the  $\frac{167}{100}$ 

respondents which are 63% of them reported that they do not know the requirement for the particular academic streaming, because the requirement is always changing. So they do not know the standard requirement.



Figure 5.5 : Knowledge about the requirement for the particular academic streaming

For second part in the questionnaire (Refer Appendix D) the analysis is done using PLS. Analysis using PLS is to show the convergent and discriminant validity of the constructs or factors. Model measurement is done using PLS, in order to specify the relationship between constructs or factors involved. Table 5.14 shows the model measurement for this study.

Construct	Item	Loading	AVE	CR
Parental + peer pressure	A5	0.611		
	A6	0.419	0.515	0.822
High expectations	B1	0.656		
	B2	0.515		
	B3	0.842	0.671	0.846
Multiple Career Option	C3	0.573		
	C4	0.561	0.567	0.827
No guide	D1	0.510		
	D3	0.530		
	D4	0.847	0.629	0.824
New industry career trends	E2	0.890		
	E4	0.562	0.726	0.753
Exam result	F1	0.522		
	F2	0.807		
	F3	0.546	0.625	0.862
Teacher advice	G1	0.793		
	G2	0.794		
	G3	0.802	0.796	0.839
Parent's aspiration	H1	0.817		
	H2	0.641	0.729	0.894
Friends	I1	0.657		
	I2	0.695		
	13	0.998		
	I4	0.527	0.719	0.820
Self-concept	J1	0.502		
	J2	0.526		
	J3	0.952		
	J4	0.954	0.733	0.866
Aptitudes	K1	0.538		
	К2	0.530		
	K4	0.723	0.598	0.880
Interest	L2	0.986		
	L3	0.547		
	L4	0.527	0.687	0.837
Skills	M3	0.610		
	M4	0.648	0.629	0.802

 Table 5.14 : Model Measurement for Data Collection 2 (Surveys)

Note: A1, A2, A3, A4, B4, C1, C2, D2, E1, E3, F4, K3, L1, M1 and M2 were deleted due to low loading.

Fornell and Larcker (1981) suggest using the average variance extracted (AVE) as a criterion of convergent validity. In an AVE analysis, the test is run to see if the square root of every AVE is much larger than any correlation among any pair of latent constructs.

In summary, all the measurements of latent variables or factors are meeting all the criteria. All the AVE values in this study are more than 0.5 indicates sufficient convergent validity, meaning that a latent variable is able to explain more than half of the variance of its indicators on average (e.g., Go<sup>--</sup>tz, Liehr-Gobbers, & Krafft, 2009). Besides, the AVE of each latent variable is greater than the latent variable's highest squared correlation with any other latent variable.

For the discriminant validity, it is prerequisite for analyzing relationships between latent variables (Henseler, Ringle, & Sarstedt, 2014). For the discriminant validity, the value is shown in Table 5.15. The criterion of the discriminant validity is usually a bit more liberal. The loading of each factor is expected to be greater than all of its cross-loadings (Chin, 1998; Go<sup>--</sup>tz et al., 2009). In order to ensure the discriminant validity, the AVE of each latent variable should be higher than the squared correlations with all other latent variables. Thereby, each latent variable shares more variance with its own block of indicators than with another latent variable representing a different block of indicators.

<b>Table 5.15:</b>	Discriminant Validity	
	·	

	Aptitudes	Exam result	Friends	High expectations	Interest	Multiple Career Option	New industry career trends	No guide	Parental + peer pressure	Parent's aspiration	Teacher advice	self- concept	skills
Aptitudes	0.773												
Exam result	0.473	0.790											
Friends	0.476	0.934	0.848										
High expectations	0.517	0.609	0.511	0.819									
Interest	0.450	1.000	0.935	0.793	0.829								
Multiple Career Option	0.522	0.859	0.518	0.298	0.881	0.753							
New industry career trends	0.528	0.307	0.437	0.648	0.296	0.437	0.852						
No guide	0.519	0.517	0.403	0.936	0.513	0.624	0.590	0.793					
Parental + peer pressure	0.453	0.521	0.386	0.381	0.520	0.616	0.371	0.305	0.717				
Parent's aspiration	0.752	0.512	0.688	0.342	0.511	0.716	0.404	0.422	0.809	0.631			
Teacher advice	0.537	0.844	0.479	0.803	0.834	0.891	0.508	0.305	0.428	0.762	0.892		
Self-concept	0.893	0.596	0.529	0.521	0.569	0.601	0.406	0.630	0.481	0.385	0.305	0.856	
Skills	0.512	0.535	0.372	0.881	0.340	0.544	0.641	0.559	0.580	0.362	0.747	0.518	0.793

 Skills
 0.512
 0.535
 0.372
 0.881
 0.340
 0.544
 0.641
 0.5

 Note: Diagonals represent the square root of the AVE while the off-diagonals are correlations.

#### 5.6 DATA ANALYSIS FOR CONTENT ANALYSIS

This analysis is conducted on student's academic performance data, to find a clustering technique that suited with the data used.

In this study, a student's performance data involves 20 available variables involving data for *Bahasa Melayu* (BM), English (BI), Mathematics and science subjects. It involves 5 tests for each subject throughout the year. These data are in the form of grades converted into numerical form based on the implementation of new grade of School Based Assessment (PBS).

Data were tested with IBM SPSS Modeler and three clustering models have been suggested based on the data used. In order to find the most appropriate clustering model, all these three models then be used to analysis the same data, and the result is compared. These models are i) Twostep; ii) Kmeans; and iii) Kohonen. Each model has certain strengths and needs to be considered as the most suitable model for the data. Numbers of clusters are not set up to ensure that the model can produce appropriate natural cluster based on the data. The result of this analysis shows three main information from the data, specifically; model summary, cluster size and predictor importance:

• *Model Summary* – The model summary shows a cluster quality of the model, including a silhouette measure of cluster cohesion and separation that is shaded to indicate poor, fair, or good results. The results of poor, fair, and good are based on the work of Kaufmann & Rousseeuw (1990) regarding interpretation of cluster structures. In the model summary view, a good result equates to data that reflects as either reasonable or strong evidence of cluster structure, fairly reflects their rating of weak evidence, and poor reflects their rating of no significant evidence.

The silhouette measure averages, over all records, (B-A) / max (A, B), where A is the record's distance to its cluster centre and B is the record's distance to the nearest cluster centre that it doesn't belong to. A silhouette coefficient of 1 would mean that all cases are located directly on their cluster centres. The scale used is between -1 and 1. Value of 1 indicates good quality, whereas, value of -1 indicates poor quality. A value of -1 would mean all cases are located in the cluster centres of some other cluster. A value of 0 means, on average, the cases are equidistant between their own cluster centre and the nearest other cluster.

- *Cluster size* Clusters have widely differing sizes, and shapes, and when the data contain large amounts of outliers. Cluster Sizes indicated how many clusters produced by a particular model. Cluster size is also shown in the form of a size for each cluster produced and also its percentage. Each cluster naturally produced based on the student's performance data. The percentage size of each cluster is shown on each slice.
- *Predictor importance*–Predictor importance is calculated from the test partition. The predictor importance is useful in fine-tuning the model. It specifies the relative importance of each input for a particular model, whereas interactions and correlations are taken into account. Since the values are relative, the sum of the values for all predictors is 1.0. Predictor importance does not relate to model

accuracy. It just relates to the importance of each predictor in making a prediction, and not based on the accuracy of the prediction.

Explanation for model summary, cluster size and predictor of importance for each model (Twostep, K-means and Kohonen) will be described further in the following section (5.61 - 5.6.3) and the analysis for these three models presented in Section 5.7.

#### 5.6.1 TwoStep

This model combines the ability of the K-Means clustering to handle a very large dataset, and the ability of the Hierarchical clustering (HCA – Hierarchical Cluster Analysis) to give a visual presentation of the results.

The TwoStep model is an exploratory tool designed to disclose natural groupings or clusters within a dataset that would otherwise not be apparent. The algorithm employed by this model has several desirable features that differentiate it from traditional clustering techniques:

- Handling of categorical and continuous variables. By assuming variables to be independent, a joint multinomial-normal distribution can be placed on categorical and continuous variables.
- Automatic selection of number of clusters. By comparing the values of a model-choice criterion across different clustering solutions, the procedure can automatically determine the optimal number of clusters.
- **Scalability.** By constructing a cluster feature (CF) tree that summarizes the records, the TwoStep algorithm allows large data files to be analyzed.

The TwoStep clustering procedure consists of the following steps:

#### 1. Pre-clustering

Pre-clusters are the clusters of original cases/objects that are used in place of raw data to reduce the size of the distance matrix between all possible pair of cases. After completing the pre-clustering, the cases in the same pre-cluster are treated as a single entity (Verma, 2013). Thus, the size of the distance matrix depends upon the number of pre-clusters instead of cases. Hierarchical clustering method is used on these pre-clusters instead of the original cases.

The procedure is implemented by constructing a modified cluster feature (CF) tree. The CF tree consists of levels of nodes, and each node contains a number of entries. A leaf entry (an entry in the leaf node) represents a final sub-cluster. The non-leaf nodes and their entries are used to guide a new record quickly into a correct leaf node. Each entry is characterized by its CF that consists of the entry's number of records, mean and variance of each range field, and counts for each category of each symbolic field. For each successive record, starting from the root node, it is recursively guided by the closest entry in the node to find the closest child node, and descends along the CF tree. Upon reaching a leaf node, it finds the closest leaf entry in the leaf node. If the record is within a threshold distance of the closest leaf entry, it is absorbed into the leaf entry and the CF of that leaf entry is updated. Otherwise it starts its own leaf entry in the leaf node. If there is no space in the leaf node to create a new leaf entry, the leaf node is split into two. The entries in the original leaf node are divided into two groups

using the farthest pair as seeds, and redistributing the remaining entries based on the closeness criterion.

If the CF tree grows beyond allowed maximum size, the CF tree is rebuilt based on the existing CF tree by increasing the threshold distance criterion (Zhang et al., 1996). The rebuilt CF tree is smaller and hence has space for new input records. This process continues until a complete data pass is finished.

All records falling in the same entry can be collectively represented by the entry's CF. When a new record is added to an entry, the new CF can be computed from this new record and the old CF without knowing the individual records in the entry. These properties of CF make it possible to maintain only the entry CFs, rather than the sets of individual records. Hence the CF-tree is much smaller than the original data and can be stored in memory more efficiently.

The structure of the constructed CF tree may depend on the input order of the cases or records. To minimize the order effect, randomly order the records before building the model.

#### 2. Outlier handling

An optional outlier-handling step is implemented in the algorithm in the process of building the CF tree. Outliers are considered as data records that do not fit well into any cluster. Data records in a leaf entry is consider as outliers if the number of records in the entry is less than a certain fraction (25% by default) of the size of the largest leaf entry in the CF tree. Before rebuilding the CF tree, the procedure checks for potential outliers and sets them aside. After rebuilding the CF tree, the procedure checks to see if these outliers can fit in without increasing the tree size. At the end of CF tree building, small entries that cannot fit in are outliers.

3. Clustering

The cluster step takes sub-clusters (non-outlier sub-clusters if outlier handling is used) resulting from the pre-cluster step as input and then groups them into the desired number of clusters. Since the number of sub-clusters is much less than the number of original records, traditional clustering methods can be used effectively. TwoStep uses an agglomerative hierarchical clustering method, because it works well with the auto-cluster method.

Hierarchical clustering refers to a process by which clusters are recursively merged, until at the end of the process only one cluster remains containing all records. The process starts by defining a starting cluster for each of the subclusters produced in the pre-cluster step. All clusters are then compared, and the pair of clusters with the smallest distance between them is selected and merged into a single cluster. After merging, the new set of clusters is compared, the closest pair is merged, and the process repeats until all clusters have been merged. (If you are familiar with the way a decision tree is built, this is a similar process, except in reverse.) Because the clusters are merged recursively in this way, it is easy to compare solutions with different numbers of clusters.



Figure 5.6 : Cluster Size and the Cluster Quality Using Twostep Technique

Based on the data, the first results to be presented are the cluster size and the cluster quality. The cluster size and the cluster quality for this method are shown in Figure 5.6. The cluster size is 2 and the cluster quality is 0.413.



**Figure 5.7 : The Cluster Size Produced by Twostep Technique** 

Next, detailed information about the cluster size is shown in Figure 5.7. The figure shows the cluster size, size of smallest and largest cluster and ratio size of the cluster.

Size of the smallest cluster is 16 which are 23.9% of the cluster size, and size of the largest cluster is 51 which is 76.1%. Whereas, the ratio of size of the largest cluster to the smallest cluster is 3.19.

Next, Figure 5.8 presents the predictor importance for the TwoStep. The figure shows relative importance of each field in estimating the model.



**Figure 5.8 : The predictor importance for the TwoStep** 

The predictor importance is, sc4 (science subject for test no 4) because it is seen as an important predictor compared to other subjects.

#### 5.6.2 K-means

The second model used is K-means. K-Means is one of the simplest unsupervised learning models used for clustering. It is the most widely-used, simple and well-known clustering algorithm, and in this research it aims to partition n instances into k clusters in which each instance belongs to the cluster with the nearest mean.

The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point, k new centroids need to re-calculate as barycenters of the clusters resulting from the previous step. After the k new centroids are calculated, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop, k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. The main idea is:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| \boldsymbol{x}_{i}^{(j)} - \boldsymbol{c}_{j} \right\|^{2}$$

Where  $\|x_i^{(j)} - c_j\|^2$  is a selected distance measure between a data point  $x_i^{(j)}$  and the cluster centre  $c_j$ , is an indicator of the distance of the *n* data points from their respective cluster centres.

The algorithm is followed the four steps below:

- Put k points into the space represented by the objects that are being clustered.
   These points represent initial group centroids.
- 2. Allocate each object to the group that has the closest centroid.
- 3. When all objects have been allocated, recalculate the positions of the k centroids.
- 4. Repeat Steps 2 and 3 until the centroids no longer change. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.


Figure 5.9 : Cluster Size and The Cluster Quality Using K-means Technique

Using k-means, the first results to be presented are the cluster size and the cluster quality. Cluster size and cluster quality for this model are shown in Figure 5.9. Cluster size is 5 and cluster quality is 0.282.



Figure 5.10 : The Cluster Size Produced by K-means Technique

Next, detailed information about cluster size is shown in Figure 5.10. The figure shows cluster size, size of smallest and largest cluster and ratio size of the cluster. Size of the  $^{181}$ 

smallest cluster is 3 which is 4.5% of the cluster size, and size of the largest cluster is 23 which is 34.3%. Whereas, the ratio of size of the largest cluster to the smallest cluster is 7.67.

Next, Figure 5.11 presents the predictor importance for the K-means. The figure shows relative importance of each field in estimating the model.



Figure 5.11 : The predictor importance for the K-means Technique

The predictor importance is, sc2 (science subject for test no 2) because it is seen as an important predictor compared to the other subjects.

## 5.6.3 Kohonen

The last model used is Kohonen. It is used because of its ability in mapping highdimensional data to lower dimensions. It is applicable to exploratory data analysis. Its combination with U-Matrix method will help to detect data clusters much easier than most classical methods. The basic idea of Kohonen comes from knowledge, that brain uses inner space data image for storing an information. At first, data received are converted to vectors which are encoded to the neural network. The output from such neural network is geometrically prepared to some arrangement, e.g. abreast, or to the rectangle and thus there is a possibility of neighbor identification. This layer is called Kohonen layer. Number of inputs entering to the network is equal to the input space dimension. Principle of cluster analysis is the ability of the algorithm in the kohonen model to set respective neurons to the clusters of submitted patterns and thus to distribute submitted patterns into the clusters.



Figure 5.12 : Cluster Size and the Cluster Quality Using Kohonen Technique

The first results to be presented using Kohonen model are cluster size and cluster quality. Cluster size and cluster quality for this model is shown in Figure 5.12. Cluster size is 11 and cluster quality is 0.273.

Next, detailed information about cluster size is shown in Figure 5.13. The figure shows cluster size, size of smallest and largest cluster and ratio size of the cluster. Size of the

smallest cluster is 1 which is 1.5% of the cluster size, and size of the largest cluster shows a big difference which is 14 or 20.9%. Whereas, the ratio of size of the largest cluster to the smallest cluster is 14.00.



Figure 5.13 : The Cluster Size Produced by Kohonen Technique

Next, Figure 5.14 resents the predictor importance for the Kohonen. The figure shows relative importance of each field in estimating the model.



Figure 5.14 : The predictor importance for the Kohonen Technique

The predictor importance is, sc5 (science subject for test no 5) because it is seen as an important predictor compared to the other subjects.

## 5.7 FINDING FOR CONTENT ANALYSIS

Based on the conducted clustering analysis, the findings discover a hidden patterns and relationships among the attributes in the educational data used. The result is summarized not only based on cluster distribution, but also clustering quality, importance of the clustering and ratio of clustering size.

It was found that the most important predictor for all the three models involved (Twostep, K-means and Kohonen) is science subject. This is because, from all the analysis that was done, it was found that the value of an important predictor for this subject is the highest when compared to the other subjects.

Besides, the analysis performed can also identify the best natural cluster model for the educational data. In this case, TwoStep provides natural cluster that is suitable for the data, which is two clusters; because the data needs to be grouped into two categories which is art streaming or science streaming. Whilst, K-means and Kohonen provide a larger cluster, which is; five clusters for K-means and 11 clusters for Kohonen.

Clustering Algorithm	Cluster Size	Clustering Quality	Importance	Ratio of Size
TwoStep	2	0.413	0.807	3.19
Kmeans	5	0.282	0.496	7.67
Kohonen	11	0.273	0.478	14.00

 Table 5.16 : Summary for the Three Models Used in the Clustering Analysis

Table 5.16 summarized all the three models used in the analysis in terms of cluster size, clustering quality, the importance of the cluster and ratio size of the cluster. The cluster quality of TwoStep is 0.413 which is close to fair quality, but the quality of K-means and Kohonen are nearly low which is 0.282 (K-means) and 0.273 (Kohonen). The scale used is between -1 and 1. Value of 1 indicates good quality, whereas, value of -1 indicates poor quality. A value of -1 would mean all cases are located in the cluster centers of some other cluster. Due to the number of clusters produced by the K-means and Kohonen, the ratio of the cluster size is also high. This is because the calculation of the ratio of cluster size is made based on the largest cluster divided smallest cluster. The small ratio shows no significant difference in the distribution of the clusters. Among the three models analyzed, TwoStep shows the smallest value of ratio of size which is 3.19. While for K-means and Kohonen the ratio of clusters produced.

Based on the analysis, it is apparent that TwoStep is the most important model that suit for the educational data of students' performance with importance value is 0.807 compared to K-means (0.496) and Kohonen (0.478). This show that based on the student's performance, TwoStep is the most suitable model to group students into two categories or academic streaming class which is: art or science.

#### 5.8 SUMMARY OF FINDINGS

Findings derived from the results of the analysis of interview and surveys are combined for comparative purposes as a mixed strategy method that used embedded strategy involved qualitative data and quantitative data that needs to be linked during the analysis phase (Angell & Townsend, 2011; Harris & Brown, 2010).

Factors	Guide	Teacher Advice	Friends	Parental & peers pressure	Self- concept	Aptitudes	Parent's aspiration
No. of Quotations	9	10	17	12	12	12	11
Factors	High expectation	Multiple career option	New industry career trends	Interest	Skills	Exam Results	
No. of Quotations	1	2	2	10	13	13	

**Table 5.17 : Numbers of quotation linked to important factors** 

shows all the 13 factors linked to the total number of quotations in an interview. Most of the respondents mentioned about "friends" factor, "skills" factor and "exam result" factor during interview session. Based on the interview with teachers, their response indicated that most of the academic streaming cases occurring in school related to the influence of friends. Besides, "exam result" and "skills" play an important role as these two factors are taken into account in determining the academic stream. The other factors identified through the survey which is "High expectation", "Multiple career option", and "New industry career trends" only have 5 calling quotes respect of this factors; 1 for the "high expectation"; 2 quotes for "Multiple career option" and 2 quotes "Career to new industry trends". It is related to their high expectation for science stream and confusing with a lot of career available. In addition, the student is also less exposed to the new industry career trends, so they cannot associate academic streaming with their future career.

Table 5.18 shows the combined finding derived from surveys and interview. The finding shows that all the factors derived from the interview, which consists of 13 factors are, supported by the factor loadings because it exceeding more than 50. The

closer it is to 1 the better the items describe the factor's variance. The cut-off factor loading is the same as the indicator reliability, which the composite reliability's value suggests be reached at  $\geq 0.6$  for the data's exploratory and  $\geq 0.7$  for confirmatory features (Bagozzi & Yi, 1988; Hair, Ringle, & Sarstedt, 2011; Hair, Sarstedt, et al., 2011). Thus, the values of the indicator reliability will affect the value of composite reliability. Based on the Table 5.18, it indicates that each factor is affecting the academic stream and need to be considered when to make a decision.

Factors	Factor	Quotations
	loading	
Guide	.629	"We are always trying to help students make decisions; which class should they choose, what area that suitable
		for them" (T3)
Teacher Advice	.796	"Teachers, which is counsellor try to provide an advice
		to the students, usually based on current market needs
		and suggest them to map the available academic
		streaming with their performance" (T4)
Friends	.719	"A lot of cases select their academic streaming because
		tollowing their friends, they want to be together with their close friend in same class" (T7)
Parental & peers pressure	.515	"There have been cases of select class because following
		their friends, some of them select class because of their parents" (T7)
Self-concept	.733	"Of course there are students who choose a field based
-		on the success of the others family members. For
		example since childhood he/she is exposed to their
		uncle, who is a pilot" (T9)
Aptitudes	.598	"There are students who are really interested in their
		direction. I believe this is because their performance or talent was nurtured" (T11)
Parent's aspiration	.729	"Yes, I admit that the success of many parents affects
		children's actions and decisions. They said: I want to be
		a doctor because my dad is a doctor" (T12)
High expectation	.671	"They actually think science can only be a doctor,
		though there are many other careers related to science"
		(T13)
Multiple career option	.567	"most of the times students are confuse to choose
		streaming because there a lot of career outside there that
XY 1 1 1	726	they can choose" (113)
New industry career trends	.726	"Students are less expose to the current career trends.
Teternet	<b>C07</b>	I his is need to be improved (114)
Interest	.687	Sometimes student's interest is not in line with what
		they choose, pernaps because there are other more
Statle	620	strong influence, such as filends (113)
SKIIIS	.029	final examination. They have their own target to get into
* ~ \		certain classes" (T16)
Evan results	676	"In order to identify a student's streaming the school is
Exam results	.020	using final exam result. Then the student will request for
		the particular class that they are interested in Lastly the
		request and the final exam result will be mapped to
		make sure that they qualify for the class" (T16)

## Table 5.18 : Finding in Surveys and Interview (Quotations)



Figure 5.15 : Distribution of Related Quotations across Respondents

Figure 5.15 shows all the 13 factors are supported by qualitative (interview) and quantitative (surveys) evaluation. Most of the respondents for these data collections agreed that "exam result" is the most important factors that influence the academic streaming.

While "high expectation" is resulting as the lowest factor that contributing to the academic streaming process. This shows that the academic streaming process are relying heavily on the student academic performance.

## 5.9 SUMMARY

Through this chapter, the data analysis and findings for the study are presented. The interview is analyzed using thematic analysis, the surveys is analyzed using PLS, and for the content analysis, the documents are analyzed using IBM SPSS Modeler. Based

on the analysis, the finding shows three results, namely finding from the interview, surveys and content analysis. Then the finding from interview and surveys are combined together, since the study is using mix-methodology which is embedded strategy. In this strategy, the finding from qualitative and quantitative needs to be combined at the end of analysis stage.

The finding from the interview showed there is no linkage between the student's performance and student's academic streaming process, causing difficulty to identify individual potential and academic streaming process becomes complicated. The next finding from the surveys is confirmed the factors that have been obtained from the previous literature. These factors are affecting the student's academic streaming. Apart from that, the finding from content analysis shows that TwoSteps model is the appropriate model that can be used to assist academic streaming. Lastly, the finding from interview and surveys shows that the academic streaming process are relying heavily on the student academic performance, and all the factors that have been identified also supported in the interview and surveys.

191

## **CHAPTER 6**

## **DISCUSSION AND CONCLUSION**

## 6.1 INTRODUCTION

The final chapter of this study is a discussion of the findings, and how they can be used to reach conclusions, or lead to the formulation of recommendations. The four main search questions posed have been explored, and the results will enable an analysis to be made that reflects the research objectives that were defined for this study. To guide the process, the research problems are restated, in order to be compared with an overview of the various investigations which took place. This analysis will be based on the conceptual framework, which was designed to focus the direction of the research, and guide the contribution to the knowledge this study can make. The impact of academic streaming is examined, in terms of the theories that exist and the practicalities of actual implementation, including a discussion of using clustering models. During the investigations, a number of limitations were revealed, the background and effects of which will be indicated. The chapter will also include a number of recommendations concerning areas where additional research would be beneficial.

## 6.2 OVERVIEW OF THE RESEARCH

The overall aim of this study was to identify the way in which academic streaming was carried out in schools, taking into consideration whether this method of grouping students had any negative impacts. From these observations, it would potentially be possible to propose amendments to the practices and resolve any problems that may exist. The required approach to academic streaming is contained in the Malaysian Ministry of Education guidelines, and any proposals that might be made would be based on the scope of this existing resource.

The purposes of this search are threefold. First, it has investigated the current streaming practices for secondary school entry and the systems that are used to monitor academic performance. Second, the study has sought to identify the factors that need to be considered when applying streaming among students, in order to develop a streaming framework. Third, the identification of a suitable clustering model, capable of grouping students in a meaningful manner in accordance with academic performance and aptitude.

The decision to investigate this area in education management arose due to problems identified in the literature in the context of streaming. The issues pointed out included insufficient relevance of the streaming practices to actual academic performance, which marred identification of the real potential of an individual student. Concerns had also been expressed in the literature about the lack of systematic determination of a particular placement in a streaming system, due to insufficient understanding of the factors that influence how academic streaming should be carried out. Very few techniques for clustering students are being used in the educational environments. However, there is room for substantial improvement, based on applying data that reflects academic performance when implementing streaming.

The data collection process that was applied in this study was based on a perceived need to resolve the current limitations of information that is available. A mixed-method approach was used, which involved semi-structured interviews, surveys, and content analysis. The interviews were conducted with teachers, to obtain an insight into the

193

academic streaming being carried out in their particular school. The purpose of the surveys was to direct questions to students to understand if they experienced problems associated with streaming. Content analysis enabled the researcher to review and analyse data about streaming and the application of clustering, which could be used to support the formulation of a proposed clustering model and streaming framework. By gathering data using these methods, it was possible to obtain a better appreciation of the issues than would have been possible by using qualitative or quantitative methods in isolation.

Two key findings that resulted from the research are now discussed. They result from the qualitative approach used in the semi-structured interviews that were conducted.

a. Lack of a compelling linkage between student academic performance and placement in a streaming system was identified. This leads to the conclusion that streaming is not carried out systematically. Further analysis showed a need for robust measurements or methods that enabled leaders of schools to objectively assign students to either science or art curriculum streams. The decisions are currently more frequently made based on assessment by the teaching staff and the school, and an absence of standardized guidelines was noted. It was found that schools had a tendency to stream the students based on the availability of classrooms and the number of students, and not always through a systematic consideration of the aptitude and performance of an individual student. The final examination results were the only performance throughout the academic year. Science streaming are typically available to only about 20% of students,

which is dependent on the facilities that are available in a particular school and the teaching staff qualified to teach these subjects.

In addition to these practical problems, which affect considerations when determining streaming, appropriate assessment of the academic potential of a student is of equal importance. If the performance of individual students is not adequately monitored, the only measure of performance may be the general standard achieved by a class as a whole. This can lead to teaching interventions that focus on the classroom performance of just a few individuals and may neglect the learning needs of others. If the individual achievement of students cannot be identified, a rational process of improving performance and understanding difficulties is hard to implement.

- b. The second key finding involves the factors on which academic streaming of students in schools are taken. The five factors that influence decisions are:
  - Educational institution: guidance and advice from teachers;
  - Peers: influence of friends;
  - Family: parental pressure, self-concept, aptitudes;
  - Historical trends: aspirations of parents; and
  - Employment market: interests, skills, and exam results.

While through the surveys that was carried out it was found that there are also five groups of factors that influenced the students in choosing the appropriate academic streaming. The groups of factors are: Education Institution factor, Peers factor, Family factor, Historical trends factor and Employment market factor.

#### 6.2.1 Restatement of the Problem

A more detailed discussion of the results is undertaken in this section. It relates the study findings to the research questions which were posed for the purposes of this thesis. The outcomes are viewed in relation to theories, the literature on processes applied in academic streaming, and the techniques which are followed. For ease of understanding, the research questions and objectives are revisited, and the individual findings are discussed.

# Research objective 1 (RO1) - To investigate current practices applied in the academic streaming and student academic performance system in Malaysia.

A description of the interviews conducted has been provided in Section 4.2, while the Section 5.2 provide information concerning the practices that are applied in achieving this research objective. The conclusion is that student academic performance has an important impact on how academic streaming decisions are taken and these are discussed in conjunction with the research questions in Section 5.3.

*RQ1.* How are the academic performance and academic streaming of students being conducted in schools?

According to a review of the literature, the learning path or choice of academic streaming will typically be associated with student performance (Gage & Berliner, 1992). In their model, the researchers also state that decisions will be influenced by the understanding of the teacher concerning the characteristics and aptitudes of the student, as well as how best to encourage academic excellence.

The results of the interviews are presented in section 5.2.1 and concur with this statement, that streaming of students is based on academic performance. However, the respondents state that only the final exam is taken into account in the decision, and performance throughout the year is ignored. The analysis that results in a streaming decision is not currently a consideration of an entire school session. The results of this study show that the final exam result is the most important factor when deciding on the streaming of a particular student. However, in other education systems, the learning pathway has been influenced by other concepts that affect individual performance at the time of an examination (Patton & McMahon, 1998). By basing decision-making on only one event, it does not fully embrace the overall academic performance throughout the year and undervalues some aspects of the study a student undertakes. This conclusion was fully appreciated by the teachers who were interviewed.

Abdullah (2006) and Budhwar & Sparrow(2002) state that academic streaming takes place based on independent data sources which are not directly related to a particular student pathway that is offered, nor are decisions carried out based solely on teacher recommendation. The opinions expressed during the interviews pointed to a connection between academic performance and streaming in the current system. This study also found that although there are systems that report on student performance over several years, they are not linked to the academic streaming process. Currently, there is no technology available which can electronically connect historic and current academic performance data. The decisions on academic streaming are taken not only based on the opinions of teachers, there are some students who request consideration to be made of their preference for a particular academic pathway. When a request is made, the relevant teachers will consider the preference, based on final exam results. Research objective 2 (RO2) - To identify the factors involved in the academic streaming process and to recommend an academic streaming framework according to the identified factors

Several factors emerged during the surveys described in section 4.3 that have an influence on academic streaming, which were subsequently reviewed and confirmed. Five main factors were identified by students that affected their choice of stream, and these are discussed in conjunction with Research Question 2 in Chapter 2.

*RQ2.* What are the factors that influence the students in their choice of appropriate streams?

As discussed in sections 5.3 and 5.5, there are several considerations that take place when a student decides which academic stream they would prefer to follow. The choice is affected by influences that were described earlier, and involve the opinion of their school, peers, family expectations or traditions, globalization, career trends and employment prospects.

The factors are:

- a) Education Institution factor Guide, Teacher Advice
- b) Peers factor Friends
- c) Family factor Parental pressure, self-concept, Aptitudes
- d) Historical trends factor Parent's aspiration, high expectation
- e) Workplace multiple career option
- f) Globalization new industry career trends
- g) Employment market factor Interest, Skills, exam results

School will exert from the advice of teachers based on academic performance, and they will ultimately have the most impact on the choice of academic stream by students. The findings of this study indicate that teacher recommendation and student preference will lead to appropriate streaming. Previous research supports this assumption (Duru-Bellat & Mingat, 1989; Zevenbergen, 2001; Gulyaev & Stonyer, 2002; Chapman & Mählck, 2004; Bardick, Bernes, Magnusson, & Witko, 2006; Designs, 2005; Kilgour, 2008; Forgasz, 2010). However, it would appear that students do not think they receive adequate guidance from the educational institution on how to reach a decision on which academic stream would be most suitable. This was confirmed by the teachers interviewed, who stated that only basic guidance was given as there was insufficient time allocated to more comprehensive study and careers advice.

The influence of peers, relates to factors where individuals offer encouragement to students that can lead to a change of decision. The pressure from peers which can affect choices is acknowledged by many researchers (Gulyaev & Stonyer, 2002; Kemple & Snipes, 2000; Kilgour,2008;, Schneider, Judy, Ebmeye, & Broda, 2014; Zevenbergen, 2001). However, there is insufficient definition of the meaning of peers and the role they play in influencing students. This study takes the standpoint that peers are people who are a similar age, share common interests, and have the same background and social status. At school, peers are usually in the same class, their association is based on the connections described, but they may have very different academic abilities. The pressure that peers can exert is frequently sufficient to change a person's opinion and behaviour. Findings show peers can significantly influence decisions by students on streaming choices, because they want to continue to be in the same social group and continue important relationships.

The inspiration of the family may originate from the identification of a parent with the child and therefore they would naturally follow the same academic stream. In the literature, it is suggested that if a parent makes a suggestion, this is often not disputed and it has a strong influence on the child. In addition, many parents are concerned about their child's education and are prepared to make sacrifices to support successful outcomes (Lam et al., 2012). The involvement of parents in the studies of students has a significant and positive impact on development (Okagaki & Sternberg, 1993).

The questions posed in the student surveys related to parental pressure, self-concept, and aptitudes, which are factors that reflect the inter-relation between students and their family members. The results showed that familial impact does affect the choice of academic stream, as family members have close contact with students and these relationships will have considerable influence.

Educational traditions within a family can also exert influence. Historical events can lead to parents expressing aspirations for students to follow in their footsteps, which can in turn affect a student's decision on which academic stream to follow. If a student experiences that their parents have been successful, emulating the current success with cause students to select the same path, rather than taking an objective view of their own academic achievements and abilities.

The final factor that will influence student choice is the state of the employment market and the perception of jobs that will be available. The opportunities in some jobs areas are not always constant and students are usually aware of discussion about future prospects in careers, and this will influence the direction in which they want to invest their efforts. The factors that need to be considered in future career-directed choices are the interests of the student, the skills they are best suited to, as well as examination results. There is an abundance of career options that need to be considered, and this can put pressure on a student who must already make a choice at secondary level which will influence their future opportunities. Current and perceived future trends in career areas of interest will have an impact on the selection of which stream to take.

## **Research objective 3 (RO3) - To identify and use an appropriate clustering model** for the distribution of students according to their academic performance.

The investigation into various clustering models that can be applied in education was based on a review of the previous literature (Fathian & Amiri, 2007; Firouzi, Sadeghi, & Niknam, 2010; Kanungo et al., 2002; Kao, Zahara, & Kao, 2008; Krishna & Murty, 1999). The observation reached was that these studies do not explicitly state how the student data should be distributed to best achieve clustering. There are clustering models that have been created based on types of data, but selecting a particular model which is suitable for specific parameters is inconclusive. The aim of this study is to identify a model based on the academic performance of students.

Clustering models are used to establish natural groups that provide an accurate portrayal of the learning pattern of students and also support the underlying teaching processes and learning targets (Hämäläinen et al., 2013). From the content analysis described in section 5.6, clustering models was decided to be applied to academic performance data. The most commonly applied models are TwoStep, K-Means, and Kohonen. These three models were tested and the results are reported. The selection of these models was due to their extensive use and suitability in the context of education. Based on the analyses that were carried out, the TwoStep model was deemed to be the most appropriate, as it

combines the K-Means and Hierarchical Cluster Analysis models, thus provide natural cluster that can be used to support the requirement for the academic streaming by the MOE. This is discussed in conjunction with Research Questions 3–4.

- *RQ3.* What clustering models can be employed for the distribution of students into suitable streams?
- *RQ4.* What particular clustering model is appropriate for an academic streaming process that makes use of data on the academic performance of students?

Previous studies have confirmed that clustering is a suitable technique for use in educational analysis (Lopez & Luna, 2012; Oyelade et al., 2010; Romero, López, Luna, & Ventura, 2013; Tair & El-halees, 2012; Zhang et al., 1996). However, among the various methods that are available, there is no clear indication of which technique is best suited for particular types of data. This study found the existing academic streaming methods can be applied using clustering analysis with the appropriate selection of data and method. Clustering is a technique used in data mining, and it could be used to analyse data on a cohort of students based on performance throughout the academic year, and still adhere to the Ministry of Education guideline for distribution to be a ratio of 60:40 of science to art respectively. Using a broader representation of student performance, the data was found to be best suited for manipulation by the TwoStep model, compared to the K-Means and Kohonen methods. The two-step approach was able to modify the data into natural clusters, suitable for the educational data that was available. The natural cluster approach provides good parameters that reflect the levels of performance over longer periods of time, rather than simply the final examination. Two clusters were prepared, based on suitability for selection to the science or art stream, and the required 60:40 distribution was achieved.

## 6.2.2 Revisited the Conceptual Framework

The conceptual framework discussed in Chapter 2 (Section 2.10) integrated concepts from the literature and Systems theory (ST) and teaching and learning theories (TLT). The conceptual framework formed the basis on which the research was conducted and was used in the overall design, including data collection and subsequent analysis. As a result of the findings, some revision to the conceptual framework was indicated, as proposed in Figure 6.1. The revised framework for use in conjunction with academic streaming is composed of the following elements:

- 1. Data Source: Student performance during academic year;
- 2. Process: Clustering technique, using the TwoStep model;
- 3. Result: Student academic streaming (Student's academic pathway); and
- 4. Factors influencing the academic streaming framework:
  - a. Education Institution : Guide and teacher Advice
  - b. Peers: Friends
  - c. Family : Parental and peers pressure, self-concept and aptitudes
  - d. Historical trends: Parent's aspiration and high expectation
  - e. Workplace : Multiple career option
  - f. Globalization : New industry career trends
  - g. Employment Market : Interest, Skills and exam results
  - h. Community groups : Gender and Race
  - i. Socio economic status : Family head occupation



Figure 6.1 : Academic Streaming Framework

There were some minor adjustments in the revised conceptual framework based on the research findings and the addition of new factors that extended its scope. The framework was based on results or outcomes that involve the prediction of the academic pathway or stream in which students should participate.

However, the research findings shows that student's academic streaming is resulting from the important factor which exam results that are related to the academic performance of the students, so the academic performance of students throughout the year can be used to ensure that the student's academic streaming is accurate. The academic performance can be assumed as an input to the academic streaming process, while the method used to implement the academic streaming is clustering techniques. The clustering technique will processed the student's academic performance data n provide clusters or streaming based on the data. These three items; academic performance data, clustering technique and academic streaming seems heavily rely to each other. Due to this condition, these three items are placed in the center of the framework as input (academic performance data), process (clustering technique) and output (academic streaming).

The factors involved consists of several groups; Education Institution (Guide and teacher Advice), Peers (Friends), Family (Parental & peers pressure, self-concept and aptitudes), Historical trends (Parent's aspiration and high expectation), Workplace (Multiple career option), Globalization (New industry career trends), Employment Market (Interest, Skills and exam results), Community groups (Gender and Race) and Socio economic status (Family head occupation). The research findings confirmed the earlier identified factors. It also was found that interaction between student's performance data and student's academic streaming occurred throughout the process to make sure that the academic streaming is done correctly. The situation can be seen as the progress from student's academic performance to identify student's academic streaming, and at the same time the interaction occurs throughout the process.

## 6.2.3 Contributions

This research has made a number of contributions to the existing knowledge on the academic streaming process carried out in schools and the technology that can successfully be used to achieve a more systematic implementation. A summary of the various contributions is now discussed.

## Identification of important factors that influence the academic streaming process

The identification and testing of factors that affect academic streaming has been carried out to assess their importance. Accompanied by documentation, these factors could be introduced to the Ministry of Education, the States, and individual education districts, so they are able to review the proposals that are put forward for potential improvements in the existing systems that are used for academic performance monitoring and the criteria for streaming selection. An additional benefit is should there be problems associated with the factors that have been identified, they can be resolved or avoided.

# Insights and outcomes of the study on academic streaming processes reflect the perspectives of teachers and students

The link between data on student academic performance and decisions on academic streaming was clearly sustained throughout this research. The intention has been to examine the basis of decisions that affect the educational paths of secondary school students and provide a better understanding of performance data could be used to validate streaming decisions, from the perspectives of both the students and teachers. Currently, decisions are reached based on the results of the final examination and the performance data during the academic year is overlooked. The contention is that a more accurate assessment of academic performance would be possible by considering the wider scope of study that students have completed in the course of educational processes. Based on findings in this study, it would be beneficial to systematically link overall performance and academic streaming, potentially resulting in the development of a superior system, as the current method lacks the integration of different performance aspects.

## Analysis of the Selected Clustering Models

The current study considered clustering approach which has been discussed in the literature. Taking into account the techniques that are suitable for educational data, steps were taken to reach a decision on how best to use clustering to improve academic

streaming. Cluster sampling is a relatively straightforward approach when there are already natural clusters or groupings within a population, which are expected to affect the outcome of data analysis. As the basis of this study is to define groups, which in this case are proportions of students that will be selected to undertake a particular academic pathway, clustering is a suitable technique. The selection of a commonly used model makes the selection process more transparent, and there is clarity about how the model has been developed and on which data the model has been based. The TwoStep model was considered to be adequate for the purposes of this study on academic performance data, as it supports the development of natural clusters which would be suitable to adhere to the Ministry of Education guidelines on the 60:40 ratio of student assigned to science or arts-based streams.

## 6.3 IMPLICATIONS

Based on the development of the conceptual framework and the subsequent revision, the outcome of this study has brought additional aspects of knowledge into the forum, based on empirical investigation, related to academic performance and streaming decisions for studies at secondary school level. The following explanatory discussion covers the potential theoretical and practical implications of the research.

## 6.3.1 Theoretical Implication

Factors from Teaching and Learning theories suggested by Huitt (1995) have formed the basis of the theoretical perspective on which this research is based. The theories have been adapted from Systems theory, the founding of which is attributed to Bertalanffy (1969). His theory is concerned with establishing principles, which in this study have been concerned with understanding experiences associated with academic streaming that teachers and students have reported. Both TLT and ST have good alignment to the types of research required in this study, which have included data collection and analysis, to elucidate a potentially innovative basis on which to make streaming decisions. Systems theory has been used to differentiate various TLT aspects, in order to better understand the academic streaming process.

The conceptual framework was developed through the methodology of literature review, with additional consideration of ST and TLT constructs. A subsequent review led to the development of the revised conceptual framework. The integration of ST constructs and TLT enabled the researcher to improve the instrument, which supported more complex research needed on academic streaming. By linking the theoretical understanding with practical aspects of the research topic, the revised framework has furthered the understanding of the scope of student academic streaming.

The factors that affect streaming decisions were categorised in the revised framework as affects which emanate from the education institution, peers, family, historical trends, workplace, as well as the effects of globalisation, employment market fluctuations, the composition of community groups, and socio-economic status. When all the factors in TLT were applied to the academic streaming environment, useful insight was possible to critically examine the current streaming practices. Thereby, this research makes a valuable contribution to enrich the utilisation of TLT, which previously had only been applied in the context of education, but has successfully been adapted to understand academic streaming.

## 6.3.2 Practical Implications

The implications of this research on the practical aspects of streaming can be categorised as follows. There is a clearer understanding of the factors associated with academic streaming. A streaming framework and clustering model have been suggested for use in the academic streaming of students. The perspectives of both students and teaching staff have been analysed, in order to develop a more holistic approach. Experience has been further developed, on the basis of conducting evidence-based research.

## A clearer understanding of factors affecting streaming

The outcome of this study has resulted in a clearer understanding of the factors that are involved in processes to determine the streaming of students based on academic performance. The factors were derived by data collection and analysis.

## Proposal of a framework and clustering model for academic streaming

A framework that can be applied to academic streaming into secondary schools and a clustering model to assess data on student performance have emerged from this research. The input, process, and output, of the framework have been described and the TwoStep clustering model proposed as a means to achieve a more systematic approach to academic streaming.

## Considerations from the perspectives of both teachers and students

As decisions on streaming take into consideration the views of students and their teachers, this research has involved both types of respondent to investigate the research topic. Therefore, the findings reported in this thesis comprehensively represent the

views of the main stakeholders involved in academic streaming. To obtain a thorough understanding, it was important to gather information on the views of the stakeholders and to analyse and evaluate the issues from their perspectives, as well as from theoretical viewpoints. As teachers are involved in handling academic streaming, their views are important. However, as the student will ultimately be affected by the decision process, understanding their angle is equally valid. As academic performance is the basis of the decision process, students to a great extent are masters of their destiny. There are, however, a range of influences that may be encountered by students when they have to reach a decision on which stream is more suitable for them. A comparison has been made of the findings of this research and opinions expressed in previous research, in order to define to what degree the current outcomes are valid.

## Experience of conducting evidence-based research

The process followed, and the evidence that has been gathered, are important aspects of this research. They can contribute to the improvement of the level of understanding of all parties involved. By gathering data from teachers and students, the scope of the research has been broadened and a comparison of the different data sources has been possible. However, it is the data from both the teachers and students which results in empirical evidence of the views and opinions. This study has attempted to bridge theory and practice related to educational streaming, and literature review has been a useful approach to achieve this.

## 6.4 LIMITATION

A number of limitations, identified in the course of this research, should be acknowledged, which can be described as follows.

## Sample size

The interview sample size was limited to 17 respondents from different locations in Malaysia. However, although this might not be considered a representative sample, the study has generated useful data and better insight and understanding of secondary school streaming. It also provided a contrasting context that could be considered in conjunction with the surveys questionnaire that was carried out.

#### Clustering models used

It was decided to assess only three clustering models, rather than a more extensive selection. This model selection is done based on suggestion that is provided by IBM SPSS Modeler; based on data used. These three clustering models are widely accepted for used for the data.

## Focused of the study

The background of this study is focused on the performance of students once the education phase has been completed, and does not consider any aspects of teaching and learning. The educational processes are not within the scope and aims of the current research.

#### Academic streaming used

The data used are academic data streaming in secondary schools in Malaysia without any comparison with other academic streaming processes used in other countries. Besides, Only the science and art streams are considered in terms of the clustering process, as they are the main academic pathways in government schools in Malaysia. Normally in Malaysia, students will be divided into this two areas. However, there are additional streams, which include pure science, religion, religion as a profession, accounting, economics, and technical (drawing).

## 6.5 FUTURE RESEARCH

The overall contribution of the research could have been improved with an iterative feedback mechanism whereby the outcome of the clustering technique can be verified against the empirical study from the current practices and the improvements can be determined or measured. Apart from that, based on some of the limitations that have been identified, there are some areas in which future research would add value to the topic specifically under discussion:

## Additional clustering models

Investigation of other additional clustering models to be used with students' performance data. All the analysis of the models used then be compared to the two-step approach. The comparative analysis might yield results that could offer any interest to the research.

## Accuracy of the cluster

It would be desirable to consider other measures to determine the accuracy of the cluster contents itself. For example, precision and recall analysis of the cluster data will be usefull. It is also very important to evaluate the effectiveness of the output and to validate whether it actually produce accurate outputs.

## Include teaching and learning

Inclusion of aspects of the teaching and learning process in the academic streaming framework, because teaching and learning take place before the academic streaming. Besides, an additional sources of student performance data such as co-curriculum and discipline can also be used, to improve the accuracy of clustering.

## International comparison

More International comparison with other academic streaming processes used in other countries (e.g. Germany or Austria in Europe, and Japan and China in Asia). A comparison among these countries might make sense in order to find any differences on the process or procedure.

## Include other academic streaming

An extension of the scope of studies on clustering, to include other academic streams, apart from science and arts, that are also available in government schools such as pure science, religion, religion as a profession, accounting, economics, and technical (drawing). When all existing streams are taking into account, students will put extra focus in identifying streams that are appropriate with their performance.

#### 6.6 SUMMARY

The primary motivation for this research was to understand how student academic streaming was carried out, the impact of performance on the process for secondary school selection, and whether the selection methods could be simplified by data mining and clustering methods. The study was also concerned with factors that affect the preferences students express for a particular stream, and other prevailing influences that affect academic streaming.

One conclusion drawn from the research is that performance data is extensively used in the streaming process, in terms of the results of the final examination. However, this data is handled in isolation and not linked electronically to any other performance data system and therefore other aspects of performance are not considered in streaming decisions. Improvements in the accuracy of streaming could potentially be adopted, which use information technology to plot clusters of proficiency or aptitude. The distribution of students based on academic strengths could be carried out in a more systematic way. The potential of data mining and clustering have been identified as suitable methods to enable some degree of automation to be available to assist teachers in the task of academic streaming.

The outcome of this research is that the result of experiments on performance data did not show clustering tendencies that would be helpful in indicating which streams students were more academically suited to follow. One of the most significant findings to emerge from this research was confirmation of the factors that influence student academic streaming resulting from the two theoretical approaches used in the investigations. The Systems and Teaching and Learning theories were selected, based on a review of the literature and the particular nature of the study area.

The TwoStep clustering model was identified as the most suitable, as it involves both K-Means and Hierarchical Cluster Analysis. The data was embedded to produce natural clusters that could be used to support the distribution of students across the two main streams of science or arts. The modelling experiment aimed to group the students into two clusters, based on their academic performance, which could potentially be used as a basis for streaming. The clustering method would be a basis by which students would also be able to identify their academic strengths and which stream would be more suitable for them when starting secondary level education. Although the study was not conclusive, the aims and objectives were fulfilled and additional knowledge about the wider effects on streaming and potential improvements that may be reconsidered in the future, have been achieved.

## 6.7 CONCLUSION

The topic that has been addressed in this research can be considered innovative, as it sought to examine the academic streaming of students based on performance data and took into consideration both teacher and student perspectives on academic streaming decision-making. It was important to the researcher to involve both parties, as students are most affected by the academic streaming, and teachers, who are responsible for managing the process, see the benefits of the most suitable selection of study pathway being selected. Both parties play equally important roles in a decision that will have a considerable effect on the academic future of the students, if not their eventual career choices.

The major benefits of completing this research are the proposing of an innovative and alternative way to understand the potential use of student performance data, and whether by linking different aspects of electronically stored information, it would assist in the transparency and accuracy of the methods used to determine academic streaming in schools.
## REFERENCES

- Abdullah, A. A. (2006). Pembangunan Dan Kemajuan Modal Insan Dalam Rancangan Malaysia Kesembilan (RMK-9): Satu Penelitian, 1–21.
- Abdullah, A. Z. (2006). *History of Education. Educational Studies*. Kuala Lumpur. Retrieved from http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ042254
- Abusabha, R., & Woelfel, M. Lou. (2003). Qualitative vs quantitative methods: two opposites that make a perfect match. *Journal of the American Dietetic Association*, *103*(5), 566–9.
- Adams, R., & Biddle, B. (1970). The classroom scene. In *Realities of Teaching*. New York: Holt, Rinehart, & Winston, Inc.
- Adawo, M. A. (2011). Has education (human capital) contributed to the economic growth of Nigeria? *Economics and International Finance*, *3*(1), 46–58.
- Agarwal, S., Pandey, G., & Tiwari, M. (2012). Data mining in education: data classification and decision tree approach. *International Journal of E-Education, E-Business, E-Management and E-Learning*, 2(2), 140–144.
- Aggarwal, C. C., & Yu, P. S. (2001). Outlier detection for high dimensional data. *ACM SIGMOD Record*, *30*(2), 37–46.
- Ajith, P., Sai, M., & Tejaswi, B. (2013). Evaluation of Student Performance: An Outlier Detection Perspective. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 2(2), 40–44.
- Alfred, R. (2008). A Data Summarisation Approach to Knowledge Discovery (Doctoral dissertation). The University of York, United Kingdom.
- Al-Radaideh, Q., Al-Shawakfa, E., & Al-Najjar, M. (2006). Mining Student Data Using Decision Trees. In International Arab Conference on Information Technology (ACIT'2006). Jordan.

- Al-Radaideh, Q., & Nagi, E. Al. (2012). Using Data Mining Techniques to Build a Classification Model for Predicting Employees Performance. (IJACSA) International Journal of Advanced Computer Science and Applications, 3(2), 144– 151.
- Angell, B., & Townsend, L. (2011). Designing and Conducting Mixed Methods Studies.
- Ashton, P. (1984). Teacher efficacy: A motivational paradigm for effective teacher education. *Journal of Teacher Education*, *35*(5), 28–32.
- Awal, N. H. J. N. M., & Bakar, K. A. (2008). The Mastery of English Language among lower secondary school students in Malaysia: A Linguistic Analysis. *European Journal of Social Sciences*, 7–2.
- Axinn, W., & Pearce, L. (2006). *Mixed method data collection strategies*. Cambridge: Cambridge University Press.
- Aziz, Z. A. (2006). Focus on human capital development in Malaysia (pp. 1–2). Kuala Lumpur.
- Azlina, R. (2001). Sistem-Sistem Dalam Pengurusan Sekolah (pp. 1–18). Selangor, Malaysia.
- Bachtiar, F. a., Eric, C. W., & Kamei, K. (2011). Student grouping by neural network based on affective factors in learning English. *Proceeding of the International Conference on E-Education, Entertainment and E-Management*, 226–229.
- Bagozzi, R., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal* of the Academy of Marketing Science, 16(2), 74–94.
- Baradwaj, B. K. (2011). Mining Educational Data to Analyze Students' Performance. International Journal of Advanced Computer Science and Applications, 2(6), 63–69.
- Bardick, A. D., Bernes, K. B., Magnusson, K. C., & Witko, K. D. (2006). Junior High School Students' Career Plans for the Future: A Canadian Perspective. *Journal of Career Development*, 32(3), 250–271.

- Becker, G. S. (1993). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education* (3rd ed.). Chicago: University of Chicago Press.
- Bellin, N., Dunge, O., & Gunzenhauser, C. (2006). The importance of class composition for reading achievement: Migration background, social composition, and instructional practices An analysis of the German 2006 PIRLS data. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 3(1), 9– 34.
- Benson, L. (2002). Serving gifted students through inclusion: A teacher's perspective. *Roeper Review*, 24(3), 126–127.
- Berita Harian. (2010, June 23). Kementerian Pelajaran Tunggu Maklum Balas Pelbagai Pihak. *Berita Harian*. Kuala Lumpur.
- Berita Harian. (2012, December 21). 15 perakuan dilaksana tambah bilangan pelajar aliran sains: Muhyiddin. *Berita Harian*. Kuala Lumpur.
- Bertalanffy, L. Von. (1969). General System Theory: Foundations, Development, Applications (Revised Edition). New York. George Braziller Inc.
- Bhullar, M. S., Iaeng, M., & Kaur, A. (2012). Use of Data Mining in Education Sector, *I*, 24–27.
- Bian, H. (2010). Clustering Student Learning Activity Data. In EDM (pp. 277–278).
- Biddle, B., & Ellena. (1964). *Continued research on teacher effectiveness*. New York:Holt, Rinehart, & Winston, Inc.
- Biddle, B. J. (1967). Methods and Concepts in Classroom Research. *Review of Educational Research*, *37*(3), 337–357.
- Bloom, B. (1971). *Mastery learning. New York*. New York: Holt, Rinehart, & Winston, Inc.
- Brady, M., Clinton, D., Sweeney, J., Peterson, M., & Poynor, H. (1977). Instructional Dimensions Study (IDS). Washington, D.C.: Kirschner Associates, Inc.

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101.
- Budhwar, P. S., & Sparrow, P. R. (2002). An Integrative Framework for Understanding Cross-National Human Resource Management Practices. *Human Resource Management Review*, 12(3), 377–403.
- Burns, R., & Mason, D. (2002). Class Composition And Student Achievement In Elementary Schools. American Educational Research Journal, 39(1), 207–233.
- Burris, C. C., & Allison, K. E. (2013). Review of Does Sorting Students Improve Test Scores?. Boulder, CO. National Education Policy Center.
- Burris, C. C., Heubert, J. P., & Levin, H. M. (2006). Accelerating Mathematics Achievement Using Heterogeneous Grouping. American Educational Research Journal, 43(1), 137–154.
- Caldwell, D. W. (2012). Educating Gifted Students in the Regular Classroom : Efficacy , Attitudes , and Differentiation of Instruction. Georgia Southern University.
- Caldwell, J., Huitt, W., & Graeber, A. (1982). Time spent in learning: Implications from research. *The Elementary School Journal*, 82(5), 471–480.
- Carroll, J. (1963). A model of school learning. Teachers College Record, 64, 723–733.
- Carroll, J. (1989). The Carroll model a 25-year retrospective and prospective view. *Educational Researcher*, 18(1), 26–31.
- Chandola, V., & Kumar, V. (2007). Summarization–compressing data into an informative representation. *Knowledge and Information Systems*. *12*(3), 355–378.
- Chandra, E., & Nandhini, K. (2010). Knowledge Mining from Student Data. *European Journal of Scientific Research*, 47(1), 156–163.
- Chang, G. (2008). Strategic planning in education: some concepts and methods. UNESCO Education Sector. Paris, France.
- Chang, G., & Radi, M. (2001). Educational planning through computer simulation. UNESCO Education Sector. Paris, France.

- Chapman, D. W., & Mählck, L. O. (2004). Adapting technology for school improvement : a global perspective.
- Chen, L., & Guo, G. (2015). Nearest neighbor classification of categorical data by attributes weighting. *Expert Systems with Applications*, 42(6), 3142–3149.
- Chopra, N., & Lal, M. (2013). Descriptive Analysis of Enrollment Data and Adaptive Educational Hypermedia. *International Journal of Information Technology*, 5(1), 531–536.
- Colquitta, G., Pritcharda, T., & McColluma, S. (2011). The Personalized System of Instruction in Fitness Education. *Journal of Physical Education, Recreation & Dance*, 82(6).
- Cooper, C., Coll, C. T. G., Bartko, W. T., Davis, H. M., & Chatman, C. (2006). Developmental Pathways Through Middle Childhood: Rethinking Contexts and Diversity as Resources (p. 368). Psychology Press.
- Cooper, D. R., & Schindler, P. S. (2011). *Business Research Methods* (9th ed., p. 761). Pennsylvania State: McGraw-Hill Irwin.
- Cordeiro, R. L. F., Faloutsos, C., & Traina Júnior, C. (2013). Data Mining in Large Sets of Complex Data, 111–116.
- Côté, S., & Healy, T. (2001). The well-being of nations: The role of human and social capital. *Paris: Organisation for Economic Co-Operation and Development*.
- Crangle, C. (2002). Text summarization in data mining. *Soft-Ware 2002: Computing in an Imperfect World*, 332–347.
- Creed, P., & Patton, W. (2003). Predicting two components of career maturity in school based adolescents. *Journal of Career Development*, 29, 277–290.
- Creswell, J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* Sage Publications.
- Cruickshank, D. R. (1986). Profile of an Effective Teacher. *Educational Horizons*, 64(2), 80–86.

- Dasgupta, P., & Weale, M. (1992). On measuring the quality of life. World Development, 20(1), 119–131.
- Delahunty, S. (2000). The Effect of Information Technology on Decision Making. delahunty.com. Retrieved from http://delahunty.com/cv/paper\_EffectIT.doc
- Delavari, N., Shirazi, M. R. ayatollahzadeh, & Beikzadeh, M. R. (2004). A new model for using data mining technology in higher educational systems. In *Proceedings of* 5th International Conference on Information Technology Based Higher Education and Training (ITHET). Istanbul, Turkey.
- Deno, S., & Reschly, A. (2009). Developing a school wide progress monitoring system. *Psychology in the ..., 46*(1).
- Designs, E. (2005). Understanding Student Differences, 94(1), 57-72.
- Devi, B. R., Rao, K. N., Setty, S. P., & Rao, M. N. (2013). Disaster Prediction System Using IBM SPSS Data Mining Tool. *International Journal of Engineering Trends* and Technology (IJETT), 4(August), 3352–3357.
- Dimmock, C., & Wildya, H. (1995). Conceptualizing curriculum management in an effective secondary school: a Western Australian case study. *The Curriculum Journal*, 6(3), 297–323.
- Dominguez, A., Yacef, K., & Curran, J. (2010). Data Mining for Individualised Hints in e-Learning. In ... *conference on educational data mining* (pp. 91–100).
- Duncan, B. (2014). Education and Social Background. American Journal of Sociology, 72(4), 363–372.
- Duncan, B., & Duncan, O. (1968). Family stability and occupational success. Soc. Probs., 1.
- Duru-Bellat, M., & Mingat, A. (1989). How do French junior secondary schools operate? Academic achievement, grading and streaming of students. *European Sociological Review*, 5(1), 47–64.

Economic Planning Unit Malaysia. (2010). Ninth Malaysia Plan 2006 - 2010. Putrajaya.

- Ekstrom, R. (1961). Experimental studies of homogeneous grouping: A critical review. *The School Review*, 69(2), 216–226.
- Esposito, D. (1973). Homogeneous and heterogeneous ability grouping: Principal findings and implications for evaluating and designing more effective educational environments. *Review of Educational Research*, *43*(2), 163–179.
- Fang, C., & Bei, H. (2009). Research on Human Capital and Talent Gravitation Model of Industrial Cluster. 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology, 1529–1534.
- Fathian, M., & Amiri, B. (2007). A honeybee-mating approach for cluster analysis. *The International Journal of Advanced Manufacturing Technology*, 38(7-8), 809–821.
- Firouzi, B., Sadeghi, M., & Niknam, T. (2010). A new hybrid algorithm based on PSO, SA, and K-means for cluster analysis. *International Journal of Innovative Computing, Information and Control*, 6(7), 4198.
- Fisher, C., Marliave, R., & Filby, N. N. (1979). Improving Teaching by Increasing" Academic Learning Time.". *Educational Leadership*, 52–55.
- Flanders, N. (1965). *Teacher influence, pupil atitude, and achievement*. Washington, D.C.
- Forgasz, H. (2010). Streaming for mathematics in years 7–10 in Victoria: An issue of equity? *Mathematics Education Research Journal*, 22(1), 57–90.
- Fornell, C., & Bookstein, F. (1982). Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. *Journal of Marketing Research*, *19*(4), 440–452.
- Gage, N., & Berliner, D. (1992). *Educational psychology* (5th ed.). New Jersey: Houghton Mifflin Company.
- Gage, N. L., & Berliner, D. (1998). *Educational Psychology* (6th ed., p. 704). New Jersey: Wadsworth Publishing.

- Gamoran, A. (2002). Standards, Inequality & Ability Grouping in Schools. CES Briefings.
- Gamoran, A., & Berends, M. (1987). The effects of stratification in secondary schools: Synthesis of survey and ethnographic research. *Review of Educational Research*, 57(4), 415–435.
- Gentry, M. L. (1999). Promoting Student Achievement and Exemplary Classroom Practices Through Cluster Grouping: A Research-Based Alternative to Heterogeneous Elementary Classrooms. United States.
- Gilbert, M. J. (2012). *The Relationship Between Pupil Control Ideology and Academic Optimism*. Seton Hall University.
- Gilman, S., Kawachi, I., Fitzmaurice, G., & Buka, S. (2003). Socio-Economic Status, Family Disruption and Residential Stability in Childhood: Relation To Onset, Recurrence and Remission of Major Depression. *Psychological Medicine*, 33, 1341–1355.
- Glascoe, F. (2000). Evidence- based approach to developmental and behavioural surveillance using parents' concerns. *Child: Care, Health and Development*, 26(2), 137–149.
- Graff, J. C. (2009). Mixed Methods Research. Jones & Bartlett Learning, LLC.
- Gregor, S. (2002). A theory of theories in information systems. *Information Systems Foundations: Building the Theoritical Base*, 1–20.
- Gulyaev, S. a., & Stonyer, H. R. (2002). Making a Map of Science: General Systems Theory as a Conceptual Framework for Tertiary Science Education. *International Journal of Science Education*, 24(7), 753–769.
- Guruler, H., Istanbullu, A., & Karahasan, M. (2010). A new student performance analysing system using knowledge discovery in higher educational databases. *Computers & Education*, 55(1), 247–254.
- Hager, P., Holland, S., & Beckett, D. (2002). Enhancing The Learning and Employability of Graduates: The Role of Generic Skills (No. 9) (p. 16). Melbourne.

- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. The Journal of Marketing Theory and Practice, 19(2), 139–152.
- Hair, J. F., Sarstedt, M., Ringle, C. M., & Mena, J. a. (2011). An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal* of the Academy of Marketing Science, 40(3), 414–433.
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, 107–145.
- Hallam, S., Rogers, L., & Ireson, J. (2008). Ability grouping in the secondary school: attitudes of teachers of practically based subjects. *International Journal of Research & Method in Education*, 1–27.
- Hämäläinen, W., Kumpulainen, V., & Mozgovoy, M. (2013). Evaluation of clustering methods for adaptive learning systems. In *Artificial Intelligence Applications in Distance Education* (pp. 1–32).
- Hamid, N., & Zaman, H. (2009). Human Capital Indicators: Towards Achieving a Knowledge Society In Malaysia. *Napsipag.org*.
- Han, J., & Kamber, M. (2001). Data Mining: Concepts and Techniques. New York: Morgan Kaufmann Publishers.
- Hansen, A., Cottle, S., Negrine, R., & Newbold, C. (1998). Mass Communication Research Methods. Palgrave Macmillan.
- Hanushek, E., & Wößmann, L. (2007). *Education Quality and Economic Growth*. Washington, DC: World Bank.
- Harlen, W., & Malcolm, H. (1997). *Book Reviews Setting And Streaming : A Research Review*. (Scottish Council for Research Education, Ed.) (pp. 73–82). Edinburgh: SCRE Publication.
- Harris, L., & Brown, G. (2010). Mixing interview and questionnaire methods: Practical problems in aligning data. *Practical Assessment Research & Evaluation*, 15(1).

- Haslinda, A. (2009). Major Challenges to The Effective Management Of Human Resource. *International Social Research*, 2(8), 11–25.
- Heer, J., Shneiderman, B., & Park, C. (2012). A taxonomy of tools that support the fluent and flexible use of visualizations. *Interactive Dynamics for Visual Analysis*, 10, 1–26.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2014). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*.
- Huitt, W. (1995). A systems model of the teaching/learning process. *Educational Psychology Interactive*.
- IBM Corporation. (2011). IBM SPSS Modeler Professional. United States.
- IBM Corporation. (2012). IBM SPSS Modeler Text Analytics 15 User's Guide. United States.
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666.
- Jennifer Stepanek. (1999). Meeting the Needs of Gifted Students: Differentiating Mathematics and Science Instruction.
- Kanungo, T., Member, S., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An Efficient k -Means Clustering Algorithm : Analysis and Implementation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 24(7), 881–892.
- Kao, Y.-T., Zahara, E., & Kao, I.-W. (2008). A hybridized approach to data clustering. *Expert Systems with Applications*, 34(3), 1754–1762.
- Katz, L. F., Murphy, K. H., Bound, J., Freeman, R., & Coldin, C. (1991). Changes in Relative Wages, 1963-1897: Supply and Demand Factors (No. 3927) (pp. 1963– 1987). Cambridge, MA.

- Kaufmann, L., & Rousseeuw, P. . (1990). Finding Groups in Data. New York: John Wiley & Sons, Inc.
- Kehrer, J., & Hauser, H. (2013). Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(3), 495–513
- Kementerian Pendidikan malaysia. (2001). *Struktur Pendidikan Malaysia*. Kuala Lumpur.
- Kementerian Pendidikan malaysia. (2012). Pelan Pembangunan Pendidikan Malaysia 2013 2025.
- Kementerian Pendidikan Malaysia. (2013). Buletin Bulanan Transformasi Kementerian Pendidikan Malaysia. *Unit Prestasi Pelaksanaan Pendidikan (PADU)*, 1–4.
- Kemple, J. J., & Snipes, J. C. (2000). Career Academies: Impacts on Students' Engagement and Performance in High School. James Kemple, Jason Snipes. New York.
- Kerckhoff, A. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review*, 51(6), 842–858.
- Khan, S. B. D. (2005). Student Information Management System: A Case Study of Sekolah Menengah Kebangsaaan Seri Mahkota. Universiti Utara Malaysia.
- Kilgour, P. (2008). Problems arising from streaming mathematics students in Australian Christian secondary schools : To stream or not to stream? (pp. 33–37).
- Klausmeier, H., & Goodwin, W. (1971). *Learning and Human Abilities*. New York: Harper & Row Publishers.
- Klein, H. K., & Myers, M. D. (1999). A Set of Principles For Conducting And Evaluating Interpretive Field Studies in Information Systems. *MIS Quarterly*, 23(1), 67–93.
- Kocherlakota, S., & Healey, C. (2005). Summarization techniques for visualization of large multidimensional datasets. Raleigh, NC.

- Krejcie, R. V, & Morgan, D. W. (1970). Determining Sample Size for Research Activities. *Educational And Psychological Measurement*, 38, 607–610.
- Krishna, K., & Murty, M. (1999). Genetic K-means algorithm. Systems, Man, and Cybernetics, 29(3), 433–439.
- Kulik, C., & Kulik, J. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, 19(3), 415–428.
- Kulik, C. L., & Kulik, J. A. (1982). Highlights from "research on ability grouping." *Educational Leadership*, 39(8), 620.
- Kyriacon, G. (1980). Level and Growth Effects of Human Capital: A Cross Country Study of the Convergence Hypothesis. New York.
- Lam, C., Ho, E., & Wong, N. (2012). Parents' beliefs and practices in education in Confucian Heritage Cultures: The Hong Kong case. *Journal of Southeast Asian Education*, 3(1), 99–114.
- Lawrence, L., Jamison, D., & Louat, L. (1991). *Impact of Education by Region*. World Bank, Washington, D.C. Processed.
- Lank, E. (1997). Leveraging Invisible Assets: The Human Factor. Long Range Planning, 30(3), 406–412.
- Laszlo, M., & Mukherjee, S. (2007). A genetic algorithm that exchanges neighboring centers for k-means clustering. *Pattern Recognition Letters*, 28(16), 2359–2366.
- Lembaga Peperiksaan. (2012). Panduan Pengurusan Pentaksiran Berasaskan Sekolah (PBS).
- Lembaga Peperiksaan Malaysia. (2011). Buku Pengurusan dan Pengendalian Pentaksiran Berasaskan Sekolah (PBS) Di Sekolah Rendah (Vol. 14, pp. 1–50).
- Leuven, E. (2004). Explaining international differences in male skill wage differentials by differences in demand and supply of skill\*. *The Economic Journal*, 114(1997), 466–486.

- Lewis, M. (1998). Iterative triangulation: a theory development process using existing case studies. *Journal of Operations Management*, *16*(4), 455–469.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry* (p. 416). Beverly Hills: Sage Publications.
- Liu, W. C., Wang, C. K. J., & Parkins, E. J. (2005). A longitudinal study of students' academic self-concept in a streamed setting: the Singapore context. *The British Journal of Educational Psychology*, 75(Pt 4), 567–86.
- Lopez, M., & Luna, J. (2012). Classification via Clustering for Predicting Final Marks Based on Student Participation in Forums. In *Proceedings of the 5th international conference on educational data minin* (pp. 148–151).
- Losada, M. (2008). Work teams and the Losada line: New results. Retrieved from from http://pos psych.com/news/guest author/20081209129
- Maki, P. (2002). Developing an assessment plan to learn about student learning. *The Journal of Academic Librarianship*, 28(1), 8–13.
- Mandeville, G. (1988). School effectiveness indices revisited: Cross-year stability. *Journal of Educational Measurement*, 25(4), 349–356.
- Maulik, U., & Bandyopadhyay, S. (2000). Genetic algorithm-based clustering technique. *Pattern Recognition*, 33(9), 1455–1465.
- McIlrath, D., & Huitt, W. (1995). *The teaching-learning process: A discussion of models*. Valdosta, GA.
- Mingers, J. (2003). The paucity of multi-method research: a review of the information systems literature. *Information Systems Journal*, *13*, 233–249.
- Ministry of Education Malaysia. (2012). Pelan Pembangunan Pendidikan Malaysia 2013-2025.
- Ministry of Education Malaysia. (2013). The Public Schooling System for Primary, Secondary and Post-secondary Levels. In Schools of Malaysia Directory (4th ed.). Kuala Lumpur: Challenger Concept (M) Sdn Bhd.

- Mittendorff, K., Jochems, W., Meijers, F., & den Brok, P. (2008). Differences and similarities in the use of the portfolio and personal development plan for career guidance in various vocational schools in The Netherlands. *Journal of Vocational Education & Training*, 60(1), 75–91.
- Mitzel, H. (1960). Teacher efectivenes. In *Encyclopedia of Educational Research* (3rd ed.). New York: Macmilan.
- Mohamad, M. I., Rashid, S. A., Suryana, A. S., Abdullah, S., Abdullah, S., Omar, R., & Shoib, S. (2009). *Sistem analisis peperiksaan sekolah (SAPS)* (pp. 1–28).
- Monecke, A., & Leisch, F. (2012). semPLS: structural equation modeling using partial least squares. *Journal of Statistical Software*, (1989).
- Multimedia Development Corporation. (2005). Malaysian Smart School Roadmap 2005-2020: An Educational Odyssey.
- Nam, J., & Tewfik, A. (1999). Dynamic video summarization and visualization. Proceedings of the Seventh ACM International Multimedia, 53–56.
- National Economic Advisory Council. (2010). New Economic Model (NEM) For Malaysia. Putrajaya.
- Natoli, R. (2008). Indicators of economic and social progress: an assessment and an alternative. Victoria University, Melbourne, Australia.
- Ndiyo, N. (2002). The Paradox of Education and Economic Growth in Nigeria: An Empirical Evidence. In *NES Proceedings*.
- Nghe, N. T., Janecek, P., & Haddawy, P. (2007). A Comparative Analysis of Techniques for Predicting Academic Performance. 2007 37Th Annual Frontiers in Education Conference, T2G–7–T2G–12.
- Niknam, T., & Amiri, B. (2010). An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis. *Applied Soft Computing*, 10(1), 183–197.
- Oakes, J. (2005). *Keeping Track: How Schools Structure Inequality* (2nd ed.). New Heaven, CT: Yale University Press.

- Odden, A., & Kelly, J. A. (2008). Strategic Management Of Human Capital in Public Education. Madison.
- Ogor, E. (2007). Student academic performance monitoring and evaluation using data mining techniques. In *Electronics, Robotics and Automotive Mechanics Conference* (pp. 0–5).
- Okagaki, L., & Sternberg, R. (1993). Parental beliefs and children's school performance. *Child Development*, 64(1), 36–56.
- Olaniyan, D., & Okemakinde, T. (2008). Human capital theory: implications for educational development. *Pakistan Journal of Social Sciences*, 24(2), 157–162.
- Onwuegbuzie, A., & Collins, K. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, *12*(2), 281–316.
- Opler, M. (1945). Themes as dynamic forces in culture. *American Journal of Sociology*, 51(3), 198–206.
- Osmanbegović, E., & Suljić, M. (2012). Data Mining Approach for Predicting Student Performance. *Economic Review (Journal of Economics and Business)*, *X*(1), 3–12.
- Oyelade, O., Oladipupo, O., & Obagbuwa, I. (2010). Application of k means clustering algorithm for prediction of students academic performance. (*IJCSIS*) International Journal of Computer Science and Information Security, 7(1), 292–295.
- Pare, V. (2004). Exploring the Conflicts Involved With Ability Grouping. Neag Center for Gifted Education and Talent Development. University of Connecticut, United States.
- Patton, W., & McMahon, M. (1998). *Career development and systems theory: A new relationship* (1st ed.). Cengage Learning.
- Petrides, L. A., & Ufsd, H. (2002). Knowledge Management for School Leaders : An Ecological Framework for Thinking Schools. *Teachers College Record*, 104(8), 1702–1717.

- Pil, F., & Leana, C. (2009). Applying organizational research to public school reform: The effects of teacher human and social capital on student performance. *Academy* of Management Journal, 52(6), 1101–1124.
- Popham, W.J. (1999). *Modern educational measurement. Practical guidelines for the education leader*. Michigan: Pearson.
- Preckel, F., Götz, T., & Frenzel, A. (2010). Ability grouping of gifted students: Effects on academic self-concept and boredom. *British Journal of Educational Psychology*, 80(3), 451–472.
- Proctor, C. P. (1984). Teacher Expectations : A Model for School Improvement. *The Elementary School Journal*, 84(4), 468–481.
- Reay, D., David, M. E., & Ball, S. (2005). *Degrees of Choice: social class, race and gender in higher education*. Trentham Books.
- Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458.
- Rodríguez, P., & Nussbaum, M. (2010). A Monitoring and Evaluation Scheme for an ICT-Supported Education Program in Schools. *Educational Technology & ..., 13*, 166–179.
- Rogers, K. B. (2002). Grouping the gifted and talented: Questions and answers. *Roeper Review*, 24(3), 103–107.
- Rohani, A. H., & Khiruddin, A. (2003). Using Performance Data To Improve Learning:
  A Case Study Of The E-Monitoring Programme Of The Terengganu State
  Education Department. In *ELTC ETeMS Conference 2003: Managing Curricular Change* (pp. 1–16).
- Rokach, L., & Maimon, O. (2008). Data Mining With Decision Trees: Theory and Applications. (H. Bunke & P. S. P. Wang, Eds.). Tel Aviv. Israel.: World Scientific Publishing Co. Pte. Ltd.

- Romero, C., López, M.-I., Luna, J.-M., & Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, 68, 458–472.
- Romero, C., & Ventura, S. (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), 12–27.
- Rosenshine, B., & Stevens, R. (1986). Teaching functions. In *Handbook of research on teaching* (3rd ed., pp. 376–391). New York: Macmilan.
- Rosenthal, R., & Jacobson, L. F. (1968). Teacher Expectation for the Disadvantaged. *Scientific American*, 218(4).
- Sandelowski, M. (1995). Focus on qualitative methods: Sample sizes in qualitative research. *Research in Nursing & Health*, 18(2), 179–183.
- Sang, M. S. (2002). Psikologi Pendidikan. Kumpulan Budiman Sdn. Bhd.
- Savickas, M. L. (1999). The Transition from School to Work: A Developmental Perspective. *Career Development Quarterly*, 47(4), 326–336.
- Schneider, B., Judy, J., Ebmeye, C. M., & Broda, M. (2014). Trust in Elementary and Secondary Urban Schools: A Pathway for Student Success and College Ambition.
  In D. Van Maele, P. B. Forsyth, & M. Van Houtte (Eds.), *Trust and School Life* (pp. 37–47). Dordrecht: Springer Netherlands.
- Schütt, F. (2003). The Importance of Human Capital for Economic Growth. *Institute for World Economics and International Management*.
- Schwandt, T. (1994). Constructivist, interpretivist approaches to human inquiry. In N.K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 118–137). Thousand Oaks, CA: Sage.
- Seifert, J. (2004). Data mining: An overview. National Security Issues.
- Sembiring, S., Zarlis, M., Hartama, D., & Wani, E. (2011). Prediction of student academic performance by an application of data mining techniques 1 1, 6, 110– 114.

- Shields, C. M. (2002). A comparison study of student attitudes and perceptions in homogeneous and heterogeneous classrooms. *Roeper Review*, 24(3), 115–119.
- Shirwaikar, P., & Rajadhyax, N. (2012). Data Mining on Educational Domain. *arXiv Preprint arXiv:1207.1535*, 1–6. Retrieved from http://arxiv.org/abs/1207.1535
- Shovon, M., Islam, H., & Haque, M. (2012). An Approach of Improving Students Academic Performance by using k means clustering algorithm and Decision tree. (IJACSA) International Journal of Advanced Computer Science and Applications, 3(8), 146–149.
- Siegle, D. (1992). What Educators Need to Know About Ability Grouping. United States.
- Slavin, R. E. (1992). The nongraded elementary school: Great potential, but keep it simple. *Educational Leadership*, 50(2), 24.
- Smith, C. M. M., & Sutherland, M. J. (2003). Setting or mixed ability? Teachers' views of the organisation of pupils for learning. *Journal of Research in Special Educational Needs*, 3(3), 141–146.
- Smith, W. (2011). Strategic Planning in an Academic Context. In *Conference of the Canadian Engineering Education Association* (p. 6).
- Squires, D., Huitt, W., & Segars, J. (1983). Effective classrooms and schools: A research-based perspective. Washington, D.C.: Washington, D.C.: Association for Supervision and Curriculum Development.
- Sukhnandan, L., & Lee, B. (1998). Streaming, Setting and Grouping by Ability, a review of literature. Berkshire: National Foundation for Educational Research.
- Sumathi, S., & Sivanandam, S. (2006). Data Mining Tasks, Techniques, and Applications. *Introduction to Data Mining and Its Applications*, *4*.
- Tai, R. H., Liu, C. Q., Maltese, A. V, & Fan, X. (2006, May). Planning Early for Careers in Science. AAAS, 312(ii), 2.

- Tair, M. M. A., & El-halees, A. M. (2012). Mining Educational Data to Improve Students' Performance: A Case Study. *International Journal of Information and Communication Technology Research*, 2(2), 140–146.
- Takada, T., & Koike, H. (2002). Tudumi: information visualization system for monitoring and auditing computer logs. *Proceedings Sixth International Conference on Information Visualisation*, 570–576.
- Tashakkori, A., & Teddlie, C. (2003). *Handbook of Mixed Methods in Social & Behavioral Research*. Thousand Oaks, CA, Sage Publications.
- Taylor, A. (2007). Pathways for Youth to the Labour Market : An Overview of High School Initiatives. Ottawa, Ontario, Canada.
- Tomida, K., & Yamaguchi, Q. (2015). Using Data Mining to Identify Patterns of Student Voltage Drop. KUROSHIO, 8(2), 138-146.
- Tomlinson, C. (2001). Differentiated Instruction in the Regular Classroom: What Does It Mean? How Does It Look? *Understanding Our Gifted*, *14*(1), 3–6.
- Trivedi, S., Pardos, Z., Sárközy, G., & Heffernan, N. (2011). Spectral Clustering in Educational Data Mining. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th International Conference on Educational Data Mining* (pp. 129–138).

UNESCO. (2003). Fundamentals of educational planning.

- Utusan Melayu. (2009, March 27). Dasar 40:60 Pelajar Sastera-Sains Belum Tercapai. *Utusan Malaysia*. Kuala Lumpur. Retrieved from: http://www.utusan.com.my/utusan/info.asp?y=2009&dt=0327&pub=Utusan\_Mala ysia&sec=Dalam\_Negeri&pg=dn\_07.htm
- Verma, J. P. (2013). Data Analysis in Management with SPSS Software. India: Springer India.
- Wahbeh, A. H., Al-radaideh, Q. a, Al-kabi, M. N., & Al-shawakfa, E. M. (2010). A Comparison Study between Data Mining Tools over some Classification Methods. *Ijacsa*, 2 (8)(Special Issue on Artificial Intelligence), 18–26.

- Waterman, A. S. (2014). Service-learning: Applications From the Research (p. 208). Routledge.
- Weinstein, R. S., Madison, S. M., & Kuklinski, M. R. (1995). Raising Expectations in Schooling: Obstacles and Opportunities for Change. American Educational Research Journal, 32, 121–159.
- Whetzel, D. (1992). *The Secretary of Labor's Commission on Achieving Necessary Skills*. Retrieved from http://www.edpsycinteractive.org//files/scansrpt.html
- Willms, J. (2000). Monitoring school performance for'standards-based reform'. *Evaluation & Research in Education*.
- Wold, H. (1985). Partial least squares. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences* (pp. 587–599). New York: Wiley.
- Wong, H. (2003). There is only one way to improve student achievement. *Retrieved November*.
- Wu, X., Zhu, X., Wu, G.-Q., & Ding, W. (2014). Data mining with big data. IEEE Transactions on Knowledge and Data Engineering, 26(1), 97–107.
- Wynne, E. A. (2011). Evaluation as a Component in Incentive Systems. *Educational Evaluation and Policy Analysis*, 6(1), 85–93.
- Yadav, S., Bharadwaj, B., & Pal, S. (2012). Mining Education Data to Predict Student's Retention: A comparative Study. *International Journal of Computer Science and Information Security*, 10(2), 113–117.
- Yin, M., Hu, Y., Yang, F., Li, X., & Gu, W. (2011). A novel hybrid K-harmonic means and gravitational search algorithm approach for clustering. *Expert Systems with Applications*, 38(8), 9319–9324.
- Yin, R. . (1994). *Case study research: Design and method* (2nd ed.). Beverly Hills, CA: Sage Publishing.

Yusof, M. (2000). Sistem Pengurusan Peperiksaan (SPP). Terengganu, Malaysia.

- Zeng, L., Xu, L., Shi, Z., Wang, M., & Wu, W. (2007). Techniques, process, and enterprise solutions of business intelligence. In 2006 IEEE Conference on Systems, Man, and Cybernetics October 8-11, 2006 (p. 4722). Taipei, Taiwan.
- Zengin, K., Esgi, N., Erginer, E., & Aksoy, M. E. (2011). A sample study on applying data mining research techniques in educational science: Developing a more meaning of data. *Procedia - Social and Behavioral Sciences*, 15, 4028–4032.
- Zevenbergen, R. (2001). Is Streaming an Equitable Practice ? Students ' Experiences of Streaming in the Middle Years of Schooling. In 24th Annual MERGA Conference (pp. 563–570). Sydney.
- Zevenbergen, R. (2003). Grouping by ability: A self-fulfilling prophecy? Australian Mathematics Teacher, 59(4), 2–7.
- Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: an efficient data clustering method for very large databases. In *Proceedings of the ACM SIGMOD Conference* on Management of Data (Vol. 1, pp. 103–114). Montreal, Canada.
- Zimek, A., Schubert, E., & Kriegel, H.-P. (2012). Outlier detection for high dimensional data. In *IEEE 12th International Conference on Data Mining (ICDM)*. Brussels, Belgium.

Zimmermann, A., & Raedt, L. (2009). Cluster-grouping: from subgroup discovery to clustering. *Machine Learning*, 77(1), 125–159.