MODELING THE RELATIONSHIP BETWEEN STATISTICAL ACHIEVEMENT AND COGNITIVE DETERMINANTS AMONG MALAYSIAN DIPLOMA STUDENTS

FOO KIEN KHENG

INSTITUTE OF GRADUATE STUDIES
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

MODELING THE RELATIONSHIP BETWEEN STATISTICAL ACHIEVEMENT AND COGNITIVE DETERMINANTS AMONG MALAYSIAN DIPLOMA STUDENTS

FOO KIEN KHENG

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

INSTITUTE OF GRADUATE STUDIES
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: FOO KIEN KHENG

Registration/Matric No: HHB070004

Name of Degree: PHD

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

MODELING THE RELATIONSHIP BETWEEN STATISTICAL ACHIEVEMENT AND COGNITIVE DETERMINANTS AMONG MALAYSIAN DIPLOMA STUDENTS

Field of Study: STATISTICS EDUCATION

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 Date: 29 May 2017

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

The main purpose of the study is to determine the relationships of selected cognitive determinants on statistical achievement and statistical reasoning. In addition it seeks to determine the direct and indirect effect of gender and language on these relationships. This study uses a survey approach to collect data on the exogenous and endogenous variables using data from a cross-section of the sample of Diploma students. A survey form was used to collect secondary and primary data. To increase the content and construct validity of the instrument, two pilot studies were carried out. The pilot studies included the use of focus groups. Item analysis was used to weed out poor items. Reliability of the instrument was measured using Cronbach alpha. The SRA has moderately good reliability index. Purposive sampling was used to select 381 students from 6 statistics classes sourced from two branch campuses of a large university in Malaysia. The survey was administered a week later and handed back to the researcher immediately. Data cleaning and screening were carried out and only 374 usable forms were keyed in using the SPSS package. Multiple linear regression (MLR) analytic procedure was used to study the complex multivariate relationships based on the different hypothesized models as suggested in this present study. The findings showed that, students achieved moderately well on prior mathematical knowledge (PMK) and statistical achievement (SA). Unfortunately, they did not do well in statistical reasoning (SR) and had a substantially high level of misconception (MC) about statistics. PMK (M = 78.54, SD = 11.72) and SA (M = 64.63, SD = 24.78) as compared to SR (M= 38.17, SD = 13.83) and MC (M = 34.44, SD = 11.56). The best regression model on statistical achievement was:

$\mathrm{SA} = 8.75 + .58\,(\mathrm{PMK}) + .27(\mathrm{SR})$ where only prior mathematical knowledge (PMK) and statistical reasoning (SR) being significant contributors. The best model on

statistical reasoning was: $\mathrm{SR} = 43.61 + 0.05(\mathrm{SA}) - 0.58(\mathrm{MC}) + 3.45(\mathrm{ENG})$ where SA, MC and ENG were significant contributors to SR. Finally the findings found that gender and language mastery did not moderate the hypothesized relationships among the various cognitive determinants on achievement or reasoning. The significance of the findings includes identifying the determinants that are directly or indirectly influencing achievement and reasoning. These are important input for educators to find ways to improve the teaching and learning process in class. The current study has also shown that statistical achievement and reasoning are complex constructs and that the determinants used are but a small subset of the population of cognitive and non-cognitive factors.

ABSTRAK

Tujuan utama kajian ini adalah mengenalpasti perhubungan factor-faktor kognitif terpilih terhadap pencapaian (SA) dan penaakulan statistik (SR). Di samping itu ia bertujuan mengkaji kesan langsung dan tidak langsung faktor jantina (GEN) dan bahasa (ENG) terhadap perhubungan-perhubungan tersebut. Kajian ini menggunakan pendekatan kuantitatif menggunakan soal selidik untuk mengumpul data pembolehubah luaran dan dalaman dari pelajar-pelajar Diploma. Borang kaji selidik yang telah digunakan untuk mengumpul data sekunder dan primer. Untuk meningkatkan kesahan kandungan dan konstruk instrumen ini, dua kajian rintis telah dijalankan dan data dianalisis untuk memperbaiki borang kaji selidik dan item-item SRA. Kaedah kajian rintis termasuk kumpulan fokus. Analisis item telah digunakan untuk menapis item yang lemah. Kebolehpercayaan instrumen ini diukur dengan menggunakan Cronbach alpha. SRA ini mempunyai Indeks kebolehpercayaan yang sederhana. Selain daripada menggunakan hasil dua kajian rintis untuk menguji kesesuaian item-item SRA, kajian-kajian perintis ini juga membantu menentukan keberkesanan prosedur pengumpulan data. Persampelan 'purposive' telah digunakan untuk memilih 381 pelajar dari 6 kelas statistik yang diperolehi daripada dua kampus cawangan universiti besar di Malaysia. Borang kaji selidik yang teruji ini ditadbir seminggu kemudian dan diserahkan kembali kepada penyelidik dengan serta-merta. Data diteliti serta diperiksa untuk kesilapan input. Dari pemeriksaan awal tersebut, borang-borang yang boleh digunakan berjumlah 374. Maklumat ini terus dimasukkan menggunakan pakej SPSS. Prosedur analitik regresi linear pelbagai (MLR) telah digunakan untuk mengkaji hubungan multivariate kompleks berdasarkan model-model sebagaimana yang disarankan dalam kajian ini. Dapatan kajian menunjukkan bahawa responden kajian ini mempunyai penguasaan pengetahuan sedia ada matematik (PMK) dan pencapaian statistik (SA) yang baik

manakala penguasaan agak lemah dalam penaakulan statistik (SR) dan mempunyai konsepsi salah statistik (MC) yang agak tinggi. PMK (M = 78.54, SD = 11.72) dan SA (M = 64.63, SD = 24.78) berbanding SR (M= 38.17, SD = 13.83) dan MC (M = 34.44, SD = 11.56). Model regresi pertama adalah:

$\mathrm{SA} = 8.75 + .58\,(\mathrm{PMK}) + .27(\mathrm{SR})$ di mana PMK dan SR merupakan faktor yang bersignifikan sahaja. Model kedua pula adalah:

$\mathrm{SR} = 43.61 + 0.05(\mathrm{SA}) - 0.58(\mathrm{MC}) + 3.45(\mathrm{ENG})$ di mana SA, MC and ENG merupakan faktor-faktor signifikan kepada SR. Kajian mendapati bahawa jantina (GEN) dan penguasaan bahasa (ENG) tidak mempunyai kesan moderasi langsung terhadap sebarang perhubungan faktor kognitif yang diselidiki. Kepentingan penemuan ini termasuk mengenal pasti faktor penentu yang secara langsung atau tidak langsung mempengaruhi pencapaian dan penaakulan statistik. Penemuan ini adalah input penting bagi pendidik untuk mencari jalan memperbaiki pengajaran dan pembelajaran dalam kelas. Kajian ini turut menunjukkan bahawa pencapaian dan penaakulan statistik adalah konstruk yang kompleks dan factor-faktor yang digunakan adalah sebahagian kecil daripada populasi faktor-faktor kognitif dan bukan kognitif.

ACKNOWLEDGEMENTS

This research can only be completed because of the dedication and competence of many people throughout my journey of academic 'enlightenment'. The acknowledgments here represent my deepest gratitude and sincere appreciation to all those who have assisted me directly or indirectly.

First and foremost to my two supervisors, Professor Dato Dr. Noraini Idris of Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, previously with Faculty of Education, University of Malaya, Kuala Lumpur and Professor Dr. Ibrahim Mohamed, Faculty of Science, University of Malaya, Kuala Lumpur who contributed much of their expertise, patience and support throughout the whole dissertation process, giving me encouragement when I needed most.

To Dirk T. Tempelaar, Ph.D. Senior lecturer, Department of Quantitative Economics, Maastricht University, School of Business and Economics, Maastricht, The Netherlands for his insightful discussions and advice that make my research meaningful and on track.

Students and staff of UiTM, Kota Samarahan, Sarawak, Malaysia and UiTM, Kuala Pilah, Negeri Sembilan, Malaysia who unselfishly gave their time and cooperation to facilitate my collection of data and gave me leaves of absence to do this research.

Not forgetting all my friends, relatives and acquaintances who advise and encourage me at all time.

Last but not least to my supportive wife and 5 beautiful children and grandson who believed in me through the ups and downs of this long journey. Your sacrifices and expectations are not in vain.

TABLE OF CONTENTS

ABSTRACT……. III

ABSTRAK……… V

ACKNOWLEDGEMENTS... VII

TABLE OF CONTENTS.. VIII

LIST OF FIGURES XIV

LIST OF TABLES XVI

LIST OF APPENDICES XXI

CHAPTER 1 : INTRODUCTION 1

1.1 Background of the study 1

1.1.1 Statistics Education Today 1

1.1.2 Assessment and Statistical Education 2

1.1.3 Mathematical and Statistical Achievement of Malaysian Students 4

1.2 Statement of the Problem 6

1.3 Conceptual Framework 8

1.3.1 Prior knowledge 9

1.3.2 Reasoning 10

1.3.3 Errors in Human Cognition 12

1.3.3.1 Approaches to the study of error 14

1.4 Model of Study 15

1.4.1 Relationship between Prior Mathematical Knowledge (PMK) and Statistical Achievement (SA) 17

1.4.2 Relationship between statistical misconception and statistics achievement 17

1.4.3 Relationship between Statistical Reasoning (SR) and Statistical Achievement (SA) 19

1.4.4 Relationship of Prior Mathematics Knowledge (PMK) and Misconception (MC) 20

1.4.5 Relationship between Prior Mathematics Knowledge and Statistical Reasoning 20

1.5 Moderating Variables 21
1.5.1 Gender Effect and Statistical Achievement 23
1.5.2 Language Effect and Statistical Acheivement 25
1.6 Purpose of the Study 27
1.7 Objectives of the study 27
1.8 Research Questions 28
1.9 Delimitations of the Study 28
1.10 Limitations of the Study 29
1.11 Definition of Terms 30
1.12 Summary 31
CHAPTER 2 : LITERATURE REVIEW 33
2.1 Introduction 33
2.2 Statistics Education in Malaysia 33
2.2.1 The teaching and learning of statistics 34
2.3 Assessment in Statistics 35
2.3.1 Purposes of assessment 35
2.3.2 Taxonomy for assessing statistics educational outcomes 36
2.3.3 Assessing Statistical Cognitive Outcomes 38
2.3.4 Designing Assessments for Statistics Classes 39
2.3.5 Different ways of assessing statistical knowledge 40
2.3.5.1 Quizzes, tests and examinations 40
2.3.5.2 Homework 42
2.3.6 Assessing Achievement in statistics class 42
2.4 Information Processing Theory (IPT) 43
2.4.1 Information Processing Model and the Computer 44
2.4.2 Stage Model of Information Processing 44
2.4.3 Basic Principles of Information processing approach 46
2.4.4 Types of Memory 47
2.4.4.1 Sensory Memory (STSS) 47
2.4.4.2 Short Term Memory (STM) 48
2.4.4.3 Difference between short-term memory and working memory 48

2.4.4.4 Long-term memory (LTM)....49
2.4.4.5 Process of storing information in LTM....50
2.4.5 Recall of Information....50
2.4.6 Mental Representations....51
2.4.7 Schema Theory....53
2.4.8 The Practical Aspect of Schema Theory- Putting Theory into Practice....56
2.4.9 Schema Theory in Education....56
2.4.10 Instructional Implications of Schema Theory....56
2.4.11 Impact of Schema Theory on Education....57
2.5 Student Achievement in Statistics Classes....58
2.5.1 Achievement of primary school students in content areas and cognitive domains from TIMSS studies....58
2.5.2 Correlation analysis between content areas and cognitive domains in three TIMSS studies....61
2.6 Statistical Reasoning....62
2.6.1 What is reasoning?....62
2.6.2 Psychological perspective on Reasoning....66
2.6.3 Educational perspective on reasoning....68
2.6.4 What is statistical reasoning?....69
2.6.5 Relationships between Statistical Reasoning, Literacy and Thinking 70
2.6.6 Statistical reasoning and its assessment....72
2.6.7 Development of the SRA by Garfield (2003)....74
2.6.8 Validity of the SRA instrument....75
2.6.9 Weaknesses of the SRA instrument....77
2.7 Misconceptions in Statistics....78
2.7.1 Studies about misconceptions in basic statistics and statistical inference....81
2.7.2 A Survey of Malaysian and Singaporean University students' misconceptions concerning statistical inference....81
2.8 Prior Knowledge and Information Processing Model (IPM)....86
2.8.1 Sensory memory....86
2.8.2 Short-term memory (STM)....87
2.8.3 Difference between short-term memory and working memory....87
2.8.4 Long-term memory (LTM)....87
2.8.5 Implications for Learning....88

2.8.6 Undergraduates' understanding of some common statistical terms 88

2.9 What are Moderators? 90

2.10 Summary 92

CHAPTER 3 : METHODOLOGY 93

3.1 Introduction 93

3.2 Research Design 93

3.3 Model Testing and Model Adequacy 95

3.3.1 *R*-squared and Adjusted *R*-squared 95

3.3.2 The *F*-test 95

3.3.3 Survey Design 96

3.4 Sampling 97

3.4.1 Rationale for Sampled Population 97

3.4.2 Descriptions of sample and sample size 98

3.5 Data Collection Instruments 99

3.6 Procedures for Implementation of Study 101

3.6.1 Preliminary study 101

3.6.2 Pilot testing 102

3.6.3 Item Analysis 103

3.6.4 Results of Principal Component Analysis for pilot testing of SRA (n = 206) 106

3.6.5 Validity and Reliability issues of SRA 110

3.6.5.1 Checking for Reliability of SRA using Cronbach Alpha ... 112

3.7 Actual study 114

3.8 Data Analysis Techniques 114

3.8.1 Statistical Software 116

3.8.2 Preliminary Analysis 116

3.8.3 Missing values 117

3.8.4 Methodological issues on the use of multiple regression analysis ... 118

3.8.5 The Choice of Software for Analysis 119

3.8.6 Screening for assumptions of multiple regression 119

3.9 Selecting the best regression model 120

3.9.1 Deciding on the best model 122

3.10 Procedure for testing moderation effect 125

3.10.1 General Guideline to assess a moderator effect in a causal relationship 125

3.11 Summary 126

CHAPTER 4 : RESULTS 127

4.1 Introduction 127

4.2 Descriptive Analysis 127

4.2.1 Description of Sample and Population 127

4.2.2 Descriptive results of cognitive variables 128

4.2.3 Correlational analysis of variables of interest 129

4.2.3.1 Pearson's correlation coefficient 129

4.3 Relationships of Students' statistical achievement with selected variables like reasoning, prior knowledge, misconception, language mastery and gender 131

4.3.1 Diagnostics on the Hypothesized Model 132

4.3.1.1 Checking for order of entry into the model using Partial Correlation Matrix Results 132

4.3.2 Assumption checks for the Regression Model 140

4.3.2.1 Assumption Checks on Normality of dataset 140

4.3.2.2 Assumption Checks on Multicollinearity of dataset 141

4.3.2.3 Checking for Outliers in the sample 142

4.3.3 Best Model for the regression analysis 144

4.4 Moderating effect of language mastery and gender on the relationships between statistical achievement and the predictors 146

4.4.1 The moderating effect of the variables language mastery (ENG) and gender (GEN) on the relationships of the other response variables like SA, SR, PMK and MC 146

4.4.1.1 Does English language mastery moderate the influence of statistical reasoning on statistical achievement? 146

4.4.1.2 Does English language mastery moderate the influence of prior mathematical knowledge on statistical achievement? 148

4.4.1.3 Does gender moderate the influence of statistical reasoning on statistical achievement? 150

4.4.1.4 Does gender moderate the influence of prior mathematical knowledge on statistical achievement? 151

4.5 Relationships of Students' statistical reasoning with selected variables like prior knowledge, misconception, language mastery and gender 153

4.5.1 Assumption checks for Regression Model 160

4.5.2 Best model for regression of cognitive determinants on Statistical Reasoning 161

4.6 Moderating effect of language mastery and gender on the relationships between statistical reasoning and the predictors 162

4.6.1.1 Does language mastery moderate the influence of misconception on 163

statistical reasoning? 163

4.6.1.2 Does gender moderate the influence of misconception on statistical reasoning? 165

4.7 Summary 167

CHAPTER 5 : DISCUSSION AND CONCLUSION 172

5.1 Introduction 172

5.2 Discussion 172

5.3 Research Design, Sample and sampling technique 179

5.4 Data collection instrument 180

5.4.1 Data analysis technique using Multiple Linear Regression approach 182

5.5 Implications 183

5.6 Future Research 188

5.7 Summary 191

REFERENCES… 193

LIST OF PUBLICATIONS AND PAPERS PRESENTED 236

LIST OF FIGURES

Figure 1.1: The Hypothesized Relationships among selected cognitive factors and statistical achievement using aggregated scores 16

Figure 1.2: A Moderator Effect Framework for a Correlational Design (Baron & Kenny, 1986) 22

Figure 2.1: Types of Memory (Plotnik & Kouyoumdjian, 2011) 47

Figure 2.2: The Information Processing model (Atkinson and Shiffrin, 1968) 50

Figure 2.3: The overlapping of the relationships between statistical literacy, reasoning and thinking (delMas, 2004a) 71

Figure 2.4: Percentages of respondents with misconceptions across 4 studies............... 83

Figure 2.5: Misconception scores across 4 studies - item by item analysis. 83

Figure 2.6: Types of Memory (Plotnik et.al, 2011) 86

Figure 2.7: A Moderator Effect Framework for a Correlational Design (Baron & Kenny, 1986) 91

Figure 3.1: Scree Plot showing the six dimensions/components 108

Figure 4.1: Residuals analysis on normality of dataset 140

Figure 4.2: Normal P-P plot on normality of dataset 140

Figure 4.3: Data points distribution in 3D plot to identify outliers 143

Figure 4.4: Scatterplot on zpred versus zresid to check for linearity, homoscedasticity and independence (Field, 2013) 143

Figure 4.5: Moderating effect of ENG on the relationship between SR and SA 146

Figure 4.6: Moderating effect of ENG on the relationship between PMK and SA 148

Figure 4.7: Moderating effect of ENG on the relationship between SR and SA 150

Figure 4.8: Moderating effect of GEN on the relationship between PMK and SA 151

Figure 4.9: Scatterplot on distribution of SA versus MC 160

Figure 4.10: Scatterplot on distribution of statistical reasoning normality check......... 160

Figure 4.11: Scatterplot on distribution of standardized residual showing, linearity, homoscedasticity and independence (Field, 2013) 161

Figure 4.12: Moderating effect of ENG on the relationship between MC and SR 163

Figure 4.13: Moderating effect of GEN on the relationship between MC and SR 165

Figure 4.14: The best model showing the relationships prior mathematical knowledge, statistical reasoning and statistical achievement 170

Figure 4.15: The best model showing the relationships between statistical achievement, misconception, language mastery and statistical reasoning ... 171

LIST OF TABLES

Table 2.1: Words used for Different Assessment Items or Tasks (delMas, 2002) 37

Table 2.2: Achievement Rubric for TIMSS studies (Mullis et al., 2008) 59

Table 2.3: Trend of the average mathematics scores of eighth grade students, by selected country 59

Table 2.4: Scores for Mathematics Content and Cognitive Domain of Eighth Grade Students, by Country in 2007 (Mullis et al., 2008; IEA, 2009) 63

Table 2.5: Grade 8 Math versus Cognitive Domains from TIMSS 2003 to 2011 (IEA, 1999, 2003, 2009; Mullis et al., 2000, 2008, 2012) 64

Table 2.6: Grade 4 Math versus Cognitive Domains from 2003 to 2011 (IEA, 1999, 2003, 2009; Mullis et al., 2000, 2008, 2012) 65

Table 2.7: Topics and distribution of items for reasoning scales in SRA 76

Table 2.8: Topics and distribution of items used in the SRA for different versions 77

Table 2.9: Average misconception scores for Malaysian and Singaporean Participants 82

Table 2.10: Malaysian and Singaporean Participants' Understanding of Statistical Concepts 90

Table 3.1: Difficulty index and Discrimination Index of SRA instrument 105

Table 3.2: Dimensions of SRA (Garfield, 2003) 107

Table 3.3: Dimensions from PCA analysis based on dataset (n=206) 107

Table 3.4: The extracted six components after rotation 109

Table 3.5: Misconceptions in Statistical Reasoning (Garfield, 2003) 110

Table 3.6: Case Processing Summary 113

Table 3.7: Reliability Statistics 113

Table 3.8: Item-Total Statistics 113

Table 4.1: Language Mastery Distribution of Sample 128

Table 4.2: *Aggregated scores for independent and dependent variables 129

Table 4.3: Analysis of Correlation Matrix 130

Table 4.4: Correlation Matrix 132

Table 4.5: Correlation matrix controlling for Prior Mathematical Knowledge 133

Table 4.6: Correlation matrix controlling for Prior Mathematical Knowledge 134

Table 4.7: Correlation matrix controlling for PMK, SR and GEN 135

Table 4.8: Correlation matrix controlling for PMK, SR, GEN and MC 135

Table 4.9: Order of entry into the regression model 136

Table 4.10: Checking for the best model 138

Table 4.11: Identifying the best regression model coefficients[a] 138

Table 4.12: Significance of the regression model 139

Table 4.13: Identifying the collinearity measures 141

Table 4.14: Residuals Statistics[a] 142

Table 4.15: Tolerance and VIF indices for checking multicollinearity 145

Table 4.16: Influence of ENG on SR and SA 147

Table 4.17: Regression Coefficients 147

Table 4.18: Influence of ENG on PMK and SA 148

Table 4.19: Regression Coefficients 149

Table 4.20: Influence of GEN on SR and SA 150

Table 4.21: Regression Coefficients 150

Table 4.22: Influence of GEN on PMK and SA 152

Table 4.23: Regression Coefficients 152

Table 4.25: Order of Entry of variables 154

Table 4.24: Correlation Matrix for the selected factors 155

Table 4.26: Summary statistics 156

Table 4.27: Coefficients of the regression model 158

Table 4.28: Residuals Checks 161

Table 4.29: Moderator Effect on language mastery on the said relationship 163

Table 4.30: Regression analysis to test for moderating effect of GEN on SR and MC.165

Table 4.31: Regression coefficients 166

Table 4.32: ANOVA table 166

LIST OF SYMBOLS AND ABBREVIATIONS

ANOVA	Analysis of Variance
ASA	America Statistical Association
CAOS	Comprehensive Assessment of Outcomes in a first Statistics course
ENG	Language Mastery
GAISE	Guidelines for Assessment and Instruction in Statistics Education
GEN	Gender
GPA	Grade Point Average
ICOTS	International Conference on Teaching Statistics
IEA	International Association for the Evaluation of Educational Achievement
IPM	Information Processing Model
IPT	Information Processing Theory
LTM	Long-term memory
MC	Misconception
MEB	National Education Blueprint
MLR	Multiple Linear Regression
NCTM	National Council of Teachers of Mathematics
NHST	Null Hypothesis Statistical Test
NUS	University of Singapore
OECD	Organisation of Economic Cooperation and Development
PCA	Principal Component Analysis
PISA	Program for International Student Assessment
PMK	Prior Mathematical Knowledge
QRQ	Quantitative Reasoning Questionnaire
SEM	Structural Equation Model
SA	Statistical Achievement

SR	Statistical Reasoning
SRA	Statistical Reasoning Assessment
STM	Short Term Memory
STSS	Sensory Memory
TIMSS	Trends in International Mathematics and Science Survey
UM	University of Malaya
VIF	Variance Inflation Factor

LIST OF APPENDICES

APPENDIX A1 - Statistical Reasoning Assessment (Garfield, 2003)........................ 213

APPENDIX A2 - Statistical Reasoning Assessment- Final Version.......................... 219

APPENDIX B - Scoring For Reasoning and Misconception Subscales…………..…225

APPENDIX C - How Aggregated Score is calculated for each of the factors………..226

APPENDIX D - Questionnaire on Hypothesis Testing and Prior Knowledge………..228

APPENDIX E - Variables entered/removed……………………………………….231

APPENDIX F - Excluded variables……………………………………………..232

APPENDIX G - Residual statistics…………………….…………………………233

APPENDIX H - ANOVA table….…………………………………………………234

APPENDIX I - Excluded variables/Residual Statistics………………………………235

CHAPTER 1 : INTRODUCTION

1.1 Background of the study

Malaysia has made major inroads into providing educational quality and accessibility to all. However there are still areas for improvements in particular mathematics and statistics. Recent reports from two international studies into the achievement of primary and secondary schoolchildren in the field of Science and Mathematics around the globe have indicated that much has to be done in Malaysia. Trends in International Mathematics and Science Survey (TIMSS) and the Program for International Student Assessment (PISA) are funded by the International Association for the Evaluation of Educational Achievement (IEA) and the Organisation of Economic Cooperation and Development (OECD) respectively (IEA, 2009,2013; Mullis et al., 2000, 2008; Mullis, Martin, Foy & Arrora, 2012; OECD, 2010, 2013). The Organisation of Economic Cooperation and Development (OECD) released the PISA 2011 (OECD, 2013) findings where Malaysia is placed at 52nd place out of 76 countries in term of 15 year old students' basic skills behind Vietnam and Thailand, its close neighbors. It also highlighted the fact that Malaysia is in the bottom third where its primary and secondary school Mathematics and Science tests are concerned. Findings from these studies are indicators of students' proficiency level in mathematics and statistics.

Changes are all around us and statistics education too follows this dynamic of uncertainties and variations with respect to environment, culture, technology and needs of the time. Thus it is no surprise that statistics educators are faced with ever-changing challenges and issues that were significantly different at the turn of the decade.

1.1.1 Statistics Education Today

Statistics is a good tool in assisting us to portray the representational and inferential properties of the data set. Statistics has high utility value in empirical studies be it in the Sciences, Economics, Business or Social Sciences. The appropriate usage and its optimization assure an output that can provide better and reliable information for solving problems and making good decisions. The ability to extract quality information from big data is a much needed skill in today's workplace. Recent studies (Chan, Zaleha & Bambang, 2014; Foo & Noraini, 2010; Noraidah, Hairulliza, Hazura & Tengku Siti Meriam, 2011; Garfield & Ben-Zvi, 2008) have found the learning of statistics difficult for many especially those with weak mathematical foundations. Many studies about how students developed statistical schemas and structures, acknowledged that learning statistics is a complicated process involving links and crossovers among many related cognitive components. These learning complexities ultimately make statistical understanding a challenging task (Garfield, 2003; Franklin & Garfield, 2006; Guidelines for Assessment and Instruction in Statistics Education (ASA, 2005a, 2005b). In addition the researchers concurred on the need for meaningful learning through new teaching and learning strategies. Acquisition of strong statistical foundation and seeing the 'big picture' hold the key to understanding statistics and its utility without which statistics remain 'a long list of terms to memorize and complex calculations to compute' (Foo & Noraini, 2010). The research findings had clearly indicated a need for revisions to a curriculum where higher-order statistical thinking skills are highly valued (ASA, 2005a, 2005b; Pfannkuch & Wild, 2003, 2004).

1.1.2 Assessment and Statistical Education

Assessment has been defined by Overton (2008) as the process of gathering information for the purpose of monitoring the learner's progress as well as to make educational decisions. It is conceptually different from the terms 'testing' or 'evaluation'. While testing is about the way one determines a learner's ability to complete a particular task or to be able to demonstrate mastery of a skill or knowledge of content, assessment on the other hand goes beyond that to include assessment techniques such as observations, interviews and behavioral monitoring. On the other hand, evaluation has both quantitative and qualitative aspect to assessing a learner. Overton (2008) sees it as a set of procedures used to determine whether the subject meets pre-set criteria i.e. such as qualifying for special education.

In this present study, the focus will be on assessment, a very crucial component of the learning process. An important goal at the end of the teaching and learning process is to know what and how much has been internalized by the learner. Thus assessment should be the source of this needed information. Some educators thought of assessment as: 1) assessment for learning, 2) assessment of learning that takes into account the active process of cognitive restructuring occurring when individuals interact with new ideas, 3) assessment of learning is about using tools or strategies to measure proficiency and assist in deciding students' future learning (Manitoba Education, Citizenship and Youth, 2006).

In many statistics classes nowadays, the traditional method of assessment is no more the primary path to getting information and feedback on learner achievement. Modern techniques are now employed to inform educators not only of the scores but the students' understanding and reasoning as well.

Gal and Garfield (1997) suggested that assessing only statistical knowledge or skill is too limiting. Assessment should provide information concerning whether students are able to understand statistical processes such as investigations, reasoning, thinking as well as being statistically literate. To achieve this, Garfield (1994) and Radke-Sharpe (1998) suggested some methods for assessing statistical knowledge and understanding among which are doing assessment tasks like quizzes, group projects, case studies, portfolios and examinations just to name a few.

The GAISE Reports published by the American Statistical Association (ASA 2005a, 2005b, 2005c) emphasizes on students to develop statistical literacy and thinking.

They further implored educators to adopt a 'frame-work' that can promote the crucial competencies for graduates to work in the modern world.

At the end of each statistics course, invariably one has to know whether the students are statistically literate, can reason well and most importantly be able to think and apply learnt skills to a data-rich environment in which one live.

1.1.3 Mathematical and Statistical Achievement of Malaysian Students

The achievement of students in statistics both in schools and higher learning centers is a cause for concern. Access to Malaysian school mathematics and statistics achievement results in particular is limited. The general picture of the situations in Malaysia can be seen at two international studies. These studies on science and mathematics achievement like TIMSS and PISA (Mullis et al., 2000, 2008, 2012; OECD, 2010, 2013) have traditionally been the main sources of data to inform the general public about how primary and secondary students in a participating country are ranked in comparison to other participant countries. Malaysia launched its Malaysia Education Blueprint (MEB) for 2013-2025 to improve access to quality education and putting the country among the top educational hubs of the region. To improve, it must rectify weaknesses in the education system. One of the identified areas for improvement was the achievement of Malaysian students in Mathematics and Science. The preliminary Blueprint report (Ministry of Education Malaysia (MOE), 2012) among others highlighted the downward trend of Malaysian secondary students from the TIMSS and PISA studies. It reported that Malaysian's achievement had slipped to below the international average where 18% of Malaysia's students failed to meet the minimum proficiency levels in Mathematics in 2007 as compared to only 7% in 2003. In addition the report said that the results from PISA 2009 (OECD, 2010) were also discouraging where Malaysia ranked in the bottom third out of 74 participating countries.

An in-depth analysis of data from the Trends in International Mathematics and Science Survey (TIMSS) reports from 1999-2011 (Mullis et al., 2000, 2008, 2012; Gonzales et al., 2004) confirmed that there is much to be done in the teaching and learning of mathematics and more so in statistics for Malaysia. The 2011 TIMSS report (Mullis et al., 2012) showed Malaysia's mathematical achievement dropped significantly as compared to 2007 while its closest neighbour Singapore recorded an increase of 18 points for the same period of time. Furthermore in the same study, it was reported that Malaysia recorded a significant drop in the 'Data and Chance' component. In 2007 Malaysian secondary school participants scored an average of 468 in four major content areas, with a standard estimate of 3.8 as compared to 429 with a standard estimate of 5.3 in the 2011 TIMSS report. The bigger standard of estimate for 2011 as compared to that of 2007 is not a good indicator of performance consistency. The performance of the 2011 cohort of Malaysian secondary school students in the section Data and Chance was lower than that of the other 3 components i.e. Number, Algebra and Geometry. All indicators taken together meant that the mathematics and statistics proficiency of the Malaysian Form 2 students are of real concern. Furthermore, there was a wide variation of abilities among the students in this cohort. This worrying trend in statistical achievement has been noted since 1999 and the present scenario seems to indicate that it is still sliding.

As for the Cognitive domain reported in these studies, a similar trend has been observed. Malaysian students' achievement in 'Reasoning', one of the three cognitive domains assessed, as expected was below the other domains like 'knowing' and 'applying'. This domain is understandably much more difficult than the other two as it involves higher-order thinking skills like analyzing, synthesizing and evaluating. The average reasoning score in all TIMSS studies attained by each of the countries mentioned above, was generally lower than those of the 'knowing' and 'applying' domains. The findings from the various TIMSS studies show clearly the route educators must pay more attention to i.e. reasoning competencies to prepare for functioning in future workplace. In this respect, statistical reasoning is a crucial higher-order thinking skill that needs to be aggressively imparted in diploma and undergraduate statistics courses without which rote memorization will probably prevail.

A more recent study by University Technology Malaysia (UTM) further provided more evidence of the weaknesses students in the tenth grade are facing in their statistics classes (Chan, Zaleha & Bambang, 2014). One of the major objectives of the UTM study was to gauge the statistical reasoning ability among the tenth-grade students in the secondary schools. Unsurprisingly the study found this random sample of 412 students from among Malaysian secondary schools, performed 'at a poor level'. There are abundant studies about statistics achievement and in particular statistics reasoning in the west but in Malaysia they are few and far in between.

Mathematics and statistics achievements in Malaysian colleges and universities are not expected to perform any better gauging from the poor achievement of Malaysian primary and secondary school students in the TIMSS and PISA studies (Mullis et al., 2000, 2008, 2012; OECD, 2010, 2013). The findings of Noraidah et al. (2011) suggested that undergraduates' statistical achievement in a Malaysian public university was only average. Statistical achievement of Malaysian Diploma students did not fare too well. This finding was corroborated by Zuraida, Foo, Rosemawati & Haslinda, (2012).

1.2 Statement of the Problem

According to the Executive Summary of the National Education Blueprint (MOE, 2012) the Malaysian government conceded that students lack "important cognitive skills, including problem-solving, reasoning, creative thinking, and innovation. This is an area where the system has historically fallen short, with students being less able than they should be in applying knowledge and thinking critically outside familiar academic contexts" (p. E-16). This statement was timely and Malaysia realizes the below-par achievement of her students in both content and cognitive domain in particular statistics.

There are very few studies aimed at measuring students' statistical competency and assessing their conceptual understanding and reasoning skills (Zamalia & Nor Hasmaniza, 2010; Watson, 1997). Many of the studies in the literature concerns undergraduates and secondary students and little about Diploma students (e.g. Garfield, 2002, 2003; Tempelaar, Van der Loeff & Gijselaers 2007; Chan, Zaleha and Bambang, 2014). The TIMSS reports on the 'Reasoning' domain as well as 'Data & Chance' domain of the Malaysian Year 4 and Form 2 students were other sources of reliable data reflecting their statistical competency as described earlier. One interesting similarity in the findings was the question of the apparent insignificant relations between achievement and reasoning where Tempelaar et al. (2006) were puzzled by the apparent low or non-existence of correlations between statistical achievement and reasoning skills.

Declining standards in statistics achievement cannot be blamed solely on reasoning skills alone. There are studies that point to other cognitive and non-cognitive determinants like student previous course of study, their grade point average, language skills, self-efficacy, student's attitude towards statistics or student perception of statistics as a tough subject (Lalonde & Gardner, 1993; Hardre et al, 2006; Chang & Cheo, 2012). Cognitive and non-cognitive determinants have varying influence on student achievement in introductory and advanced statistical courses. Lalonde and Gardner (1993) found among psychology students that achievement was related to aptitude, anxiety, attitudes and motivation to learn statistics while Hardre et al, (2006) found a mix of cognitive and non-cognitive factors influencing the achievement among her respondents. Some of which were academic ability, motivation, support, gender, age, race and motivation to learn.

A recent study found that students' pre-university grade is the most important determinant in undergraduates' achievement. The type of pre-university program taken prior to university admission, and ethnicity were found to be important determinants among University of Malaya students (Chang & Cheo, 2012).

Research has indicated that achievement in statistics was directly predicted by a variety of cognitive and non-cognitive factors (Tremblay, Gardner & Heipel, 2000; Nasser, 2004; Chiesi & Primi, 2010). Additionally a literature review highlighted an obviously complex relationship among the various cognitive and non-cognitive factors with statistical achievement. Based on these grounds, this research attempts to determine the effect of only selected cognitive factors on statistical achievement and reasoning in Diploma in Science students in a major Malaysian public university using multiple regression model. Among the factors to consider are cognitive determinants like prior mathematical knowledge, reasoning skills, and misconceptions on student achievement in statistics. In addition this study seeks to determine whether demographic factors like language mastery and gender have any interaction effect on the relationship mentioned earlier.

1.3 Conceptual Framework

Learning is partly a cognitive process and partly a socio-affective process. Through these processes one acquires concepts, ideas, knowledge structures, skills and competencies, attitudes and beliefs. Learning involves not only cognitive faculty but other faculties like feeling, experience and of course a context for all these to happen. An understanding of the processes involving learning can be illuminated through an understanding provided by cognitive psychology.

At the very heart of cognitive psychology is the idea of information processing. A cognitive psychologist sees a person as a processor of information, just like how a computer processes information following the direction given out by a program. The approach used by cognitive scientists to study the complex cognitive processes of the human brain is similar to the way a person seeks to understand the complex algorithms executed by a computer (Anderson, 1982, 1996).

McLeod, (2008) opined that information is being transformed by the senses upon entering the human brain through 'mental programs' with behavioural responses as the output.

Cognitive psychology has influenced and integrated with many other approaches and areas of study. Its perspective is reductionist in nature thus able to reduce complex mental processes into their smaller and simpler components to facilitate scientific inquiry (Anderson, 1982, 1996).

Cognitive development theories are developed to understand and explain complex thinking like reasoning, judgement, decision making and problem solving. According to Riegler and Riegler (2004), reasoning, judgement and decision making are complex thought processes that utilize all the component parts of cognition and are found to be closely related.

As these three processes are highly related, it is very difficult to study the complexities of their relationships. Thus this study takes a reductionist view by focusing specifically on the reasoning aspect, the errors the students frequently make while reasoning, prior knowledge and the influence of gender and language.

1.3.1 Prior knowledge

Cognitive theories see prior knowledge as residing in the long-term memory. Psychologists hypothesized this knowledge has been encoded in the form of mental representations or cognitive representations. These representations are theoretical constructs of cognitive scientists in their attempt to explain mental processes and their manifestations in the form of behaviors. Some studies have shown that prior knowledge is an important determinant of undergraduates' academic performance (Chang & Cheo, 2012). Equally important in measuring prior knowledge is to establish the mathematics content as required in any introductory statistics course. Chiesi and Primi (2010) identified pertinent mathematics content that they felt important to 'measure accurately the mathematics ability needed by psychology students enrolling in introductory statistics courses'. They defined these contents as those basic mathematical skills to solve statistics problems. The domains so identified were: Operations, Fractions, Set theory, first order Equations, Relations and Probability. In this study, the prior mathematical knowledge score calculated for each respondent is an aggregated score using the results of a few courses that tested the mastery of the student in these topics.

1.3.2 Reasoning

Reasoning, noted Galotti (2008) involves cognitive processes that turn bits and bytes of data into useful information so that the person can come to a conclusion. Reasoning covers either thinking that uses a well-defined system of logic and/or thinking on a small set of very well-defined tasks. Reasoning involves drawing conclusions based on some given information and in accordance with certain boundary conditions specified by the tasks. Mercier and Sperber (2011) see reasoning as a way of improving our store of knowledge and in turn it helps to make better decisions.

From a psychological perspective, reasoning is thought to be a mental process to derive inferences or conclusion from information known as premises. Reasoning helps to generate new knowledge and organize prior knowledge, so that it can be used in future work.

Reasoning is important as this is the key to successful decision making and problem solving. Reasoning helps to generate new knowledge and to organize existing knowledge, rendering it more usable for future mental work such as scientific, critical, and creative thinking, argumentation, problem solving, and decision making. Each of these more complex forms of thought can employ inductive, deductive, and abductive reasoning. Sometimes we use a procedure that employs shortcuts or heuristics to yield a solution. Heuristics are rules of thumb or mental short-cuts that reduce the number of steps we would normally use to solve a problem. It is fast and efficient but tends to be error ridden.

Baron (2004) suggests three psychological models to evaluate how people reason or make decision – normative, descriptive and prescriptive. The normative model tells us what people will do under ideal circumstances and unlimited time and knowledge. We create a benchmark to compare all other measures. The descriptive model tells us how we actually think. In a tossing of a fair coin experiment, after tossing four times this sequence was recorded 'HTHH" what is most probable to appear in the next toss- a tail or a head? Using the normative approach, both outcomes are likely but using a descriptive approach, a tail. Thus using the second approach incurs an error called the representative bias. The prescriptive model offers a realistic scenario, and is benchmarked against a set of realistic measures for which a person's decisions can be evaluated. It takes into consideration the constraints on their time, knowledge, energy and other priorities. knowledge is limited and this places pragmatic constraints on how well we reason (Johnson-Laird, 2006). Classical models of reasoning using logic or laws of probability usually assume people to be an ideal reasoner with a good supply of cognitive resources. Unfortunately this is not the case as reiterated by Gigerenzer and Goldstein (1996) who noted that humans display bounded rationality with constraints due to factors like limited capacity of working memory and our cognitive goals. Often one reasons just to achieve acceptable solution and not for optimal outcome. A new theory of reasoning has recently been put forth to explain why people do that. Their theory though still controversial, seeks to answer the puzzle of why at times we are so amazingly bad at reasoning yet there are times we are so good. This issue had been argued and debated by cognitive psychologists for decades. Mercier (2013) argued that we had been totally convinced thus far that reasoning can assist a person to be a better decision maker or believer following which we should improve in our reasoning capacity and do well in logical problems and statistics at large. There is ample evidence from studies that reasoning does not do all these very well.

From a psychological and education perspective, reasoning does not seem to function very well if done individually for abstract topics like mathematics or physics but if carried out collaboratively or in teams, the outcome of the reasoning and decision making processes are much better.

1.3.3 Errors in Human Cognition

Human cognition is very susceptible to errors. The sources of errors may arise from the decision making processes, conceptual base, beliefs, behaviors, social interactions or memory (Kahneman and Tversky, 1973). 'Error is the price we pay for quick and efficient processing of problem solving and decision making' (Riegler & Riegler, 2004). From a psychological perspective, errors are categorized as cognitive biases as explained by Riegler and Riegler, (2004). They are systematic errors related to issues of rationality or good judgement. There is much interest in the study of human cognitive errors? Kahnemann (1991) explained the emphasis one places on studying errors is for informativeness - i.e. understanding the conditions under which the thinking fails, can reveal important aspects of the human cognitive processes. Theories of memory distortions and the nature of automaticity revealed that we are susceptible to action slips. Olivier (1989) commented that from an "educational perspective, misconceptions are crucially important to learning and teaching, because misconceptions form part of a pupil's conceptual structure that will interact with new concepts, and influence new learning, mostly in a negative way, because misconceptions generate errors" (p.3). Olivier went on to 'distinguish between slips, errors and misconceptions'. Slips, he said are wrong answers due to the way we process information and they are characterized by carelessness, easily detected, not systemic and easily corrected. Errors on the other hand are incorrect answers that crop up during the planning stage. They are systemic and repeatedly appear under the same circumstances. Misconceptions are systemic conceptual errors caused by underlying contrary beliefs and principles deeply ingrained in the students' cognitive structures. Lèonard and Sackur-Grisvard (1987) provided a succinct explanation of the persistency of misconceptions among novices and even experts. They said, "Erroneous conceptions are so stable because they are not always incorrect. A conception that fails all the time cannot persist. It is because there is a local consistency and a local efficiency in a limited area, that those incorrect conceptions have stability" (p.444). In a study by Konold (1995) students correctly identified the different sequences of coin tosses that had equal chances of occurring. However, when asked differently i.e. which of the sequences was least likely to happen, they chose various sequences that were incorrect when in reality the answer for both questions should be the same. Interestingly enough Konold (1989) attributed this error to students who know the answer to the first question but when the question is rephrased, they use a different conceptual structure to answer. In other words, rote memorization has occurred but conceptual understanding is sadly missing. The students' incorrect intuitions are rather stable and it is really difficult to convince them otherwise (Konold, 1995; delMas & Garfield, 1991).

From an Information Processing point of view, reasoning rely very much upon the thought process and thereby causing the internal information to run into problems that sometimes give rise to misconceptions (Levitin, 2002). In his study on errors and incorrect intuitions, he found that the fundamental problems like lack of completeness of information in most real tasks; lack of precision; inability to keep up with change as internal information is very fluid and dynamic; heavy memory load in complex situation where retrieval of large amount of information is involved and finally a heavy computational load, would contribute to the frequency of making mistakes.

1.3.3.1 Approaches to the study of error

Two approaches have been proposed to measure the degree of error– normative and descriptive (Riegler & Riegler, 2004). Normative approach informs how one should think in a given situation as one will create a benchmark to compare all measures. The descriptive approach tells how a person actually thinks. Using these approaches, psychologists were able to study errors and misconceptions that people usually make. Heuristics or mental shortcuts afford a learner fast and efficient reasoning but sometimes they give rise to biases like representative biases, availability biases or confirmation biases. The representative biasness involves the tendency to assume that the characteristics of a sample should look like that of its population. An interesting item is given in this probability test item.

Which of the following sequences is most likely to result from flipping a fair coin 5 times?

a. H, H, H, T, T
b. T, H, H, T, H
c. T, H, T, T, T
d. H, T, H, T, H
e. All four sequences are equally likely

If a student chooses the options a, b or d, this student is not alone for these are some of the popular selections by undergraduate students. The answer is actually e. According to the Laws of Probability, the sequences given above have the exact same probability of happening. Law of Probability says that the probability of getting a head or a tail is 50-50. Unfortunately due to some misunderstanding with this law, we infer wrongly from the same law that in all the sequences given above (samples drawn from the same population), the number of heads and the number of tails should be roughly equal. Consequently we will most likely to choose options a, b or d as these sequences give a more balanced distribution of heads and tails. This biasness or misconceptions is known as the representative biasness. On the other hand, availability biases are due to errors in making the correct estimations. Generally it is assumed that objects in a category which come easily to mind are the objects that are considered more probable. Thus we tend to overestimate its chance of occurrence. Confirmation biases come about due to the tendency to find support for the hypothesis without considering other possibilities. One special case is function fixedness bias--the tendency to adhere to a single approach or a single way of using an object (Kalat, 2011). This issue was flagged earlier by the works of Mercier and Sperber (2009).

Errors happen for different reasons. People can reason well but still have a decision work out badly or we can reason badly yet still luck out into a good outcome. Kahneman and Tversky (1973) reasoned that prior knowledge and beliefs can retard the progress of valid reasoning as they showed with the availability bias and representative bias. Task and learner characteristics too do have some impact on the reasoning process (Schoenfeld, 1985).

Human error research has a lot of randomness or variations in results as information is never complete. This has been clearly shown in many studies concerning statistical misconceptions where findings are not conclusive with varying results (Garfield, 2003; Garfield & Ben-Zvi, 2008; Liu, 1998; Tempelar, 2004, 2007; Zuraida et al., 2012).

1.4 Model of Study

The a priori model for this study is primarily based on substantive literature review concerning the influence of three major cognitive determinants namely: prior mathematics knowledge, statistical reasoning and misconceptions held by Diploma students in a Malaysian university on statistical achievement. Figure 1.1 illustrates that statistical achievement of students is determined by three cognitive factors in a hypothesized manner as indicated by the one direction or bi-direction arrows. This study seeks to shed light on whether there is a production of a cause and effect (causation) as exemplified by the model. It does not seek to establish causality which can only be determined using a true experimental design.

Building this model takes into account the number of explanatory variables to use. It is important that the number is capped at a reasonable size to give the model enough explanatory power. Two approaches in determining selection of explanatory variables in this study are: 1) include only enough to make the model useful for theoretical purposes and to get enough predictive power. This is usually done through a thorough literature review. 2) For the purpose of counterbalancing the above, the researcher will keep the model simple as adding irrelevant variables only add little predictive power and causes multicollinearity. The model complexity is very much dependent upon the number of explanatory variables decided upon and this will determine the sample size.

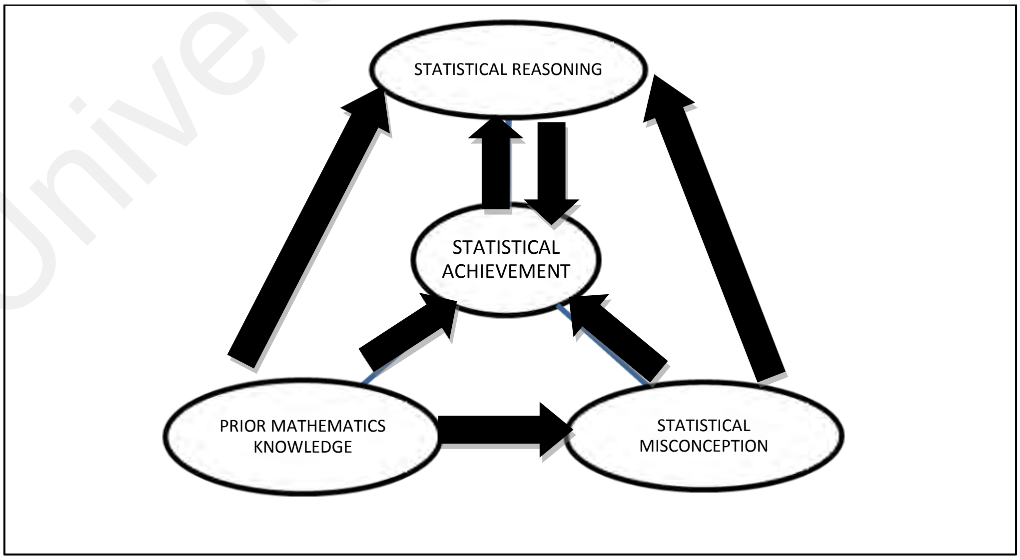

Figure 1.1: The Hypothesized Relationships among selected cognitive factors and statistical achievement using aggregated scores

1.4.1 Relationship between Prior Mathematical Knowledge (PMK) and Statistical Achievement (SA)

According to Wilkins and Ma (2002), there is evidence indicating a strong relationship between quantitative literacy i.e. abilities to perform quantitative tasks and statistical literacy. Another study found a positive correlation between highest mathematics grade-level completed, mathematics achievement and performance among students in an introductory statistics course (Lalonde & Gardner, 1993). Hulsizer and Woolf (2008) found a significant relationship between mathematics abilities and performance in statistics course and this has been reported in other studies (Nasser, 2004; Tremblay et al., 2000). Outcomes from studies by Chiesi, Primi and Morsanyi (2009); Chiesi and Primi (2010) and Zuraida et al. (2012) concurred with the above findings. Specifically what type of prior mathematical knowledge has the greatest impact on statistical achievement? Galagedera (1998) argued that basic working knowledge of algebra and set theory may be necessary though not sufficient. He added that authors of statistics books often indicated that a basic course in algebra is adequate to learn statistics concepts. Giraud (1997) using basic algebra test items to test students' readiness to learn college level statistics courses led him to the same conclusion.

These findings lend strong support to the impact of prior mathematics knowledge in particular algebra, on statistics course achievement. Curiously enough Noraidah et al. (2011) found that pre-university achievements do not affect their students' statistical achievement.

1.4.2 Relationship between statistical misconception and statistics achievement

Misconceptions in psychology or sciences are generally defined as preconceived ideas or intuitions where what one knows or believes to be true does not match what is correct scientifically. Misconceptions occur due to the reasoning process used when drawing conclusions from the premises or given information. The output from the inference process of reasoning can only be valid if the premises or information are valid. Faulty premises or errors in inference affect the truth or validity of the conclusions drawn. These wrong conclusions are one of the main sources of misconceptions. In reference to scientific misconceptions, psychologists McCutcheon (1991) and Best (1982) did not find any significant relationships between psychology course grades and their scores on misconceptions tests. On the other hand, Gutman (1979) found that there is a moderate correlation (r = .35) between grades in psychology and scores on a misconception-in-psychology test. Many researchers like McCutcheon (1991) and Best (1982) view misconceptions as an 'alternative perspective' of viewing the same construct. This happens when the perspective that one subscribed to does not match the current scientific view. From the constructivist point of view misconceptions are not that easy to 'erase' from the memory. Even with repeated teaching, the problems tend to resurface again. This is because the faults or errors have been integrated into part of the conceptual schema that will interact with new concepts and affect new knowledge in a negative way. In this respect, students who have developed misconceptions will inevitably face serious understanding issues in statistics classes. Many learners enter their classes armed with prior informal reasoning skills as explained by Schoenfeld, (1985). If these skill sets do not contradict with accepted statistical ideas then the learning process will be smooth. However they may come in with preconceptions that are intuitive and faulty then they are more likely to develop misconceptions (Schoenfeld, 1985).

Studies by Garfield (2003) and Tempelaar et al. (2007) have consistently shown that correlation between statistical misconceptions and course outcomes are non-existent and in the best scenario to be low. Evidence indicates different scales of the statistical reasoning scores by Garfield (2003) and Tempelaar et al. (2007) affect the course grades differently. This implies scores on SRA items are probably being moderated by some variables. One probable explanation would be that differing forms of misconceptions are affecting the students' achievement differently based on topics. It is a common fact that students are less confident in probability as compared to statistics. Topics like combination and permutations, conditional probability, probability distribution functions, sampling, variation and variability, uncertainty, randomness and many others are not favourite topics for many. Students coming in with faulty preconceptions in these topics do not help in their attempts to understand the topics.

1.4.3 Relationship between Statistical Reasoning (SR) and Statistical Achievement (SA)

Sedlmeier (1999) commented that perhaps if one is not to be condemned as poor probabilists one must seek solutions to improve one's reasoning process. Piattelli-Palmarini (1994) illustrated poor reasoners existed among politicians, generals, surgeons, and economists as much as among vendors of salami and ditch diggers. Sedlmeier (1999) defined reasoning as judgement under uncertainty while Garfield and Chance (2000) defined it as the way people reason with statistical ideas and make sense of the information. In statistics, learners are required to use reasoning to reach a conclusion after examining, manipulating and analyzing given information. It would seem logical to conclude that reasoning is a function of statistical achievement. Those with better reasoning ability should perform better in exams as compared to those who lack reasoning skills. However this was not the case. Research findings by Tempelaar (2004) and Garfield (2002, 2003) found little correlation between reasoning and achievement in statistics. Students may do well in exam, quizzes and class tasks but do rather badly on statistical reasoning tests. This has been attributed to surface learning and an apparent lack of understanding. Zuraida et al. (2012) found this no-relationship as with Tempelaar and co-researchers' 2007 study using aggregated scores. However they found low to moderate relation of Statistical Reasoning on course achievement. It seems to imply that statistical reasoning is content-dependent.

1.4.4 Relationship of Prior Mathematics Knowledge (PMK) and Misconception (MC)

Misconceptions are systematic conceptual errors cause by underlying contrary beliefs and principles deeply ingrained in the students' cognitive structures (Olivier, 1989). Students entering an introductory statistics course usually bring with them statistical reasoning as part of their 'prior knowledge'. These preconceptions are primal, intuitive knowledge comprising both declarative and procedural knowledge. Such knowledge is stored as 'true prior knowledge' in the long-term memory and can be accessed by the working memory when needed. If new knowledge were to merge with these errors, misconceptions are produced which unfortunately are stable over time and very difficult to 'erase' (Garfield & Ahlgren, 1988; Shaughnessy, 1992). Even with successful teaching of the correct statistical concepts, there is no guarantee that these misconceptions will not reappear under different circumstances. Students who perform well in computations and possess good statistical knowledge but shallow understanding are possible candidates for failure in reasoning.

In summary, among the more common misconceptions that will be studied are: 1) Misconceptions involving averages (mean, mode and median, 2) Outcome orientation (Konold, 1989), 3) Misconception about 'good samples have to represent a high percentage of the population', 4) Law of small numbers, 5) Representativeness bias (Kahneman, Slovic, & Tversky, 1982), 6) Equiprobability bias i.e. 'events of unequal chance tend to be viewed as equally likely' (Lecoutre,1992), 7) Availability bias and 8) Confirmation bias (Kahneman et al., 1982; Mercier & Sperber, 2011).

1.4.5 Relationship between Prior Mathematics Knowledge and Statistical Reasoning

Students entering introductory statistics course usually bring with them informal reasoning as part of 'prior knowledge' package (Olivier, 1989). Research carried out by Brown (1980,1990) provides some evidence that prior knowledge facilitates causal reasoning. Pragmatic knowledge is known to improve deductive reasoning on some conditional tasks (Cheng & Holyoak, 1985). Garfield (2002) studied the relationship between grades in statistics and statistical reasoning and found a significant association. However she noted that traditional homework problems do not correlate strongly with statistical reasoning scores. In other words, surface understanding in statistics is not enough for success in reasoning. A recent study by Tempelaar et al. (2007) found varying degree of associations between aggregated and disaggregated statistical reasoning scores with different mathematics course grades taken previously. He noted that the impact of prior mathematics education on both correct statistical conceptions and misconceptions were small. There was a higher conception score with more advanced mathematics programs. Analysis of disaggregated reasoning scores with different levels of mathematics courses taken previously do show some low to moderate correlations. Zuraida et al. (2012) found a moderate association between prior math knowledge and statistics reasoning (r = .56) using aggregated reasoning and achievement scores.

1.5 Moderating Variables

Higher cognitive processes, of which reasoning, problem solving and decision making are some examples, depends not only on their intrinsic characteristics, but also between the processes and the owner of the process acting in a social context (Schoenfeld, 1985). This implies that the learner characteristics and the social setting will have an impact on the reasoning process. The current study intends to look at two characteristics of the learner i.e. gender and the language mastery that is hypothesized to moderate the proposed model of study. Moderating factors are variables that influence the strength of the association of an independent variable on the dependent variable. Moderating variables can be discrete or continuous data.

Hair, Anderson, Tatham and Black, (1999) defines moderator as a variable that can cause the relationship between a dependent/independent variable pair to change, depending on the value of the moderator variable. This moderator effect is commonly known as interaction effect as it is known in ANOVA.

According to Baron and Kenny (1986) they stated that a variable is a moderator (i.e. qualitative or quantitative variable) if it affects the direction and/or strength of the relation between an independent and a dependent variable. In a correlational design, a moderator is a third variable that influences the correlation between the IV and DV. A suitable moderation framework can be diagrammed as shown in Figure 1.2.

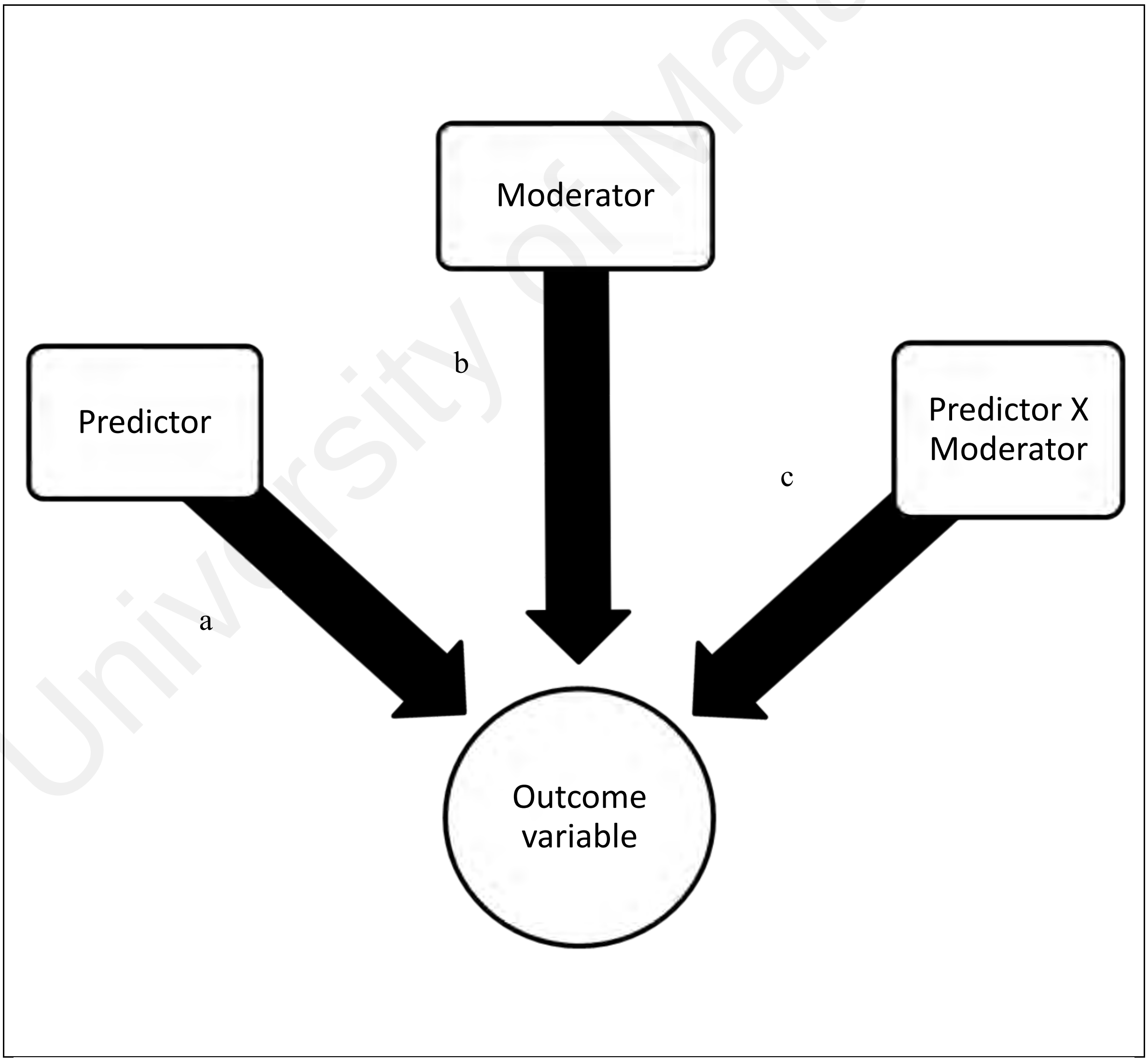

Figure 1.2: A Moderator Effect Framework for a Correlational Design (Baron & Kenny, 1986)

The diagram shows three causal paths linking to the DV which is the outcome variable. Each path is signified by an alphabet. Path ‘a’ indicates the effect of the predictor variable on the outcome variable. Path ‘b’ shows the influence of the moderator on the outcome variable while path ‘c’ shows the effect of the product of the predictor and moderator on the outcome variable. A moderation effect is considered present if path c is significant statistically. The significance of path ‘a’ and path ‘b’ is not important when testing for moderation in this framework.

Dawson (2014) showed how a moderation effect can be tested and interpreted for an Ordinary Least Square Regression model. Assuming this equation,

$$Y = b_0 + b_1X + b_2Z + b_3XZ + \mathcal{E} \qquad \textbf{1.1}$$

where Y is the outcome, X the predictor, Z the moderator and XZ the product. To test this two-way interaction, one only needs to check if the product effect is significant. This can be done by calculating the ratio of the coefficient b_3 to the standard error of XZ with a known distribution (in some cases, a t-distribution with n - k degrees of freedom, where n is the sample size.

According to Dawson (2014) it is important to make a logical choice of using the X and Z variables in their original format or to mean-center the two variables. He went on to explain that it makes little difference in its moderation effect detection in most cases.

One of the approaches discussed above is used to test if gender and language moderate the various relationships as visualized in Figure 1.1.

1.5.1 Gender Effect and Statistical Achievement

Studies on the effect of gender and mathematical achievement have been inconclusive as the discussion below will show. Brooks (1987) and Elmore and Vasu (1986) found that female students did better in mathematics grades over time. However, Buck (1985) did not find any significant influence of gender on introductory and advanced undergraduate statistics course grades over 13 semesters. A meta-analysis carried out by Schram (1996) based on 13 articles came to a conclusion that men did better than women when examinations were used as the criterion for overall achievement scores. On the other hand, females did better if formative assessment was used to aggregate the final achievement score. In more recent studies, Noor Azina and Azmah (2008) found mixed results among undergraduates in a Malaysian public university with no clear distinction of academic abilities between male and female. Another study by Chang and Cheo (2012) showed that gender does not play a role in academic achievement of Economics major students in NUS and UM. Ding, Song, & Richardsons (2006) found that both male and female students demonstrated the same growth trend in mathematics achievement over time, but females' mathematics grade-point average was significantly higher than males'.

Liu (1998) found that gender affects statistical reasoning on performance for Taiwan respondents but not for USA students. Other studies by Garfield (1998); Garfield and Chance (2000) and Tempelaar et al. (2006) similarly detected significant gender influence. Tempelaar et al. (2006) noted that gender effect was identified despite similar educational background. Martin (2013) found evidence of the existence of a gender gap in statistics in which males Canadian outperform the females.

Reilly (2012) found many of the cognitive skills show an interaction effect between gender and socioeconomic status. Hyde, Lindberg, Linn, Ellis and Williams (2008) examined gender differences in mathematics from the second grade to the eleventh grade drawing their samples from US population. Their results revealed the relationship between gender and mathematics was relatively insignificant. In another meta-analytic study by Reilly (2012) with secondary data sourced from 65 OECD countries participating in the PISA survey, stated that: 1) 'Gender differences in mathematics literacy were comparatively larger for the United States than those found across other OECD nations. This difference is most apparent when examining student attainment of the highest proficiency level in mathematics, with double the amount of boys than girls reaching this stage…'

Men and women do not differ in IQ scores, vocabulary tests, or reasoning tasks (Levitin, 2002). He went on to explain that the nature of the sex differences depends on how cognitive skills are measured whether tests are measured using spatial tests, oral tests, objective tests, essay tests or mental tests.

The mounting number of conflicting findings implies a clear lack of conclusive answer as to whether there is any interaction effect of gender on the relationship described. Hence this study hopes to contribute to the body of evidence on gender effect and interactions with cognitive factors used in this study.

1.5.2 Language Effect and Statistical Acheivement

Cognition means thinking and using knowledge (Kalat, 2011). This is the realm of cognitive psychologists who are interested in understanding the cognitive/mental processes by which stimuli from outside are transformed into meaningful information, stored, retrieved, applied and communicated to others.

The product of thinking is known as thoughts. Language is a medium for a person to communicate one's thoughts through the use of complicated rules that helps to form and string together symbols thus generating meaningful sentences or utterances.

Thoughts and language are two closely related cognitive processes that are dynamic and complex. Language facilitates and expresses those thoughts through sound and symbols (Bransford, Brown & Cocking, 1999).

Language is defined as a special form of communication that combines symbols and words, guided by a set of complex rules to form meaningful sentences or sounds. The success of this form of communication is attributed to two simple but amazing principles – words and grammar. The medium of instruction in an introductory statistics course will obviously have significant influence on statistical performance especially if the medium is not the native language. Therefore the effect of prior linguistic knowledge of learners on the comprehension of a context-laden text needs further research (Reed, 2013*)*. By extension, learner comprehension of the text will affect their achievement in the exam papers.

An important aspect to thinking is the question of the relation between language and general intelligence and whether one can develop intelligence without language or learn language without certain aspects of intelligence. Intelligence is commonly taken to refer to the ability to understand information, plan, learn, use language and solve problems with the assistance of complex cognitive processes like reasoning and related thinking. Psychologists have discovered that one can still develop one's intelligence independent of language (Kalat, 2011).

According to Dwyer (1973), boys are generally poorer readers and writers than girls in reading literacy. This fact is further strengthened in the meta-analytic study by Reilly (2012). He suggests within the United States, girls outperformed boys in overall reading. Studies also reported similar findings; girls were better readers than boys across most nations. The items in a reasoning test like the Statistical Reasoning Assessment (SRA) are fairly long and worded in technical terms which need a good degree of comprehension and interpretative skills. Girotto (2004) asserted that much of the difficulty of reasoning lies with understanding the language. Reed (2011) noted that organization of the text in an item or the story structure has an effect on performance. Shaughnessy (1992) added if the context of the test item is abstract, the achievement on this item is much lower but if put into familiar context the success rate increased significantly. The mathematical language that is employed in test items also influence the success rate in solving reasoning tasks. Gigerenzer and Hoffrage (1995) presented a well-known Bayesian inference task to a group of students using two formats – one using probability format and the other using frequency format. The frequency format yielded better results than using the probability format. A similar study by Cosmides and Tooby (1996) concurred with the findings described earlier. Items in probability format are viewed to be ‘mathematical’ while the frequency format was more ‘ordinary-looking’ i.e. a format in a layman’s term.

As thinking and language mastery are closely linked psychologically with gender differences (Ding, Song, & Richardson, 2006), it is inevitable to hypothesize that language mastery plays a moderating role in the relationship between the cognitive determinants.

1.6 Purpose of the Study

This study aims to investigate the various relationships of cognitive determinants such as prior knowledge, statistical reasoning and statistical misconceptions among others that had been identified a priori to influence statistical achievement of Malaysian Diploma students in an introductory course. In addition, this study attempts to identify factors (e.g. gender, language mastery) that are hypothesized to have an indirect effect on the various relationships between the independent variables like prior mathematical knowledge (PMK), statistical reasoning (SR) and misconception (MC) on the dependent variable; statistical achievement (SA).

1.7 Objectives of the study

This study is designed to achieve the following objectives:

i. To determine the relationships between statistical achievement and the predictors (i.e. prior mathematical knowledge, statistical reasoning and statistical misconception)

ii. To assess the effect of gender and language mastery on the relationships as mentioned in the objective above.

iii. To determine the relationships between statistical reasoning and the predictors (i.e. prior mathematical knowledge, statistical misconception)

iv. To assess the influence of gender and language mastery on the relationships as mentioned in the objective above.

1.8 Research Questions

i. What cognitive determinants affect the students' statistical achievement in an introductory statistical course?

ii. What is the regression model that expresses the relationships among the cognitive determinants that affect students' statistical achievement in an introductory statistical course?

iii. What cognitive determinants affect the students' statistical reasoning in an introductory statistical course?

iv. What is the regression model that expresses the relationships among the cognitive determinants that affect students' statistical reasoning in an introductory statistical course?

v. What is the moderating effect of gender on the relationships among the cognitive determinants?

vi. What is the moderating effect of language mastery on the relationships among the cognitive determinants?

1.9 Delimitations of the Study

This section describes the scope and the boundaries set when designing this study. Thus the important delimitations are described below:

i) The participants were selected using a purposive sampling technique. This was due to the ease of accessibility and proximity of the participants to the researcher.

The sample is not representative of the population chosen for this study. In this sense, the research findings were limited in its generalizability to the population.

ii) The participants were all *Bumiputera* or the indigenous people of the land. All of them spoke Malay language, the national language but used English as the medium of study.

iii) The instrument SRA were monitored, piloted and verified by the researcher but the out-of-class assessment scores like *SPM* results, past semester examination results were entirely self-reported

iv) The topics covered and the questions asked in the quizzes, tests that formed part of the scores of their statistical achievement covered some basic algebra skills and introductory statistics taught in the students' secondary education.

v) Multiple regression analysis was considered a more suitable tool for this study among many other techniques like Structural Equation Modeling where measurement errors of the variables of interest can be ignored in regression analysis. In addition, multiple regression is used because of the constraints arising from the SRA instrument and the nature of data to be collected of which will be discussed in the later chapters.

1.10 Limitations of the Study

Limitations are shortcomings, conditions or influences that cannot be controlled. They can place restrictions on the methodology and conclusions reached at the end of the study. The key limitations are discussed below:

i. The findings in this study cannot establish causality. All relationships in this study are hypothesized from literature review. Great care had to be taken in interpreting the outcomes of the linear regressions as establishing causal relationships. While regressions of cross sectional data can reveal associations, they usually do not document time order. Thus the findings indicate only associations and to determine causality from observational data is difficult.

ii. The findings may not be generalized beyond a similar population where this sample had been chosen. The demographics of this university diploma students are fairly unique and homogeneous

iii. The findings cannot be generalized to other courses except for introductory statistics

iv. Some of the data were collected from a self-reported survey form.

1.11 Definition of Terms

The key terms to be used in this study are defined as in the following:

i. Cognitive Determinant

Cognitive determinant is a factor that is used to characterize an individual's learning and achievement. It serves to modulate the person's performance (Danili & Reid, 2006). This factor pertains to mental processes such as perceiving, knowing, remembering, thinking, problem solving, and decision making. In the context of this study, three main cognitive determinants are identified as prior mathematical knowledge, statistical reasoning and statistical misconception in the multiple regression model.

ii. Statistical Achievement

Statistical achievement is defined as the ability of a student to master the basic statistical skills and knowledge over time that enable them to progress to a higher level of statistical literacy, reasoning and thinking (Miller, 1999). This can be measured using grades through both formative and summative assessment like quiz, test and examination that serve as proxy to learning outcomes and competencies (Kooi & Ping, 2006; York, Gibson & Rankin, 2015). An aggregated score calculated from marks collected from the respondent's quizzes, tests and final examination taken for the semester will be used to represent a student's statistical achievement in the Regression Model.

iii. Statistical Reasoning

Statistical reasoning is defined as the way students reason with statistical ideas and make sense of statistical information (Garfield, 2003). According to Garfield (2003) the Statistical Reasoning Assessment (SRA) instrument can be used to collect information about a student's reasoning ability.

iv. Prior Mathematical Knowledge

Prior Mathematical knowledge represents knowledge that encompasses both declarative and procedural mathematical knowledge; and is relevant to the achievement of the objectives of the learning outcomes in a particular mathematical course. The knowledge to be considered is both subject-oriented prior knowledge and domain-specific prior knowledge (Hailikari, 2009). In this study, to measure prior mathematical knowledge collectively, the grades that a student received in their finals during their university years and secondary school years are employed as representative of their prior knowledge.

v. Statistical Misconceptions

Misconceptions are systematic conceptual errors caused by underlying contrary beliefs and principles deeply ingrained in the students' cognitive structures (Olivier, 1989). Although this is a complex construct for the purpose of this study, the method used by Garfield (2003); Tempelaar (2006) and Martin (2013) in scoring a student's misconception through the SRA instrument will be employed.

1.12 Summary

Studies have shown the lack of real understanding among students who have 'passed' introductory statistics or quantitative methods courses but are still weak in statistical reasoning and thinking. This can be seen in the recent 2011 TIMSS report on Fourth and Eighth Grade students in Mathematics. Malaysia continues to show a decline in the mathematics achievement with the component of 'Data' and 'Chance' section faring the worst. This international survey (Mullis et al., 2012) found a strong positive correlation between content domain and cognitive domain. Statistical reasoning is a crucial cognitive skill to master and it is related to the content knowledge of the students. Nonetheless present efforts by psychologists and statistics educators still could not unravel the varied learning difficulties inherent in the complexities of statistical knowledge and understanding. Statistical learning difficulties are related to a multitude of factors. Some factors of concern in this study focus on the cognitive domain like reasoning, misconceptions and prior knowledge. This study aims to determine the various cognitive determinants that affect how students perform in probability and statistics while concurrently testing to see if other factors like language mastery and gender exert any influence on the determinants.

CHAPTER 2 : LITERATURE REVIEW

2.1 Introduction

Statistics is a highly sophisticated process to express the representational and inferential properties of the data both numerically and visually. The appropriate usage and optimal utilization of statistics assures results that can provide useful information for solving problems and making good decisions. Mathematics can be highly abstract yet still comprehendible. However, this cannot be said of statistics for it requires a context to frame the problem meaningfully. Sometimes a student may do well mathematically but not so with probabilistic thinking. Students and even mathematics teachers find the topics in probability to be comparatively difficult to handle and sometimes even baffling. For instance, in algebra $A = 3, B = 5$, therefore, $A + B = 8$. In probability on the other hand, $P\ (A) = 0.3, P\ (B) = 0.5$ but P (A U B) is sometimes equal to 0.8 but sometimes it is not (Foo & Noraini, 2010).

Theoretical probability cannot be proven to be absolutely true even after running hundreds of trials. At times students develop conflicts trying to assimilate probability ideas into developed mathematics concepts in statistics class (Foo & Noraini, 2010). The next section discusses about the teaching and learning of statistics in Malaysia and in particular statistical literacy, reasoning and thinking.

2.2 Statistics Education in Malaysia

Students in Malaysia are taught basic statistics at the age of 9 and continue to the age of 17 covering data handling, presentation of data using tables, pictures or chart and concept of average in the primary education. In the secondary years, topics include frequency using tally chart and frequency table, data collection methods and basic ideas about probability and statistics. A-Level Mathematics or its Malaysian equivalent covers more complex concepts in data description, probability and statistics. Advanced topics that are offered as optional include discrete and continuous probability distributions, sampling and estimation, correlation and regression in addition to time series and index numbers.

2.2.1 The teaching and learning of statistics

Statistics and its related process skills are very much needed now in the 21st century where data and information rules the world of information technology. Moore (1990) observed that, ''Statistics is a general intellectual method that applies wherever data, variation, and chance appear. It is a fundamental method because data, variation, and chance are omnipresent in modern life" (p. 134). Data management skill has garnered enough attention lately in many schools in various countries. With this realization, curriculum changes at the school level in many countries are happening (Watson, 2009). The new curricular changes are deemphasizing computations and fact memorization and instead providing more hours for active learning, understanding and thinking using real data and context. In addition, learning goals are designed from the bottom up where input from teachers and educators are taken into account into curriculum design (American Statistical Association, 2005a, 2007).

Undoubtedly statistics is a difficult subject matter in classes. It can be difficult to understand. Students may even show good command of propositional and procedural knowledge in tests and examinations, but the fact remains-many students find it difficult to interrelate and structure their knowledge (Broers, 2009). These students lack strong statistical foundation because of weak conceptual understanding.

To facilitate the learning process, educators and researchers are beginning to understand students' statistical knowledge structures and conceptions as well as how these concepts develop (Roseth, Garfield, & Ben-Zvi, 2008). In addition, psychologists studying reasoning realized the advantages of this approach to learning of reasoning in the classroom (Mercier & Sperber, 2011).

2.3 Assessment in Statistics

A recurring educational issue across many countries in Asia is the problem of exam-oriented teaching. In a paper by Foo & Noraini, (2010), it was said that Asian society valued excellent examination result too highly, giving emphasis to more focus in answering examination questions. A consequence of this approach is that 'difficult' topics are compromised and understanding of students 'short-changed'. If nothing is done to correct the situation at the primary and secondary level, the task of equipping undergraduates with strong statistical foundation and skills so that they are able to utilize statistics effectively is difficult.

2.3.1 Purposes of assessment

Traditionally, assessment had placed too much focus on summative aspects like tests and examinations while giving less weightage to formative forms of assessment. With changing views concerning assessment in today's curriculum, emphasis has moved to developing strategies to evaluate students' understanding and reasoning processes as well as their learning skills. Ben-Zvi and Garfield (1999) saw assessment as encompassing the following purposes: promote growth, improve instruction, recognize accomplishment and modify program through strategies like monitoring of students' progress, making good instructional decisions, evaluate students' achievements and evaluate program effectiveness. Educationists viewed assessment in broader term stating that the purpose of assessment includes: a) to assist learning, b) to measure individual achievement and c) to evaluate program (Pellegrino, Chudowsky, & Glaser, 2001). The basic elements underlying assessment are cognition, observation and interpretation. These three foundational elements according to Pellegrino et al. (2001) must be present in all formative and summative assessment in an integrated and connected whole.

2.3.2 Taxonomy for assessing statistics educational outcomes

The widely-used model to measure cognitive abilities in education is the Bloom's Taxonomy developed in 1956 and still considered to be one of the best classification approaches for educational outcomes. Educational outcomes are products of the learning process and can be measured by Bloom's classification of educational outcomes. He classified the outcomes into the following: Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation.

To differentiate the hierarchy of cognitive objectives, educationists use specific words to characterize them. These words form the basis for constructing test items at each level. For example, at the Knowledge level, one knows one is evaluating the cognitive ability of students at this level, if they can answer questions that used these words- arrange, define, describe, duplicate, identify, label, list and match. At the Comprehension level - classify, convert, defend, describe, discuss, distinguish, estimate, explain and generalize. At the Application level - apply, change, choose, compute, demonstrate, discover, illustrate, interpret, and operate while at Analysis level - analyze, appraise, breakdown, calculate, categorize, compare, contrast, criticize...etc. At the Synthesis level- arrange, assemble, categorize, construct, design, develop, formulate and generate. Finally Evaluation level- appraise, argue, assess, explain, rationalize, predict, judge, interpret, justify. When one compares across the six categories one will find that some words or their synonyms are not exclusive to any one category. This overlapping makes the taxonomy difficult to use.

The Bloom Taxonomy reflects a hierarchy of abilities starting from the lowest cognitive ability (Knowledge) to the highest thinking outcome (Evaluation). This taxonomy has been used to design test items for evaluating cognitive objectives (Garfield & Ben-Zvi, 2008). Although useful, this taxonomy has been criticized by item developers for its many constraints and limitations, one of which is the difficulty to place certain cognitive objectives into their correct levels. This is due to the overlapping between categories (Seddon, 1978). Statistics educators suggest alternative but simpler taxonomy to statistics item building (Garfield & Ben-Zvi, 2008). They have found that building statistics items according to the types of statistical cognitive processes is viable. They believe that all statistical mental processes can be separated into: a) statistical literacy, b) statistical reasoning and c) statistical thinking.

Statistical literacy refers to an understanding and using of basic language and tools of statistics: know basic statistical terms, understand basic statistical symbols, recognize and interpret visual and graphic representations of data (Rumsey, 2002)

Statistical reasoning refers to the way people reason with statistics and makes sense of statistics information: connecting concepts, understanding statistical ideas and concepts at a deeper level than statistical literacy (Garfield, 2002).

Statistical thinking refers to higher order statistical mental processes compared to literacy or reasoning: thinking usually done by professional statisticians, deep understanding of the theories underlying statistical process and methods (Wild & Pfannkuch, 1999).

delMas (2002) provided a list of words that characterized test items for literacy, reasoning and thinking as parallel to that given in the Bloom's Taxonomy as listed in Table 2.1.

Table 2.1: Words used for Different Assessment Items or Tasks (delMas, 2002)

Literacy	Reasoning	Thinking
Identify	Explain why	Apply
Describe	Explain how	Critique
Translate		Evaluate
Interpret		Generalize
Read		
Compute		

Literacy here is equivalent to Bloom's ‘Knowledge’ level while Reasoning is similar to ‘Comprehension’. Statistical thinking category is equivalent to ‘Application’, ‘Analysis’, ‘Synthesis’ and ‘Evaluation’ in Bloom's Taxonomy (Garfield & Ben-Zvi, 2008). Since deMas’s (2002) taxonomy is parallel to Bloom’s thus, it is predicted to inherit some of its limitations as well.

This problem is compounded by disagreements among statistics educators over the meanings of each of these terms (Rumsey, 2002; deMas, 2002; Garfield & Ben-Zvi, 2008; Sedlmeier, 1999; Tempelaar, 2006).

For the purpose of this study, the terms are defined accordingly to the ones agreed upon by many of statistics educators (deMas, 2002; Garfield & Ben-Zvi, 2008) and this study investigates the Reasoning category which is comparatively defined in its usage among statisticians compared to the other two categories that are still being hotly debated as to their precise definitions.

2.3.3 Assessing Statistical Cognitive Outcomes

Statisticians had always stressed on conceptual understanding and a variety of strategies to achieve good grades in statistical outcomes (deMas, 2002; Garfield & Ben-Zvi, 2008). Unfortunately the question of how to assess statistical cognitive outcomes took a backseat during this period. The importance of knowing how students think about probability and identifying effective instructional approaches seem to take precedence over developing valid and reliable methods of assessment that measure students' conceptual understanding (Shaughnessy, 1992). Other researchers too reiterated the fact that there were clearly less emphasis given to instructional methods or assessments (Konold , Pollstek, Well, Lohmeier & Lipson, 1993; Lipson, 1990; Garfield & Ben-Zvi, 2004).

Attention now has since shifted to a more equitable share between understanding, learning approaches and assessment. Traditional methods of assessing using quizzes, tests or examinations are increasingly coming under attack (Martin, 2013). The reason is that students are provided with only single summary scores to reflect their achievements over a long span of learning. Undoubtedly this assessment of the students' learning experience is inadequate. Due to the intrinsic weaknesses, statistics educators have recommended a move to more inclusive strategies and approaches that can reflect learning outcomes comprehensively. It is thus a challenge for statistics educators to construct and test out assessment tools that can measure effectively the different kinds of conceptual understanding in a statistics class. In addition, most introductory courses in statistics cater for a large number of students making it mandatory that administration of any assessment must be easy to manage, economical, time- and cost- effective. A good example of such an assessment instrument is the Statistical Reasoning Assessment (SRA) by Garfield (2003) that contains 20 multiple choice test items to measure the reasoning abilities and misconceptions of the students.

The SRA assessment tool has distinct advantages over traditional assessment in that it measures statistical development and achievement, is easy to score, covers a wide range of statistical content and can be given to large classes. The present study seeks to use this instrument to measure statistical reasoning and misconceptions.

2.3.4 Designing Assessments for Statistics Classes

The National Council of Teachers of Mathematics (NCTM, 1995) outlines six assessment standards that place greater importance on how one assesses mathematical and statistical content and the thinking processes. Consequently designing any assessment plan needs to take into considerations the following when preparing the framework (Garfield, 1994): a) what is to be assessed (the concept, skill, attitude or belief); b) the purposes of the assessment (to give a grade, to improve the teaching and learning process, or to identify errors in conceptual understanding); c) who does the assessment (self-assessment, instructor assessment or national assessment); d) the method of administering the assessment (quizzes, tests, examinations, project): e) the follow-up actions or feedback that are to be implemented after the assessment. These aspects are important factors to consider when designing an assessment tool to ensure it is aligned to the course goals and provide optimal information for the follow-up activities.

2.3.5 Different ways of assessing statistical knowledge

Statistical knowledge can be measured by way of traditional assessment methods like quizzes, tests, examinations. Although this approach is very much alive today, there is a distinct trend towards measuring higher mental statistical thinking that requires different assessment approaches. Alternative methods are available but Garfield and Ben-Zvi (2004) opined that a combination of both traditional and alternative methods allows instructors to assess a student's understanding at a deeper level and at the same time identify common misconceptions in probability and statistics that are hampering their advancement in achieving higher-order thinking. Garfield (1994) and Garfield and Ben-Zvi (2008) suggested possible assessment methods which include: homework, quizzes, minute papers, group projects, case studies or authentic tasks, critiques, concept mapping, portfolios, lab reports, and reflective journal writing. Some of the methods used in their study are elaborated as follows:

2.3.5.1 Quizzes, tests and examinations

Traditionally in any courses these three methods are used to assess how students are progressing and what they had achieved at the end of the courses. These methods are invaluable assessment tools. According to Garfield and Ben-Zvi (2008) quizzes as a form of formative assessment can provide timely information to instructors on how their students are progressing with respect to their procedural and conceptual understanding. Short quizzes or pop quizzes can be important assessment tools to keep students focus and pay attention. Well-designed quizzes or tests can be very helpful in providing students with the required experience to answer the types of questions asked in the examinations.

According to Hubbard (1997), setting questions for an exam can be a challenging task especially for novice instructors. These instructors have to take various matters into considerations namely - aligning test items to the course objectives, providing meaningful context to each item, and constructing items that assess higher order thinking skills. Tests and examinations do not necessarily ask for open-ended questions but can be given in the multiple choice format. Cobb (1998) suggested techniques to construct items that can be used to evaluate higher order thinking and reasoning. If the task to design good items is beyond the ability of instructors, there are ample selections of good statistical items available online in the ARTIST website for members but they are not freely obtainable for students (Garfield & Ben-Zvi, 2008).

As the main instrument used in this current study, the SRA is a multiple-choice test, pilot study is necessary to assess its suitability to the local population and local context before administrating it in the real study. To improve an instrument's validity and reliability, it is important to investigate the appropriateness and soundness of the constructed items. According to Wild, Triggs, and Pfannkuch (1997) multiple choice statistics items can test higher order thinking skills as well as identify common misconceptions, interpret data, select correct techniques for data analysis and make inferences. However they cautioned that these items cannot assess thinking processes qualitatively nor evaluate open-ended questions. Garfield and Ben-Zvi (2004) provided guidelines for developing items for quizzes and examinations. The guideline will be used to assess the soundness of the items in SRA (Garfield, 2003) during the pilot stage of this study.

- items must be able to assess students' reasoning and thinking as well as demonstrate their use of statistical language

- each item ideally should have 3-4 options. Make full use of each option to reflect the different reasoning or thinking processes that are correct and incorrect. The options should be able to help identify students' errors and misconceptions. Try to avoid options like 'none of the above'
- make sure there is a contextual basis to the items and avoid turning the items into computational questions.
- build the items from existing data of relevant research study which may be of interest to the respondents.

2.3.5.2 Homework

Homework assignments are means to reinforce the skills and knowledge that were learnt recently. They serve to provide constant practices in the usage of terms and computational processes to give students understanding and confidence. The assignments must not be limited to memorizing and computing but include opportunity to answer application and conceptual questions to reflect the problem-solving process.

Grading of these assignments is essential as it gives valuable feedback that students can use to apply to other similar assignments and get an idea of how grading is done in the exams (Garfield & Ben-Zvi, 2008). Paired or collaborative assignments should be encouraged as more learning will occur directly or indirectly as students argue, debate and rationalize their responses and finally the students come to a common conclusion. This support or scaffolding structure not only provide increased learning opportunity but also alleviate anxiety of assignments, quizzes and tests.

In conclusion, using a range of continuing assessment methods together with tests and examinations can efficiently measure statistical achievement.

2.3.6 Assessing Achievement in statistics class

The term achievement has been used loosely and has given rise to different interpretations when used in different contexts or by different authors. Achievement is synonymous with terms such as performance, competency, ability or accomplishment. In education, the general term educationists are more familiar with is academic achievement. Pinilla and Munoz (2005) explained that academic achievement takes into account grades, time in an educational institution and number of related courses taken per year while Allen (2005) sees academic achievement as the summed total of the final grades a student achieved with respect to course content and knowledge. Similarly, Kooi and Ping (2006) considered Grade Point Average (GPA) as the basis for a student's academic achievement. Academic achievement is differentiated from academic performance in the context of this study. Achievement is the outcome from an academic endeavor while performance is the process leading to an achievement.

Darling-Hammond and Adamson (2010) see achievement assessment as not a traditional multiple-choice testing where facts and computations are emphasized. The assessment of statistical reasoning in this study used an instrument that consists of analytically-oriented multiple choice response items while statistical achievement is assessed based primarily on scores obtained throughout the semester through the administration of assignments, homework, quizzes, tests and final examination.

2.4 Information Processing Theory (IPT)

Information Theory was an important breakthrough for the field of cognitive psychology. It suggested that information was communicated by sending a signal through a sequence of stages or transformations. This concept about human perception and memory was new and revolutionizing. This was the start of the information processing approach—the theory that cognition could be perceived as a flow of information within the organism is a concept that still continues to dominate cognitive psychology. Perhaps the first major theoretical effort in information processing psychology was Donald Broadbent's Perception and Communication (1958). Broadbent‘s hypothesis about the transfer of information from short- to long-term memory, became the important point of the dual memory models developed in the 1970s. Another aspect of Information theory that attracted psychologist's interest was a quantitative measure of information in terms of bits as used by George Miller in his widely cited 1956 paper (Miller, 1956). These were among some of the important mileposts in the development of IPT

2.4.1 Information Processing Model and the Computer

IPT is a theory used by cognitive psychologists to analyze, describe and elucidate the mental processes (Anderson, 1977). The model finds parallels in the working of a computer. Like a computer, the mind receives information externally, organizes and stores it in a form that can be accessed at a later time. Data or information is keyed in using a keyboard or scanner. In humans, the input devices are the sensory organs like the eye, ear, nose, skin or tongue. It is through these organs that a person receives information about its surroundings. The computer's Central Processing Unit is equivalent to the Working Memory or Short-Term Memory. In human, all information is stored for a brief moment, giving the brain enough time to be used, discarded, or transferred into long-term memory (LTM). Information stored on a hard disk is equivalent to that stored in the long-term memory. Information kept in the LTM is stored for a long period of time. A computer processes information and displays its results on a screen or in the form of a printout while results of human processing of information are translated into various forms of behavior or action.

2.4.2 Stage Model of Information Processing

One significant but difficult area of research in cognitive psychology is the empirical study of memory. Present day cognitive psychologists are still holding to the dominant view of the "stage theory" by Atkinson and Shiffrin (1968). This was an important theory to assist researchers to understand the relationship between learning and memory which is closely related but could not be verified or observed visually. Learning and memory are complex but necessary cognitive functions. The brain processes millions of data each second and stores them away in the form of useful information. It keeps evolving and changing every second as a person learns and takes in new information.

Memory is the ability to retain information over time through three processes – encoding, storing and retrieving. Encoding is the process of making mental images of the information so that one can keep in one's memories. Storing is where a person puts the encoded information in locations where one can retrieve when needed. Retrieving is the process of recalling that information from the short-term or long-term storage (Plotnik & Kouyoumdjian, 2011). Human memory can be visualized as consisting of components in Figure 2.1.

Recent studies by cognitive psychologists have indicated that the sequential information processing proposed by Atkinson and Shiffrin (1968) may be too simplistic to explain complex mental processes like reasoning, decision making and higher order thinking. Two other models currently in contention as alternatives are the parallel-distributed processing model and the connectionist model which suggest that information is processed concurrently at several parts of the memory locations (Huitt, 2003). The connectionistic model expounded by Rumelhart and McClelland (1986) is an expanded version of the parallel-distributed model. This model proposes that information is not stored in one location only but rather at multiple locations throughout the networks of connections in the brain. Brain research by Rumelhart and McClelland (1986) has found that the more connections a particular idea or concept has to other neural networks, the more likely it is to be remembered. Importantly this model propounds the principle that the brain learns through experience with constant exposure to stimuli from the outside world.

2.4.3 Basic Principles of Information processing approach

The information processing approach is based on a number of principles, including:

I. The memory capacity of the brain is limited at some locations of the system such as the sensory memory and working memory that leads to serious constrictions to the flow of information for processing (see Figure 2.2).

II. The processing units in the brain that attend to encoding, transformation, storage, retrieval and synthesis of information must be monitored and coordinated by a control mechanism.

III. In the attempt to make sense of the world around a person, the brain employs a 'two-way flow of information' (Huitt, 2003) known as 'bottom-up processing' and 'top-down processing' depending on whether the information is from outside or information retrieves from the long-term memory.

IV. The brain's processing system changes information in a systematic way as all human are genetically engineered to process and organize information in a specified manner. Research in language development among infants has provided convincing proof (Huitt, 2003; Rumelhart and McClelland, 1986)

2.4.4 Types of Memory

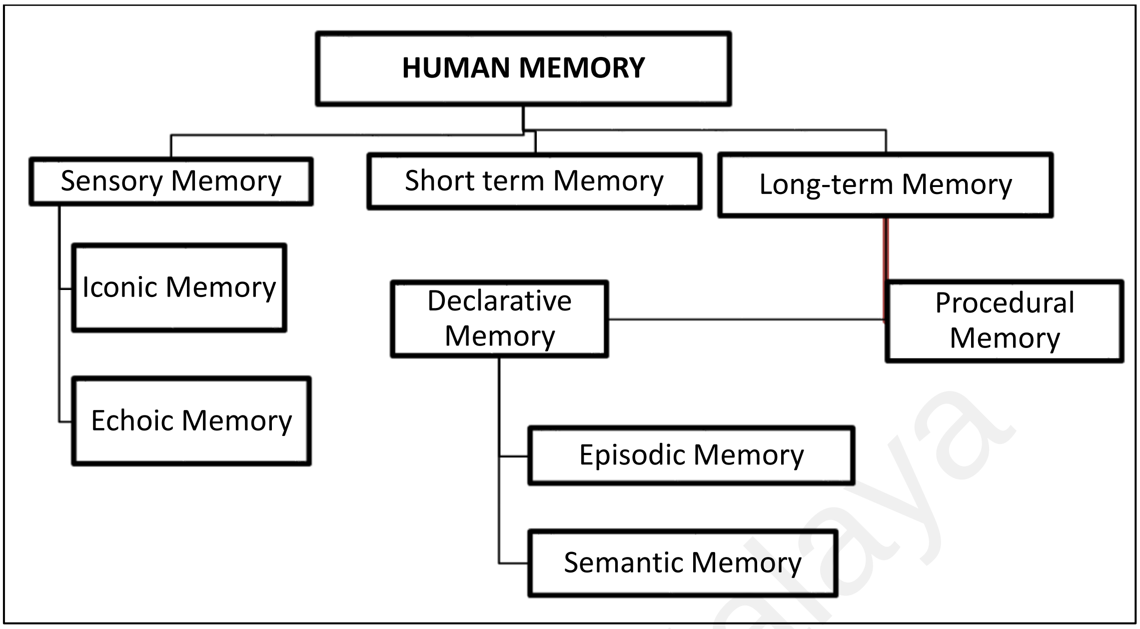

Figure 2.1: Types of Memory (Plotnik & Kouyoumdjian, 2011)

2.4.4.1 Sensory Memory (STSS)

This memory is like a video recorder that automatically record and hold sensory information for a very brief time (from an instant to a few seconds for an individual to decide whether to pay attention or just ignore it. It acts as a buffer for the senses. Scientists have identified two types of sensory memory – iconic and echoic memories.

According to Kalat (2011) iconic memory hold visual information for a very brief period of time but as soon as you stop paying attention to it, then it disappears while echoic memory holds auditory information for one to two seconds. Once the information is given attention, it is passed from here to the short-term memory.

In addition, the sensory memory serves the following functions:

i) It serves as a stimuli filter so that one is not overwhelmed by an influx of sensory stimuli bombarding from outside.

ii) It serves as a buffer to give a person time to decide – accept or reject the stimuli.

iii) Finally it serves to provide stability, playback, and recognition.

(Plotnik & Kouyoumdjian, 2011)

Cognitive psychologists believe in two major approaches to facilitate the input of information into Short Term Memory (STM). Firstly if the information has an interesting feature then the brain will pay more attention to this stimulus. Secondly, a person is more likely to pay attention if the stimulus provokes a previously learned pattern.

2.4.4.2 Short Term Memory (STM)

Short-term memory is also termed working memory and is associated with the thoughts at any given moment in time. In Freudian terms, this is a conscious memory. It is formed when one focuses on an external input, internal thinking patterns, or both.

There are two major strategies for keeping information in STM i.e. organization and repetition. IPT psychologists believed that there are four major types of organization namely: Component (part/whole)--classification by category or concept (e.g., the components of the teaching/learning model like concepts, facts, ideas, classification, taxonomy, concept map, mind map and other graphical illustrations); Sequential – time sequencing; cause/effect; processes (e.g., making a cake, writing a report, constructing a flowchart, doing mind mapping…); Relevance -- central idea or concepts (e.g., basic principles in teaching and learning, strategies for preparation of examination); Transitional (connective) -- connecting words or phrases used to show change across time (e.g., stages in Piaget's or Erikson's stages of socio-emotional development; Stage Theory of Memory, Maslow’s Theory). Sousa (2008) postulates that short-term memory can process a limited number of chunks at any one time. This number is obviously dependent on the age and ability of the person.

2.4.4.3 Difference between short-term memory and working memory

Some cognitive psychologists use these two terms interchangeably. However, short-term memory is distinct from working memory (Kalat, 2011). Working memory refers to structures and processes used for temporarily storing and manipulating information. The most prominent distinction between working memory and STM is that information stored in working memory does not have to be new and it does not have to be on the way to the long-term memory.

Working memory has been hypothesized to contain two components – a phonological loop and a visuo-spatial sketchpad. The loop stores and rehearses speech information and the sketchpad temporarily keeps and retrieves visual and spatial information.

Brain researchers like Sousa (2008), presented alternative views about memory theory in particular short-term memory. He sees short-term memory as comprising of two components – immediate memory and working memory. Immediate memory functions subconsciously or consciously holding data up to only 30 seconds while working memory involves conscious processing working on a limited number of chunks of information at any one time.

2.4.4.4 Long-term memory (LTM)

Long-term memory on the other hand, contains a seemingly unlimited capacity for storing an indefinite amount of information. It is where established relationships among the elements of information are stored. According to the dual-store memory theory by Atkinson and Shiffrin (1968), information can be stored indefinitely in the long-term memory. LTM is crucial for functioning of cognition.

The process of storing information here can be divided into three stages – encoding, storage and retrieval. It has been found that the longer an item is able to stay in STM through rehearsing, the stronger the associations of items and thus allow them to stay longer in LTM. The transfer of information from STM to LTM is known as consolidation.

2.4.4.5 Process of storing information in LTM

The self-explaining Figure 2.2 illustrates the process by which new information is being encoded, rehearsed and retrieved using the Information Processing Model by Atkinson and Shiffrin (1968

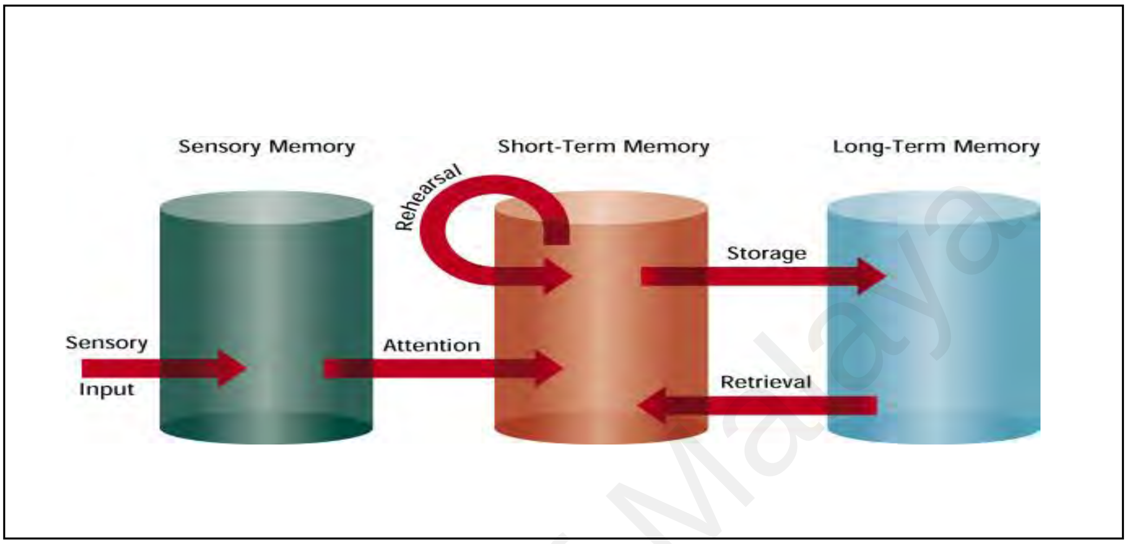

Figure 2.2: The Information Processing model **(**Atkinson and Shiffrin, 1968)

2.4.5 Recall of Information

How does one retrieve vital information from the Long Term Memory? Information processing theory informs that there are a few ways to help in this respect. The three major techniques are i) Free recall, ii) Cued recall, iii) Serial recall

Psychologists like Atkinson and Shiffrin, (1968) and Anderson (1977) have made extensive research in serial recall and these efforts have yielded several general rules:

-More recent experiences are more easily remembered in order;

-Recall of events decreases as the list of objects or sequence increases;

-A person is more likely to remember a list of recently acquired items correctly but maybe in a different order

-When an object is remembered wrongly, there is a tendency for the brain to react by providing memory of a different object which surprisingly resembles the original object in some way.

2.4.6 Mental Representations

According to Anderson (1977), representations stand for something - concrete or abstract. Physical representations stand for objects of which one can perceive with one's five senses while mental representations are totally abstract and only exist in the mind. Mental representations or cognitive representations are theoretical constructs of cognitive scientists in trying to explain mental processes and their manifestations in the form of behaviors. The study of mental representations involves ideas like concept, proposition, schema, script, mental model, image and cognitive map.

a. Concept

Plotnik & Kouyoumdjian (2011) defined a concept as a method of grouping together objects, events or people based on some common features, traits or characteristics.

b. Proposition

A proposition is the smallest unit of knowledge that can stand as an assertation. It is either true or false.

c. Schema

Schemas are knowledge structures about categories of objects, events and people. These cognitive representations can be conceived as a set of related propositions just as concepts can be conceived as a set of related words. Schemas organize related concepts and integrate past events.

More details about schema and the Schema Theory will be discussed later.

d. Mental models

A lot of the times one depends on mental models to transfer learning from one situation to another. Let us take for example playing board games like chess, checker, monopoly or scrabble. When one learns the rules and principles guiding the game like chess one would have built a mental model of this game. When one wants to learn to play

Chinese chess a person recalls the mental model of playing chess and consequently learning to play Chinese chess is much easier and efficient.

e. Mental images

When people daydream or visualize an object in their mind, they are invoking mental images.

Re-enacting these imageries are voluntary and conscious acts. According to Pinker (1999), he claimed that the experiences are stored as mental images that can be compared, contrasted and synthesized to form completely new images. These new images enable a person to form theories or hypotheses. This is how complex cognitive processes occur.

In addition, these images can be expressed in the form of auditory, olfactory and visual images. One form of visual mental images is known as cognitive maps.

f. Cognitive maps

A visual mental model is called cognitive map and it serves to provide information about relative locations and attributes of phenomenon related to the spatial environment.

This mental mapping schema assist in the construction and gathering of spatial knowledge, reduce cognitive load when visualizing images, improve the recall ability and learning.

Thinking and mental processes involve manipulations of mental representations. Varying level of complexities of these processes begin with categorization, attention, mental imagery to highly complex cognitive processes like reasoning and problem solving (Anderson, 1982, 1996).

Problem-solving and reasoning are skills that one develops so that one can act independently as adults. Adults must acquire abilities to source for information, analyze it, and then make reasonable decisions in a rich data-driven environment.

2.4.7 Schema Theory

The schema theory was one of the leading learning theories about thinking and human cognition. It 1932, Bartlett introduced this theory and Richard Anderson further developed it in the ‘70s (Anderson, 1970). A paper by Axelrod (1973) was clearly one of the leading papers to expound on the use of this definitive theory though sometimes been considered abstract by modern psychologists. Axelrod defined the schema as a ‘pre-existing assumption about the way the world is organized. Any new information will attempt to fit into the pre-existing schema but if it cannot then reconstructive cognitive measures are taken to balance the new situation as what Bartlett would called active reconstructive process rather than a passive reproductive one. In addition, Rumelhart believes that: '. . . schemata truly are the building blocks of cognition. They are the fundamental elements upon which all information processing depends. Schemata are employed in the process of interpreting sensory data (both linguistic and nonlinguistic) in retrieving information from memory, in organizing actions, in determining goals and subgoals, in guiding the flow of processing in the system.' (Rumelhart, 1980, pp. 33-34)

According to schema theory how information is processed, and the way it acts in specific settings, are determined to a significant extent by relevant previous knowledge stored in the memory. Such knowledge is said to be organized in the form of schemas – cognitive structures that provide a framework for organizing information about the world, events, people and actions

According to Eysenck and Keane (2015), this theory, schemas function to:

-organize information in the memory

-activate other schemas, often automatically, to increase information-processing efficiency

-influence social perception and behaviour, often when automatically activated

-lead to distortions and mistakes when the wrong schemas is activated

The schema is activated either through 'top-down' i.e. from the whole to the part or "bottom-up" i.e. from the parts to the whole. For example, if on seeing the word "car", one thinks of the parts, e.g. bumper, dashboard, boot, etc., that is "top-down" or "conceptually driven" whereas if one thinks of a collection of words like "swallow, eagle, swift, sparrow, kingfisher" it will produce the concept of 'birds' i.e. 'bottom-up' schema. (Pappas, 2014; Eysenck & Keane, 2015; Fischbein, 1999; Fischbein and Grossman, 1997).

Schema theorists like Fischbein and Grossman (1997) and, Eysenck and Keane (2015) differentiate the schema into various categories namely:

1. Social schema

Social schema is generated by an event (e.g. meeting up with friends in a restaurant).

2. Ideological schema

The ideological schema comes about when a person experiences situations that are generated by differing ideas, attitudes or opinions on issues of the day.

3. Formal schema

The formal schema is related to the stylistic structure of a given text.

4. Linguistic schema

The linguistic schema is the knowledge structure for a person to understand how words are organised and 'stitched' together in a sentence that is understandable either in spoken or written form.

5. Content schema

The content schema refers to knowledge representations about the content of a text. In conclusion, cognitive psychologists are of the view that the schema has four important characteristics:

i. A person can memorize and use a schema automatically.

ii. Once a schema is developed, it tends to be stable over a long period.

iii. Human uses schemata to organize, recall, and encode large amount of important information.

iv. Schemata are accumulated over time and through different experiences

In summary, Schema Theory shows its strengths in explaining how the brain works in terms of explanations to complex cognitive processes and acquisition of experiences, knowledge and memory. As Crane and Hannibal (2009) said, “The theory is useful for understanding how people categorize information, interpreted stories, make inferences and make logic among other things” (p. 72). In addition the theory helps educationists understand distorted memory with respect to social cognition and most importantly the mechanisms of stereotyping and prejudice. Darley and Gross (1983) in their research has found that schema theory has proved to be very useful in explaining processes like perception, reconstructive memory, misconceptions, stereotyping and reasoning. Two terms of importance in the current research that are related to misconceptions are: memory distortion and reconstructive memory.

Memory distortion is about the difference between what is reported and what actually occurred. Memory is the storage of the sum of a person’s experiences. The accuracy of the recording of these experiences depends on the following: i) the level of attention paid to the original event, ii) the time that passes after the original encoding, iii) the match between encoding and retrieval contexts, and iv) the presence of competing and interfering information in memory (Loftus, 2003). In essence, memory does not store the exact duplicates of information. It abstracts the gist and essential components only and fit them into schemas that make sense to the receiver of the information. Reconstructive memory suggests that in the absence of all information, one fills in the gaps to make more sense of what happened. This is why reconstructive memory contains distortions, deletions and omissions (McLeod, 2009; Bartlett, 1932)

However, critics of this theory viewed the theory as too simplistic to be of much value in explaining how complex cognitive processes are developed and used. Some cognitive psychologists were of the opinion that this concept of schema was too vague to be useful and does not explain how schemata are acquired (Cohen, 1993 as cited in McLeod, 2009). The ideas of reconstructive memory and memory distortions are important to the understanding about memory but unfortunately they lack empirical and theoretical strengths to be convincing.

2.4.8 The Practical Aspect of Schema Theory- Putting Theory into Practice

In educational context, teachers are responsible for helping students to develop new schemata and making connections between them. This is to improve their memory. Importantly Eysenck and Keane (2015) found that schema theory helps to improve teaching and learning in area, such as:

i. Mathematical problem solving;

ii. Motor learning;

iii. Reading comprehension.

2.4.9 Schema Theory in Education

Anderson (1977) stated that schemata helped in giving a form of representational structures for complex knowledge and that the construct might influence the acquisition of new knowledge. Schema theory was used to understand and improve the reading process. The schema theory approaches to reading place emphasis on reading that involves both the bottom-up information and the use of top-down knowledge to construct a meaningful schema of the content of the text.

2.4.10 Instructional Implications of Schema Theory

Cognitive psychologists (Eysenck & Keane, 2015; Fischbein, 1999; Fischbein & Grossman, 1997) have suggested that appropriate schemata should be activated just before reading; that teachers should try to provide relevant prior knowledge; and that special attention be given to teaching complex comprehension processes as well as other cognitive processes like reasoning, problem solving and decision making. Schema theory intends to provide a theoretical and empirical background for the teaching and learning process that some experienced teachers have been doing all this while.

From the different definitions of a schema above, one can gather some conclusions about how schema should be represented to be able to turn this abstract and complex term into something concrete that can be studied and taught in ways that is understandable.

In the words of Fischbein, (1999) he interprets a schema as: a program which enables the individual to: a) record, process, control and mentally integrate information, and b) to react meaningfully and efficiently to the environmental stimuli. He sees it as a sort of computer program that has been written in an established procedure that ends with a definite purpose. In this sense, if one can write a computer program to solve a problem, the brain could be similarly using a 'brainware' that helps it solves problems and make informed decisions with good judgement. This brainware is the schema.

2.4.11 Impact of Schema Theory on Education

Schema theory provided educators (Pappas, 2014; Eysenck & Keane, 2015) with an alternative approach to think and deliver representations of various forms of complicated ideas/concepts and knowledge. It has placed importance on the role prior knowledge in acquiring new knowledge. The impact of this theory is immerse in terms of trying to understand the complex processes like prior knowledge, memory (e.g. reconstructive memory and memory distortions), reasoning, problem solving or decision making that are hypothesized to occur through the stages of the Atkinson and Shiffrin model. The schema theory in this respect represents an approach for educationists to view and interpret abstract 'brainware' by comparing its working to a computer software. This in turn, helps the educationists to breakdown highly complex cognitive processes into palatable units for the purpose of understanding how the 'brainware' works. The idea of brainware first mooted by Dennett (1998) in his discussion about the theory of Connectivism, Artificial Intelligence (AI) and the concept of parallel processing "…what is more important is that at a more abstract level the systems and elements—whether or not they resemble any known brainware—are of recognizable biological types. The most obvious and familiar abstract feature shared by most of these models is a high degree of parallel processing..." (p. 226).

2.5 Student Achievement in Statistics Classes

It is a well-known fact that many students find it difficult to grasp statistical concepts and as anticipated acquire misconceptions resulting in statistical errors that compounded their difficulties in understanding more complex concepts and processes (Carmona, 2004; Gal, Ginsburg & Schau, 1997; Onwuegbuzie & Seaman, 1995). The cumulative effects from these problems can be seen in their low achievements in the statistics courses as well as low self-esteem, attitude towards statistics, motivation and confidence level (Dempster & McCorry, 2009; Nasser, 1999; Gal, Ginsburg & Schau, 1997. The next section looks at students' achievement in statistics classes.

2.5.1 Achievement of primary school students in content areas and cognitive domains from TIMSS studies

A comparison of the achievement of general mathematical and cognitive skills of primary and secondary school students from different countries can give an indication of students' achievement in the development of good mathematical or statistical understanding and reasoning. The Trends in International Mathematics and Science Study (TIMSS) is a joint international effort to study the academic competencies of students from participating countries. It seeks to 'measure over time the mathematics and science knowledge and skills' (Mullis et al., 2000) of fourth (Primary 4) and eighth-graders

(Form 2). The scaling procedure starts with the raw score of an individual. It is recalibrated through an estimation process and standardized to a mean of 500 and standard deviation of 100. Table 2.2 gives is an example of the achievement rubric to measure and compare statistics achievement between students and countries.

Table 2.2: Achievement Rubric for TIMSS studies (Mullis et al., 2008)

Advanced (625 cut point)	Students can organize and draw conclusions from information, make generalizations, and solve non-routine problems. Students can derive and use data from several sources to solve multistep problems.
High (550 cut point)	Students can apply their understanding and knowledge in a variety of relatively complex situations. They can interpret data in a variety of graphs and table and solve simple problems involving probability.
Intermediate (475 cut point)	Students can apply basic mathematical knowledge in straightforward situations. They can read and interpret graphs and tables. They recognize basic notions of likelihood
Low (400 cut point)	Students have some knowledge of whole numbers and decimals, operations, and basic graphs.

Table 2.3: Trend of the average mathematics scores of eighth grade students, by selected country from 1999-2007 (IEA, 1999, 2003, 2009; Mullis et al., 2000, 2008)

Country	1999	2003	2007
Singapore	604	605	593
Malaysia	519	508	474
United States	502	504	508
Australia	525	505	496
Russian Federation	526	508	512
South Africa	275	264	-
International Median	487	466	463

Based on the International benchmarks for Mathematics (Table 2.2 & Table 2.3), Singapore is emplaced in the 'High' band implying that an average student in Singapore is able to apply understanding and knowledge to a range of relatively difficult mathematics situations. Malaysia is placed in the ‘Intermediate’ band together with the United States, Australia and Russian Federation. It means that ‘an average student can apply basic mathematical knowledge in straightforward situations’. This level of achievement is sadly insufficient to produce thinking and reasoning students in the near future. The reasoning skill achievement of Malaysian respondents will be discussed later.

TIMSS also provides an overall mathematics scale score for the content and cognitive domain at each grade level. The cognitive domains are classified under ‘Knowing’, ‘Applying’ and ‘Reasoning’. Knowing and applying domains basically parallel Bloom’s Cognitive Objective Taxonomy of “Knowledge, Comprehension and Application”. Reasoning goes beyond the cognitive processes involved ‘in solving routine problems to include unfamiliar situations, complex contexts, and multistep problems’. An analysis of each country’s achievement for 2007 according to content and cognitive domains is shown in Table 2.4. The content domains comprise Number, Algebra, Geometry and Data and chance while the cognitive domains consisted of Knowing, Applying and Reasoning. The content domain of ‘Data and chance’ is compared to the other three areas of mathematics represents the main focus here. Singaporean students did well with the score of 574 for Data and chance section. United States with an average score of 531 and Australia with 525 did comparatively well as their students showed a better mastery of statistics and probability relative to the other content areas. As for the cognitive domains, ‘reasoning’ being a much more difficult skill to acquire was generally lower than that of the ‘knowing’ and ‘applying’ domains in all the countries used for comparison.

The TIMSS studies show that there is much to do about improving students’ reasoning competency.

2.5.2 Correlation analysis between content areas and cognitive domains in three TIMSS studies.

A correlation matrix analysis was generated from secondary data collected from three TIMSS studies (Mullis et al., 2000, 2008, 2012). The aggregated scores were abstracted from four mathematics content areas (Algebra, Numbers, Geometry, Data and Chance) and three cognitive domains (Knowing, Applying and Reasoning) for all the countries who took part in the three consecutive TIMSS studies. Table 2.5 and Table 2.6 indicate that all the math content areas were strongly correlated with each of the cognitive domains providing evidence of the strong relationships between mathematical knowledge and cognitive skills for both the fourth and eighth grades.

Table 2.5 and 2.6 indicate very high correlation indices among all the mathematical content areas tested in the TIMSS studies. This can be taken to imply that good students perform well in all areas while weak students do not do well in any of the areas of mathematics tested. It is thus highly likely that prior mathematical knowledge is a highly connected network of declarative and procedural knowledge comprising of the many fields of mathematics. Ignoring a particular content domain may not bode well in building a good mathematical foundation in the student's later mathematical development.

On closer examination, over the three studies reasoning domain showed lower correlation across all the mathematics content areas as compared with 'knowing' and 'applying' domains implying reasoning domain to be a much more complex domain to acquire.

In conclusion, what is alarming in the recent 2011 TIMSS report is the overall achievement of Malaysia's Eighth Grader in mathematics. There was a significant drop of 34 points from 474 (Year 2007 aggregated score) to 440 (Year 2013 aggregated score) while the closest neighbour Singapore recorded an increase of 18 points from 593 in 2007 to 611 in 2013. Furthermore there is a drop in the aggregated score for the Data Analysis domain. This slide in achievement understandably will have some unwelcome effect on statistical achievements of students in years to come. The slide in achievement among Malaysian students may be arrested by taking steps to improve the teaching and learning of statistics and placing greater emphasis to statistical thinking and reasoning in any curricular revision.

2.6 Statistical Reasoning

2.6.1 What is reasoning?

Reasoning refers to a set of cognitive processes that transform information so that a person can come to a conclusion (Galotti, 2008). Reasoning covers either thinking that uses a well-defined system of logic and/or thinking on a small set of very well-defined tasks. Reasoning involves drawing conclusions based on some given information and in accordance with certain boundary conditions specified by the tasks. Discussion of reasoning cannot exclude other related higher order thinking such as judgment and decision making. A discussion about reasoning from the psychologist point of view is insufficient and incomplete for an understanding of the wide ramifications of the effect of reasoning on human functioning especially in the context of learning where the educational perspective must be sought. Educational perspective deals with issues of practice while psychological perspective deals with issues of theory. Unfortunately the psychological and educational perspectives are not often brought together so that the first one can inform the other (Anderson & Lebiere, 1998). The next section will discuss these perspectives.

Table 2.4: Scores for Mathematics Content and Cognitive Domain of Eighth Grade Students, by Country in 2007 (Mullis et al., 2008; IEA, 2009)

		Content domain								Cognitive domain					
		Number		Algebra		Geometry		Data and chance		Knowing		Applying		Reasoning	
Country	*N*	Average score*	*SD*	Average score*	*SD*	Average score*	*SD*	Average score*	*SD*	Average score*	*SD*	Average score*	*SD*	Average score	*SD*
Singapore	4599	597	3.5	579	3.7	578	3.4	574	3.9	593	3.6	581	3.4	579	4.1
Malaysia	4466	491	5.1	454	4.3	477	5.6	469	4.1	478	4.9	477	4.8	468	3.8
United States	7377	510	2.7	501	2.7	480	2.5	531	2.8	503	2.9	514	2.6	505	2.4
Australia	4069	503	3.7	471	3.7	487	3.6	525	3.2	500	3.4	487	3.3	502	3.3
Russian Federation	4472	507	3.8	518	4.5	510	4.1	487	3.8	510	3.7	521	3.9	497	3.6
#Botswana	4208	366	2.9	394	2.2	325	3.2	384	2.6	351	2.6	376*	2.1	—	†

* TIMSS Scale Average is 500
— Not available.
† Not applicable.
s.e. Standard error.
\# Botswana was chosen to replace South Africa as it was not listed in the 2007 report.

Table 2.5: Grade 8 Math versus Cognitive Domains from TIMSS 2003 to 2011 (IEA, 1999, 2003, 2009; Mullis et al., 2000, 2008, 2012)

		number	**algebra**	**geometry**	**data display**	**knowing**	**applying**	**reasoning**
	Pearson Correlation	1	$.935^{**}$	$.955^{**}$	$.954^{**}$	$.982^{**}$	$.991^{**}$	$.967^{**}$
Number	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000
	N	56	56	56	56	56	56	49
	Pearson Correlation	$.935^{**}$	1	$.930^{**}$	$.872^{**}$	$.978^{**}$	$.945^{**}$	$.930^{**}$
Algebra	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000
	N	56	56	56	56	56	56	49
	Pearson Correlation	$.955^{**}$	$.930^{**}$	1	$.892^{**}$	$.958^{**}$	$.980^{**}$	$.954^{**}$
Geometry	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000
	N	56	56	56	56	56	56	49
	Pearson Correlation	$.954^{**}$	$.872^{**}$	$.892^{**}$	1	$.929^{**}$	$.948^{**}$	$.957^{**}$
data display	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000
	N	56	56	56	56	56	56	49
	Pearson Correlation	$.982^{**}$	$.978^{**}$	$.958^{**}$	$.929^{**}$	1	$.981^{**}$	$.956^{**}$
Knowing	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000
	N	56	56	56	56	56	56	49
	Pearson Correlation	$.991^{**}$	$.945^{**}$	$.980^{**}$	$.948^{**}$	$.981^{**}$	1	$.982^{**}$
applying	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000
	N	56	56	56	56	56	56	49
	Pearson Correlation	$.967^{**}$	$.930^{**}$	$.954^{**}$	$.957^{**}$	$.956^{**}$	$.982^{**}$	1
reasoning	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	
	N	49	49	49	49	49	49	49

**. Correlation is significant at the 0.01 level (2-tailed).

Table 2.6: Grade 4 Math versus Cognitive Domains from 2003 to 2011 (IEA, 1999, 2003, 2009; Mullis et al., 2000, 2008, 2012)

		number	**geometric shape**	**data display**	**knowing**	**applying**	**reasoning**
number	Pearson Correlation	1	.961**	.935**	.994**	.983**	.796**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	42	42	42	42	42	39
geometric shape	Pearson Correlation	.961**	1	.977**	.970**	.991**	.848**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	42	42	42	42	42	39
data display	Pearson Correlation	.935**	.977**	1	.944**	.978**	.872**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	42	42	42	42	42	39
knowing	Pearson Correlation	.994**	.970**	.944**	1	.982**	.804**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	42	42	42	42	42	39
applying	Pearson Correlation	.983**	.991**	.978**	.982**	1	.853**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	42	42	42	42	42	39
reasoning	Pearson Correlation	.796**	.848**	.872**	.804**	.853**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	39	39	39	39	39	39

**. Correlation is significant at the 0.01 level (2-tailed).

2.6.2 Psychological perspective on Reasoning

Chapter 1 briefly presented the definition and concept of reasoning from a psychologist perspective. Hardman and Macchi (2003) explained that reasoning, judgement and decision making are closely related and overlapping as talking about one will invoke the others. In other words, psychologists agreed that when individuals reason about something, invariably they will need to make a judgment call as well as make some kind of decision after considering all the options opened to them. In some particular circumstance, normative theories could predict what rational thinkers would do when they reason, judge or make a decision. Psychologists were puzzled why many a time thinkers are not really rational. This irrationality has given rise to errors in human cognition, human biasness, dubious conceptual understanding and consequently misconceptions (Evans, 2007; Kahneman et al. 1982; Simon, 1956). Many theories have been put forth to explain this discrepancy. Simon opined that this is due to human's bounded rationality.

Evans and Over (1996) and Stanovich (1999) entertained the idea of dual processing in thinking and reasoning. According to the researchers, there are two types of thinking - implicit or explicit that involves either intuitive processing or deliberate processing. Implicit thinking or System 1 thinking provides automatic input to the brain to act pragmatically utilizing knowledge and beliefs residing in the long-term memory of which Stanovich called it fundamental computational bias which is the basis to resort to heuristics to reason or solve problems. Heuristics work sometimes but most of the time causes biasness and errors in human cognition. The other type of thinking - explicit thinking or System 2 thinking is seen to be related to language and reflective skills. This skills provide the basis for reasoning (Evans, 2008). System 2 operation requires large space in the limited working memory where information is processed linearly. It has been established that effective functioning of this system is related to the IQ. However, due to the inherent 'inefficiency' of this site to process large amount of information, there is a tendency that most of us will fall back to System 1 regularly. Generally psychologists tend to agree that reasoning involves deliberate processes (consisting of conscious, controlled application of rules and computations) and the intuitive processes that functions automatically and without conscious control (Evans, 2007; Glöckner & Witteman, 2010).

From the eyes of a psychologist, reasoning involves a set of cognitive processes used to derive an inference or conclusion using the information available. It helps to generate new knowledge and organize existing knowledge so that this knowledge is more usable for future mental work (Mercier & Sperber, 2011). Thus reasoning is seen as a means to improve knowledge and helps us make better decisions. Unfortunately ample evidence has shown that it is not what it is made out to be (Mercier & Sperber, 2011). Brewer and Samarapungavan (1991) stated that there is seldom an ideal reasoner. In reality all of us are constrained by the 'bounded rationality' due to factors like limited working memory and the cognitive goals where one often look for an acceptable solution rather than a 'best' solution. In recent years, another revolutionary theory has emerged to explain the phenomenon of why some people are such bad reasoners sometimes and the link between these phenomena with the confirmation bias. The Argumentative Theory of Reasoning put forth by Mercier and Sperber (2011) hypothesizes that human reasoning was designed to help us win arguments and not to seek the truth. The researchers argued that poor achievement is the result of the lack of an argumentative context. The researchers opined that people basically reason to find rationale and support for their views and the truth elements in those views are secondary. The researchers found some support for their views from well-known psychologists and educators like Gerd Gigenrezer and Steven Pinker. Works by Kersten, Mamassian and Yuille (2004) and Wolpert and Kawato (1998) were quoted as the basis for some of the arguments put forth especially in the area of inferences, prior knowledge, conceptual thinking and perceptions. The researchers use their theory to explain the notorious confirmation bias as an example. The researchers reiterate that this bias is not a flaw of reasoning but rather it is a feature of human reasoning where winning an argument takes precedence over getting at the truth!

2.6.3 Educational perspective on reasoning

Reasoning being a higher order thinking skill is required for many of the thought processes in learning. Theories served up different terms and definitions for reasoning - informal reasoning versus formal reasoning, implicit vs. explicit reasoning, deductive vs. inductive reasoning, spatial reasoning, geometrical reasoning, proportional reasoning, argumentative reasoning, abductive reasoning, analogical reasoning and many more. The abundance of different definitions of reasoning clouds psychologists' ability to clearly defined what is meant by reasoning or it may well be reasoning is too complex to define unambiguously. The problem is analogous to the different types of intelligences introduced by Howard Gardner. Humans need different reasoning for different cognitive processes. The many different forms reasoning take could very well be due to the humans' limited understanding of this thought process and so one seeks to pigeon hole this highly complex and dynamic construct into defined compartments which is impossible. Educationists had reiterated that reasoning in its various forms is partially dependent on innate intelligence. This implies that reasoning can be taught and learned; it can be practiced and improved (Schwartz, 2001).

The Argumentative theory seeks confirmation of its applicability in the field of education through the confirmation bias problem. Mercier and Sperber (2011) found that novices tend to fall back on heuristics more often than professionals. In the earlier chapter under the section 'Errors in human cognition', heuristics or mental shortcuts had been shown to give rise to biases such as representative biases, availability biases or confirmation biases. Confirmation biases come about due to the tendency to find support for the hypothesis without considering other possibilities. The theory says that humans reason through argument and they do it best in groups. They opined that using collaborative learning to understand difficult and abstract topics would be relevant for reasoning to be practised where deliberation, discussion, sharing and criticizing each other's point of view have a 'natural habitat' to occur.

From the numerous statistics education studies on reasoning, findings have consistently shown that students take time to develop statistical ideas and concepts. Repeated practicing in examining, interpreting, discussing and comparing are important processes to reinforce concepts, procedures and reasoning. It is important to provide opportunity for students to build their own intuitive ideas as inventing informal language for concepts or ideas that they have not encountered formally (Garfield & Ben-Zvi, 2008, Bakker & Gravemeijer, 2004, Pfannkuch, 2005, delMas, Garfield & Ooms, 2005). The studies also indicated that the sequencing of ideas to build one on top of the other in a hierarchical form. The most important message according to statistics educators is that statistics teachers need to be aware of the difficulties students have with developing statistical ideas and concepts (Gal & Garfield, 1997, Gal, 2004). Since researchers have seen a variety of approaches to the study of human reasoning and the varied interpretations by psychologists and educators in different fields of study, in the next section, we will be looking at reasoning in statistics, its relationships to statistical literacy and thinking and how statistics educators assess statistical achievement.

2.6.4 What is statistical reasoning?

Statistical reasoning is defined as the way students reason with statistical ideas and make sense of statistical information (Garfield, 2003). Statistics reasoning is based on the knowledge and understanding of concepts such as data, distribution, graphical representations, measures of centrality and variation, association, randomness, sampling and inference and prediction. Research presently are focused on what really constitute the term 'statistical reasoning' rather than referring to such general constructs like the psychologists' version of reasoning or mathematical reasoning for that matter. The direction and trend are towards understanding reasoning and how it impacts the learning of statistics (del Mas, 2002; Reading, 2002).

In the words of Garfield (2002) who is at the forefront of research into reasoning and learning in statistics, agreed to the many different ways it is defined can cause problems but "…it appears to be universally accepted as a goal for students in statistics classes." that makes it necessary to teach the students. Undoubtedly it has a complex relationship with other cognitive processes like prior knowledge and errors in cognition. There is a need to understand how prior knowledge or preconceptions are related to reasoning especially prior reasoning skills that students bring along to class. If preconceptions correspond to true knowledge then learning can proceed smoothly. If preconceptions are misconceptions, however, then teaching for conceptual understanding is retarded depending on the seriousness and the number of misconceptions. Brandsford, Brown and Cocking (2000) warned of similar consequences when students developed wrong preconceptions. Garfield (2002) called for more research perhaps more classroom-based situations to look at the types of reasoning, the prior knowledge and skills for each type of reasoning to better understand the process of how correct statistical reasoning develops.

2.6.5 Relationships between Statistical Reasoning, Literacy and Thinking

Higher mental processes are necessary for success in studying statistics. Statistics educators agree that three overlapping constructs are crucial to the understanding and application of statistics in very diverse fields in economy, social sciences, applied sciences, mathematical sciences and management. The earlier chapter has discussed briefly these three constructs. Statistical literacy refers to the understanding and the knowledge of terms, concepts, symbols and graphical representations. Statistical reasoning is the way one makes sense of statistical information while statistical thinking is about the why and how of doing statistical investigations. delMas (2002) believed that these three constructs are not distinct but there is some overlap in their cognitive outcomes. He opined that there is a hierarchical structure to the relationships as illustrated in Figure 2.3.

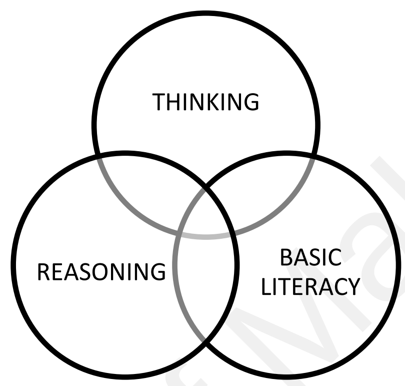

Figure 2.3: The overlapping of the relationships between statistical literacy, reasoning and thinking (delMas, 2004a)

Many statisticians agreed on the importance of acquiring these abilities (Chance & Garfield, 2002; delMas, 2002; Garfield, 2002; Rumsey, 2002; Garfield & Ben-Zvi, 2008) but there is less consensus as to their actual use and operationalization of those constructs (Ben-Zvi & Garfield, 2004a,2004b; delMas, 2004a; Garfield & Ben-Zvi, 2008).

Due to the difficulties in making clear distinctions among these three terms, studies have been mainly focused on one of these higher-order thinking processes i.e. statistical reasoning. This study seeks to investigate the relationships of statistical reasoning to other cognitive factors such as misconceptions and prior mathematical knowledge and statistical achievement.

2.6.6 Statistical reasoning and its assessment

In the earlier section on assessment, educators have recommended a move to more inclusive strategies and approaches that 1) can reflect learning outcomes comprehensively, 2) can measure more effectively the different kinds of conceptual understanding in a statistics class, 3) cater to a large number of students, and 4) easy to administer, economical, time and cost effective. Martin (2013) commented on the multiple facets of statistical reasoning making assessment of the reasoning complicated. Her study used SRA to measure statistical reasoning. She concluded that statistical reasoning improved with experience but achievement is dependent upon both cognitive and non-cognitive abilities.

Many instruments for assessing statistical reasoning, both quantitatively and qualitatively, had been developed according to the purpose of assessment as discussed previously. In terms of assessing the reasoning levels of students in large classes, ease of administering the test, ease of scoring and analyzing, SRA would be a perfect choice. The effectiveness and relative success of this instrument in measuring reasoning skills and misconceptions had spurred many statistics educators to design tests and assessment tools along the same line. Ooms (2005) had developed together with other statisticians an instrument known as the Comprehensive Assessment of Outcomes in a first Statistics course (CAOS) focusing on testing students' ability in conceptual understanding of basic statistics. This test had been extensively tested online to improve its reliability and validity (delMas, Ooms, Garfield & Chance, 2006).

Another instrument is the Quantitative Reasoning Questionnaire (QRQ) developed by Sundre (2003). Sundre considered the SRA as 'a welcome step forward in the design of instructional-friendly assessment tools'. The ability of the SRA to measure reasoning and misconceptions represented a major step in the teaching and learning of statistics as well as its capacity to provide meaningful feedback to both the educators and students. Some of the items in the QRQ closely imitated the SRA items but some were redesigned to overcome some of the inherent weaknesses of the SRA instrument as suggested by Garfield herself i.e. low internal consistency, item format and scoring omitted potentially important information, difficulty in scoring and inability to assess reasoning and misconceptions scales fully. The final version of the QRQ consisted of 43 items measuring 11 quantitative reasoning skills and 15 quantitative misconceptions and skill deficiencies. To score, there are two scoring rubrics for the open-ended items and the scoring for the multiple choice items follow the SRA technique. However this instrument is unpopular as it had too many items, was difficult to score and was too time consuming.

Hirsch and O'Donnell (2001) took up the issues of SRA validity and reliability. In their attempt to improve Garfield's instrument they designed a 16 item multiple-choice test where each item has two parts. This format replicated part of the instrument originally developed by Konold. The Konold format was chosen as the items constructed took advantage of the efficiency of multiple choice test items and at the same time measures the students' rationales behind their choice of answers. In Konold's instrument, the first part asked a question similar to SRA items however the second part of each item was supplemented with different reasoning options that partially reflect a range of possible reasoning skills of the respondents. The choices are scored for reasoning abilities and misconceptions. Results of this study showed that this instrument had higher validity and reliability compared to the SRA and this format provided invaluable diagnostic information concerning students' errors. However this instrument is not as popular as the SRA because of problems in administering a large item set to a large student population and scoring on the two-part items was comparatively difficult as it required scoring rubrics and subjective scoring. Analyzing the data takes a lot of time and effort.

The popularity of the SRA lies in its ability to measure different areas of statistical understanding within a single instrument and could be administered to a large group (Martin, 2013) although the issues of moderate reliability have been raised.

2.6.7 Development of the SRA by Garfield (2003)

The Statistical Reasoning Assessment instrument was developed by Garfield (2003). The content of this 20-item multiple-choice test comprises of statistics and probability problems. Each item has several choices of responses or options that are both correct and incorrect. The correct option taps into the reasoning power of the respondents while the rest of the options measure their misconceptions. Each option is a statement explaining the rationale for the respondents' choice thus tapping into their thinking about the problem asked. The original objective of SRA is to evaluate the curricular content areas and approaches apart from measuring the level of the students' statistical reasoning (Garfield, 2003). The first step in the designing of the instrument was to identify the types of reasoning skills students are expected to develop. The reasoning skills encompass: a) reasoning about data, b) reasoning about representation of data, c) reasoning about statistical measures, d) reasoning about uncertainty, e) reasoning about samples and f) reasoning about association.

In addition, the SRA also measures the incorrect reasoning or misconceptions. They included: a) misconceptions involving averages, b) the outcome orientation, c) good samples have to represent a high percentage of the population. d) 'law of small numbers', e) the representativeness misconception and f) the equiprobability bias (Garfield, 2002). The instrument went through several rounds of refinement using the conventional item analysis approach. As this instrument is a multiple choice test, issues related to the construction of appropriate options to capture reasoning and misconception were resolved before submitting the items to a pilot study.

2.6.8 Validity of the SRA instrument

Content validity of the SRA items was assured by choosing and adjusting items to match selected topics representing sections of the curricular content to be assessed. The items constructed were deemed to be sufficient though not complete to measure the reasoning skills of students who were taking their first course in statistics. Table 2.7 shows the list of topics and distribution of items being examined in three versions for comparison purpose.

Table 2.7 makes a comparison of three studies carried out at different times. It compares the study by Garfield (2003), Zuraida et al., (2012) and the current study (2016) on the topics and distribution of items in each of the different versions of the SRA instrument as it evolved. The items measure different aspects of statistical reasoning such as interpreting probabilities, understanding about central tendency, compute probabilities, understanding the concepts of independence or the importance of large samples and correlation as causation. They are categorized using symbols CC1 – CC7. For this current study, there are only 6 categories of interest due to the fact that the respondents are not taught concepts related to CC7.

As the SRA instrument also measures misconceptions of the respondents, Table 2.8 compares the different categories of misconceptions as proposed by Garfield (2003) namely MC1 – MC9 in the original instrument but in the present study, the categories of interest are limited to MC1-MC5 due to the characteristics of the sample chosen. The misconceptions selected to be studied cover common errors like misconceptions involving averages, outcome orientation, law of small numbers, equiprobability bias and representative bias.

Table 2.7: **Topics and distribution of items for reasoning scales in SRA**

Garfield (2003)	Zuraida et al, (2012)	Current study
CC1 - Correctly interprets probabilities Items 2,3	CC1 – Correctly interprets probabilities Items 2,3	CC1 - Correctly interprets probabilities Items 2, 3
CC2- Understands how to select an appropriate average Items 1,4,17	CC2- Understands how to select an appropriate average Items 1, 4, 12	CC2- Understands how to select an appropriate average Items 1, 4, 12
CC3- Correctly computes probability Items 8,13,18,19,20	CC3- Correctly computes probability Items 5 10, 13, 14, 15	CC3- Correctly computes probability Items 5, 10, 14, 15
CC4-Understands independence Items 9,10,11	CC4-Understands independence Items 6, 7,8	CC4-Understands independence Items 6, 7, 8
CC5- Understands sampling variability Items 14,15	CC5- Understands sampling variability Item 11	CC5- Understands sampling variability Item 11
CC8- Understands the importance of large samples Items 6 ,12	CC6- Understands the importance of large samples Item – 9	CC6 -Understands the importance of large samples Item- 9
CC6 -Correlation implies causation Items 16	CC7 - no item	CC7 – no item
CC7-Interprets two-way tables Items 1,5 – Not investigated/not in syllabus		

The changes in the items can be compared using the SRA in Appendix A1 and Appendix A2

Table 2.7 shows the dimensions and items that were adapted from the original SRA items by Garfield (2003).

Table 2.8: Topics and distribution of items used in the SRA for different versions

Garfield (2003)	Zuraida et al, (2012)	**Current** study
MC1- Misconceptions involving averages Items 1a, 1c, 12a	MC1- Misconceptions involving averages Items 1a, 1c, 12a	MC1- Misconceptions involving averages Items 1a, 1c, 12a
MC2- Outcome orientation Items 2e, 3ab, 8abd, 9c, 10b	MC2- Outcome orientation Items 2e, 3ab, 8abd, 9c, 10b	MC2- Outcome orientation Items 2e, 3ab, 8abd, 9c, 10b
MC3- Good samples have to represent a high percentage of the population– NOT INVESTIGATED	MC7- Good samples have to represent a high percentage of the population– NOT INVESTIGATED	MC7- Good samples have to represent a high percentage of the population– NOT INVESTIGATED
MC4- Law of small numbers Items 9a, 11c	MC3- Law of small numbers Items 9a, 11c	MC3- Law of small numbers Items 9a, 11c
MC5- Representativeness misconception Items 6abd, 7d, 8c	MC4- Representativeness misconception Items 6abd, 7d, 8c	MC4- Representativeness misconception Items 6abd, 7d, 8c
MC7-Equiprobability bias Items 10c, 13a, 14d, 15d	MC5-Equiprobability bias Items 10c, 13a, 14d, 15d	MC5-Equiprobability bias Items 10c, 13a, 14d, 15d
MC8- Groups can only be compared if they have the same size–– NOT INVESTIGATED	MC8- Groups can only be compared if they have the same size– NOT INVESTIGATED	MC8- Groups can only be compared if they have the same size– NOT INVESTIGATED
MC9- Correlation implies causation – NOT INVESTIGATED	MC9- Correlation implies causation– NOT INVESTIGATED	MC9- Correlation implies causation– NOT INVESTIGATED

2.6.9 Weaknesses of the SRA instrument

Many studies have attested to the problem of validity and reliability of this instrument. Garfield (2003) reiterated that there is still much work to be done to increase the validity and reliability indices of the SRA. among which are: low internal consistency, item format and scoring omitted potentially important information, difficulty in scoring and inability to assess reasoning and misconceptions scales fully. Construct was rarely reported in many of the earlier studies.

2.7 Misconceptions in Statistics

In educational research, the term misconception is subjected to a variety of interpretations. On the one hand, 'authors often consider a broad definition of the word, using it to label different concepts such as preconception, misunderstanding, misuse, or misinterpretation interchangeably' (Smith, diSessa & Roschelle, 1993). Misconceptions are sometimes 'seen in a more restrictive way, as misunderstandings generated during instruction, emphasizing a distinction with alternative conceptions resulting from ordinary life and experience' (Guzzetti, Snyder, Glass & Gamas, 1993). A more complete form considers misconceptions as 'any sort of fallacies, misunderstandings, misuses, or misinterpretations of concepts, provided that they result in a documented systematic pattern of error' (Cohen, Smith, Chechile, Burns, & Tsai, 1996). This definition from a psychological perspective is sufficient but Olivier (1989) commented that from an educational perspective, 'misconceptions are crucially important to learning and teaching, because misconceptions form part of a pupil's conceptual structure that will interact with new concepts, and influence new learning, mostly in a negative way, because misconceptions generate errors'. Misconceptions are systematic conceptual errors caused by underlying contrary beliefs and principles that are deeply ingrained in the students' cognitive structures. This will be the interpretation of the term misconception henceforth in this study.

Some of the most common misconceptions are 1) equiprobability bias i.e. the tendency to consider several outcomes of an experiment as equally likely. 2) representativeness misconception i.e. the tendency of students to wrongly think that samples which look similar to the population distribution are more probable than samples which do not.

Newton (2000) sees failure to understand leads to misconception. Much literature has found mounting proof of students' learning problems in statistics and probability. At basic level, students have problems with concepts like average, variance, law of small number, sample representativeness and variability (Gardner & Hudson, 1999; Garfield, 2002; Foo, 2011; Konold, 1989; Lipson, 2002; Schau & Mattern, 1997; Ware & Chastain, 1991).

Misconceptions in probability and statistics have been a popular research pursuit of many statistics educators and psychologists (e.g. Konold, 1989, 1991; Nisbett & Ross, 1980; Shaughnessy, 1981 Tversky & Kahneman, 1971). Shaughnessy (1981) looked at the misconceptions students have with learning probability and how it influenced their understanding in statistical inference in their later years. From his research and experience in teaching students, he found that the misconceptions they had were more psychological in nature than anything else. His hypothesis concurred with other related studies by psychologists like Kahneman and Tversky (1972) and, Cohen et al., (1996). Kahneman and Tversky (1972) claimed that some of the more serious misconceptions arising from the learning of probability among students came from the usage of two simplifying techniques in the face of complicated probability tasks. The techniques were named 'representativeness' and 'availability' strategies. Students' dependence on these faulty strategies, the study cautioned, can lead to even more understanding-related problems in their later encounter with advanced statistics. Common errors that were particularly important to take notice were: 1) insensitivity to prior probability and disregard for population proportions, 2) insensitivity to the effects of sample size on predictive accuracy, 3) unwarranted confidence in a prediction that is based on invalid input data, 4) misconceptions of chance such as the gambler's fallacy and finally 5) misconceptions about the tendency for data to regress to the mean.

Mere exposure to probability concepts does not prevent students from relying on representativeness or availability. The problem goes deeper than what they had suspected. He went on to explain that 'our intuition of probabilistic thinking has been distorted by an overemphasis on deterministic models' like the axioms of geometry or Newton's Law of Gravity. Students found it particularly hard to rationalize and adapt to two seemingly contrary perspectives (i.e., deterministic versus probabilistic thinking). This issue has already been raised by Kahneman and Tversky (1972) and again by Konold (1989). Their studies were concerned with understanding of sampling. Kahneman and Tversky found that their subjects focused on the singular rather than distributional perspective when making judgement under uncertainty. Konold (1989) upheld Shaughnessy's argument that statistically weak students still hold 'certainty' or 'deterministic' view in solving complicated probability problems. Both researchers agreed that it was really difficult to change deep-rooted misconceptions even after repeatedly giving evidence to the contrary. In other related studies (Gigerenzer, 1998, 1993; Hertwig & Gigerenzer, 1999) found that when their respondents were given a set of tasks to answer involving distribution of sample statistics, they showed similar misconceptions. Unfortunately a good number of the students treated the tasks as though they were about individual samples. The students had taken what they called as a 'singular' perspective would directly influence their ability to comprehend and apply the concepts of sampling representativeness and sampling variability.

According to Rubin, Bruce and Tenney (1991) the reasoning behind statistical inference entails the balancing of these two seemingly conflicting concepts. The researchers found that their subjects tend to choose either one of the two ideas in solving different sampling and inference tasks based on their own 'understanding'. Schwartz, Goldman, Vye and Barron (1998) addressed the same difficulty by suggesting that students can be taught to understand and overcome the contradictions as described by Rubin, Bruce and Tenney (1991). Saldanha (2004) commented that "students experienced significant difficulties coordinating and composing multiple objects and actions entailed in a resampling scenario into a coherent and stable scheme of interrelationships that might underlie a powerful conception of sampling distribution..." A good understanding of sampling distribution is the cornerstone to comprehending statistical inference.

It is thus appropriate at this juncture to look at some major misconceptions in NHST in relation to sampling distribution and statistical inference to better understand the structural problems experienced by some students, educators and researchers.

2.7.1 Studies about misconceptions in basic statistics and statistical inference

The following discussion summarizes the root causes and misconceptions of sampling distribution and hypothesis testing from a meta-analysis of 17 different studies that provide empirical evidence of misconceptions. The studies selected for analysis were all published from 1990 to the beginning of 2006. Their analysis covers three major topics namely sampling distributions, hypotheses tests and confidence intervals tracing the misconceptions in these topics to weak understanding of basic statistics.

Briefly, the researchers found weak understanding and persistent confusions in some underlying concepts and relationships (Foo, 2011; Sotos, Vanhoof, Van den Noortgate & Onghena 2007).

Misconception studies in the Asian countries are few. Findings about students' difficulties with learning of statistics and misconceptions are mostly situated in a western context. However a recent study about the misconceptions in statistical inference (Foo, 2011) will be discussed next to provide a background of the status of the learning difficulties and misconceptions with introductory statistics in higher education in Malaysia and Singapore.

2.7.2 A Survey of Malaysian and Singaporean University students' misconceptions concerning statistical inference

A study was conducted in mid-2008 to look at misconceptions among researchers, undergraduates and postgraduates students (Foo, 2011). This study was envisioned in part to answer (Shaughnessy, 1981)'s concern regarding the generalizability of research findings from the West with regard to statistical misconceptions. The author was curious to know if these findings were just artifacts of cultures or the problems do exist in other parts of the world. Misinterpretations and incomplete statistical understanding can be real obstacles to appreciating, reasoning and applying the complex hypothesis testing procedure. Hence this exploratory study was conceived to find out what misconceptions and how widespread they were. This study looked at NHST misconceptions amongst Malaysians and Singaporean respondents (Foo, 2011).

The results from the quantitative analysis found that that 95.5% of the 179 participants surveyed had significant degree of misconceptions. The average misconception score for Malaysian respondents was significantly higher than that of Singapore as can be seen in Table 2.9.

Table 2.9: Average misconception scores for Malaysian and Singaporean Participants

Country	n	Mean	Std. Error	Median
Malaysia	115	65.79	2.32	66.70
Singapore	64	51.30	3.00	50.00

As seen in Table 2.9, while the Singaporean sample performed much better than the Malaysians and in fact, many other countries, they did have problems with NHST just like the others. Mastery of basic statistical concepts is obviously a prerequisite for understanding NHST but apparently insufficient to cope with the intricacies of NHST. In addition, it was also found that high percentages of respondents still harbour differing degree of misconceptions among respondents sampled in USA (Oakes, 1986), Germany (Haller & Krauss, 2002), Malaysia (Foo, 2011) and Singapore (Foo, 2011).

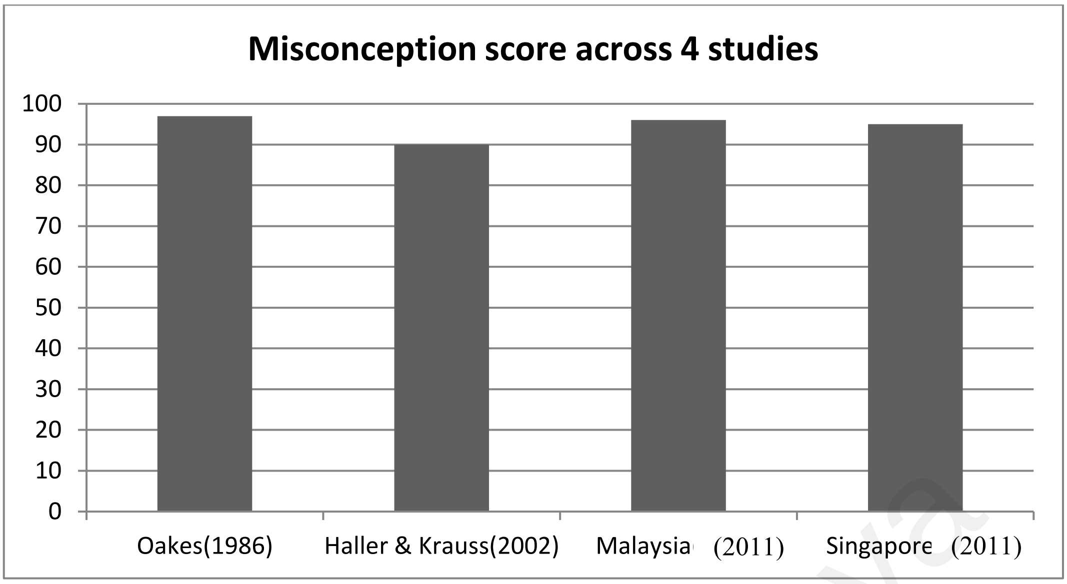

Figure 2.4: Percentages of respondents with misconceptions across 4 studies

Over a span of 20 years beginning with Oakes (1986) experiment to the Malaysian and Singaporean study in 2009, there seemed to be little change in the way people think and reason about statistics. The question boils down to "Is it correct to conclude that teaching of inferential statistics and probability theory represent some of the educational failures and thus are deemed to be 'unteachable'?, a scenario that educators would be hard to imagine.

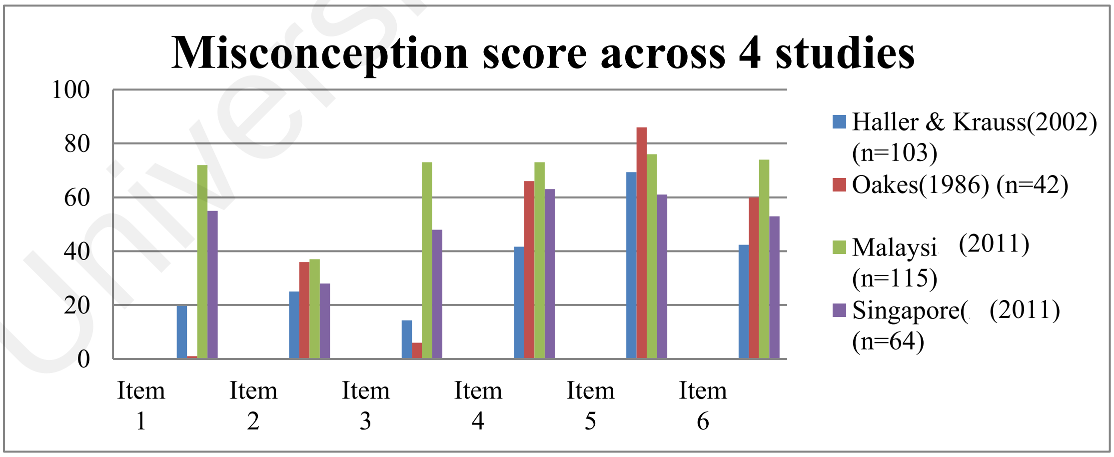

Figure 2.5: Misconception scores across 4 studies - item by item analysis.

Next, we look at various types of misconceptions by analysing item by item in the survey. As seen in Figure 2.5, the difficulty level of each of these selected items is compared across the four studies. Malaysian participants found it especially difficult to detect the falsity of each item except for item 2. Item 5 seems to be the most difficult statement to understand. The conditional logic used in the structure of the sentence and language mastery of the readers had a lot to do with the confusion and probably because of the moderate mastery of statistical knowledge too has compounded the problem. Nevertheless, other studies carried out in the West similarly recorded high incidence of misconceptions among their participants for this particular item. In a way the problem of understanding and its related issues are not unique to the Malaysian context but rather it can be considered a global phenomenon. As had been explained earlier, the train of the reasoning process gets really confusing in this particular item compared to others. In his preface, Sedlmeier (1999) opined that good statistical reasoning was rarely well taught.

Newton (2000) reasons that students' failure to understand is due to 'a failure to construct an adequate, coherent mental representation of the information in a situation', lack of prior knowledge, excessive mental demand of the situation, failure to notice relevant relationships between the new information and prior knowledge, inability to manipulate a mental representation, lack of rules or guidelines to look at relationships and a host of other reasons. He suggested general guidelines that are systemic or holistic in approach. Strategies should stem from building up a strong statistical foundation.

TIMSS studies (Mullis et al., 2000, 2008, 2012) have clearly indicated that many countries do not perform well in the Data and Chance section. Shaughnessy (1981) stated that 'misconceptions students harbored were more psychological in nature than anything else'. This view is shared by other psychologists like Kahneman and Tversky (1972) and Cohen et al., (1996). Kahneman and Tversky (1972) claimed that some of the more serious misconceptions arising from the learning of probability among students came from the usage of two simplifying techniques in the face of complicated probability tasks. The techniques were named 'representativeness' and 'availability' strategies. Due to students' dependence on these faulty strategies, the study cautioned on the possibility of these students facing more understanding-related problems in their later encounter in advanced statistics courses. One good advice on how to avoid this problem is to expose students to different situations where the techniques work and when they do not. Huck (2004) in his book "Reading Statistics and Research" pays serious attention to common misconceptions in each of his chapters. It is rather uncommon to read statistics books that took pain to explain and highlight the difficulties students face as they attempt to understand inferential statistics especially when it comes to difficult concepts. Huck was well aware of the problems that misconceptions will pose to students in later chapters if these errors are not correct in the earlier topics. These discussions are key points that readers can pay particular attention to avoid misuses and misunderstanding stemming from the incorrect interpretations of statistical concepts and relationships.

Much has been said about how and why the students acquire those misconceptions. Evidently nothing much has been done probably due to the controversy that is still very much alive leaving us with little productive time to move on. All is not lost for there are many forces of positive changes from the works of concerned statistics educators and psychologists. This is succinctly put by Gigerenzer (1993) "…it is our duty to inform our students about the many good roads to statistical inference that exist and to teach them how to use informed judgment to decide which one to follow for a particular problem" (p. 335). In looking for a good solution to the problem of overcoming misconceptions and designing a simple but effective assessment tool to identify these misconceptions should represent the main thrust of statistics researchers in the years to come.

2.8 Prior Knowledge and Information Processing Model (IPM)

Prior knowledge is located in the memory. Memory in IPM consists of three components– sensory memory, short-term memory and long-term memory (see Fig 2.6).

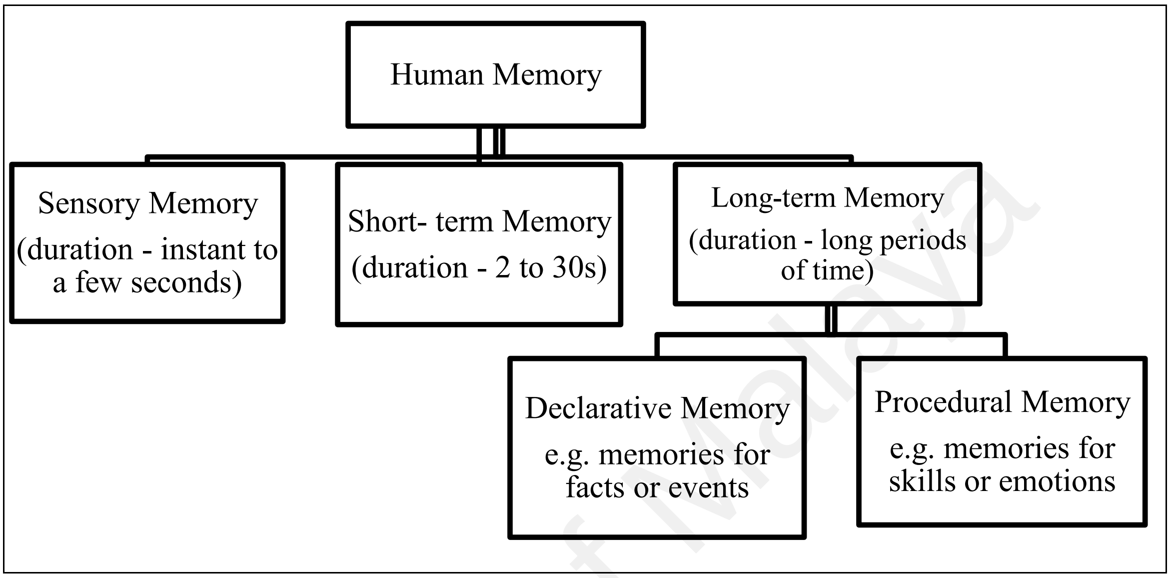

Figure 2.6: Types of Memory (Plotnik et.al, 2011)

2.8.1 Sensory memory

Plotnik & Kouyoumdjian (2011) liken sensory memory to a video recorder that automatically record and hold sensory information for a very brief time (from an instant to a few seconds) for one to decide whether one wants to pay attention or just ignore it. It acts as a buffer for the senses. Scientists have identified two types of sensory memory – iconic and echoic memories. Iconic memory holds visual information for a very brief period of time but as soon as one stops paying attention to it, then it disappears while echoic memory holds auditory information for one to two seconds. Once the information is given attention, it is passed from here to the short-term memory. In addition, the sensory memory serves the following functions:

1) It serves as a stimuli filter so that humans are not overwhelmed by an influx of sensory stimuli bombarding from outside.

2) It serves as a buffer to give us time to decide – accept or reject the stimuli

3) It serves to provide stability, playback, and recognition

(Plotnik & Kouyoumdjian, 2011)

2.8.2 Short-term memory (STM)

Sometimes called active or primary memory, the short-term memory is the ability of this storage to hold a small amount of information in an active and easily retrievable form for just a short period. This type of memory is characterized by its duration and capacity. According to Plotnik & Kouyoumdjian (2011), the duration has been quoted to be between 2 to 30 seconds. Afterwards the information decays over time. However researchers had shown that one could keep the information there longer through the technique of maintenance rehearsal. It refers to the intentional rehearsal or repetition of the elements of information one wants to commit to the short term memory. It has been reported that with rehearsal information can be kept for another 15-20 seconds.

Chunking can also help in storing more information within the capacity of the primary memory storage. Chunking is the process of grouping individual elements into meaningful patterns or clusters.

2.8.3 Difference between short-term memory and working memory

Short-term memory is distinct from working memory (Kalat, 2011). Working memory refers to structures and processes used for temporarily storing and manipulating information. One significant difference is that working memory is the information that a person is using does not have to be new and it does not have to be on the way to the long-term memory (Kalat, 2011).

2.8.4 Long-term memory (LTM)

According to the dual-store memory theory by Atkinson and Shiffrin (quoted in Kalat, 2011), information can be stored indefinitely in the long-term memory. LTM is crucial for functioning of cognition. The process of storing information here can be divided into three stages – encoding, storage and retrieval. It has been found that the longer an item is able to stay in STM through rehearsing, the stronger the associations of items and thus allow them to stay longer in LTM. The transfer of information from STM to LTM is known as consolidation. It is interesting to note that the brain does not keep all the memories in one location. They noted that each task imposes cognitive load which must either be met by using available cognitive resources or strategies like selective attention and automaticity.

2.8.5 Implications for Learning

The information processing model highlighted four important implications for the designing of the model. Firstly the storage capacities of sensory and short-term memory are extremely limited. Consequently one has to resort to some strategies to help learners cope with the limited capacity. Selective attention and automaticity are some good strategies while in language learning comprehension monitoring is being practiced (Orey, 2001; Schraw, Flowerday & Lehman, 2001; Sternberg, 2001). Suthers (1996) pointed out that the model highlighted some good learning principles which should be implemented in the classrooms.

1) Gain students' attention before content is presented
2) Review prior learning
3) Present content in a systematic and organized manner
4) Materials should be presented from simple to complex
5) Teach strategies like chunking, categorizing, reasoning, elaborating, making connections, comparing, coding, memorizing, repeating, drilling and over-learning.

2.8.6 Undergraduates' understanding of some common statistical terms

Due to a lack of local studies into the status of prior knowledge of undergraduates entering their first introductory statistics courses, a small but significant descriptive study was carried out (Foo, 2011) among Malaysian and Singaporean undergraduates. A checklist of terms was distributed to the participants to gauge their perception of their understanding of 47 statistical terms (see Appendix D). Some 56 completed forms from the Malaysian participants and 45 from Singaporeans were used for the analysis. The perceived understanding of each respondent was measured using a 4-point Likert scale ranging from 'no understanding' to 'a good understanding' of the concepts. An understanding score was then calculated based on the student's perceived level of understanding. An overall score of each item is then aggregated for each country and is labelled as degree of understanding. To standardize the mean score from each country, only similar items from the two checklists were used in the scoring. Results indicated that more familiar terms like parameter, mean, variance, skewness, normal distribution, sampling distribution, estimation, variation and probability distribution were perceived to be relatively simple as compared to more complex terms such as frequentist interpretation, posterior probability, Cohen *d*, Eta squared, Law of Likelihood approach, Bayesian approach, Fisherian approach or Neyman-Pearson approach (see Table 2.3 for a comparison across the two countries). Less than 25% of the respondents indicated a moderate to good level of understanding about these complex terms. It is pretty obvious that the respondents had little exposure and experience with this set of concepts as compared to the earlier list of terms. Students also find it moderately difficult to make sense of inference concepts like confidence intervals, *p*-value, sampling distribution, Central Limit Theorem, Type 1 and Type 11 errors and effect size. Many of these concepts are complex and conceptual understanding among these students is rather low. This is to be expected as a shallow understanding of the basic statistical terms will deter the construction of higher level statistical concepts meaningfully. Together with evidence from TIMSS studies, there are indications that prior knowledge will play a large part in students' test or examination outcomes.

Table 2.10: Malaysian and Singaporean Participants' Understanding of Statistical Concepts

No	Statistical Concepts	Degree of Understanding-Malaysia	Degree of Understanding-Singapore
1	Bayesian interpretation	23.26	8.00
2	Frequentist interpretation	13.95	8.33
3	Posterior probability	13.95	23.08
4	Strength of evidence	20.93	24.00
5	Statistical Testing Selection Skill	27.91	12.00
6	Cohen d	4.88	12.50
7	Deductive inference	18.60	12.00
8	Inductive inference	18.60	16.00
9	Statistical noise and signal	11.63	24.00
10	Eta square	6.98	12.50
11	Law of Likelihood approach	9.30	20.00

This study was exploratory in nature. It possessed limited generalizability since voluntary convenience sampling was used. The survey methodology design was considered fairly weak; however this design is sufficient to reflect the status about the perception of their statistical understanding among Malaysian and Singaporean graduates. In any event, comparisons of perceived understanding and misconceptions between Malaysia and Singapore respondents need to be interpreted within these limitations.

2.9 What are Moderators?

According to Baron and Kenny (1986) a moderator is a variable (i.e. qualitative or quantitative variable) that affects the direction and/or strength of the relation between an independent and a dependent variable. In a correlational design, a moderator is a third variable that influence the correlation between the Independent Variable (IV) and Dependent variable (DV). Figure 2.7 illustrates the framework for a moderator to function.

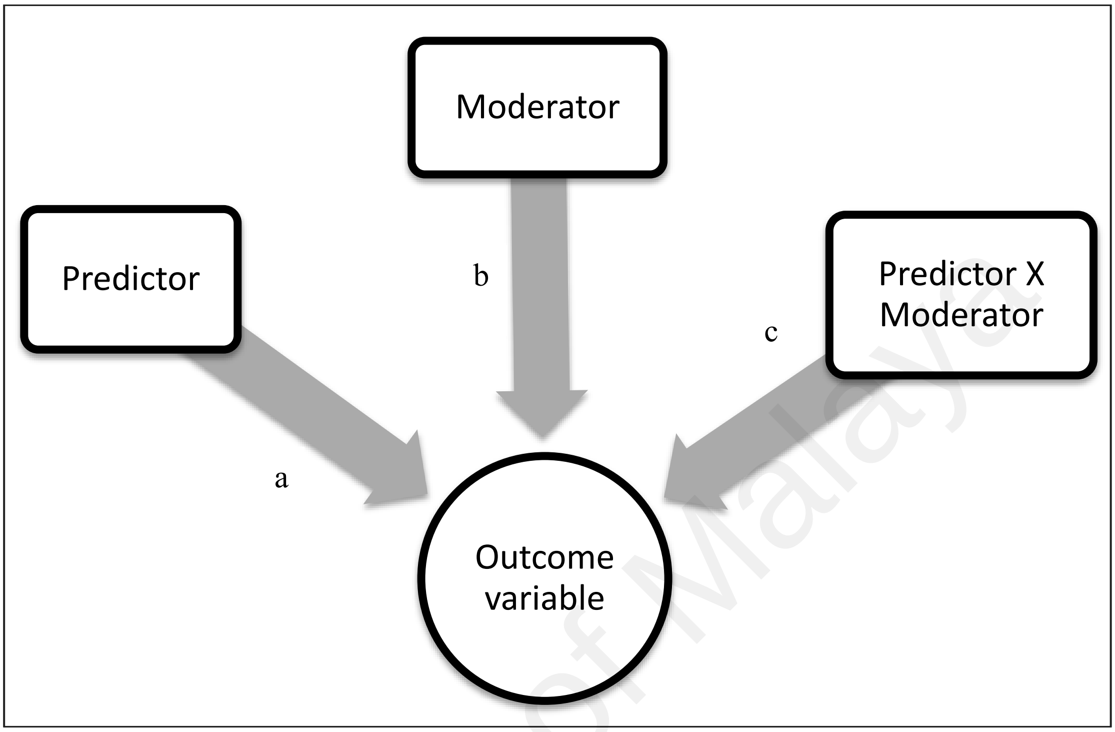

Figure 2.7: A Moderator Effect Framework for a Correlational Design (Baron & Kenny, 1986)

Figure 2.7 shows three causal paths linking to the DV which is the outcome variable. Each path is signified by an alphabet. Path ‘a’ indicates the effect of the predictor variable on the outcome variable. Path ‘b’ shows the influence of the moderator on the outcome variable while path ‘c’ shows the effect of the product of the predictor and moderator on the outcome variable. A moderation effect is considered present if path c is significant statistically. The significance of path ‘a’ and path ‘b’ is not important when testing for moderation in this framework.

The moderator is a variable that modifies a causal relationship. A simple analogy for a moderator is the volume knob of a radio that adjusts the loudness of the sound emitting from the speaker. In many case, this moderation effect is more commonly known ANOVA or MLR as ‘‘interaction’’ effect where the strength or direction of an IV on the DV depends on the level or the value of the other IV (Wu & Zumbo, 2008).

However, it is important to point out the there is a statistical distinction between moderation effect and interaction effect. Interaction analysis has been extensively applied to both correlational and experimental data. On the other hand, the term ‘‘moderation effect’’ has continuously been reserved for models that intend to make causal links. Namely, a moderation effect is a special case of an interaction effect, a causal interaction effect, which requires a causal theory and design behind the data. In other words, a moderation effect is certainly an interaction effect, but an interaction effect is not necessarily a moderation effect (Wu and Zumbo, 2008).

2.10 Summary

Literature review has shown that research into statistics education in the last two decades have leaned heavily on the teaching and learning of statistics but recently there is a clear call to look into better assessment techniques to learn more about learning difficulties in statistics and especially misconceptions.

Much has been said about how and the reasons for students acquiring those misconceptions. This chapter highlighted the problems of misplaced confidence of students when they learn statistics and paying too much emphasis on how to calculate according to a specific procedure and at the end of the routine make an interpretation of the results without really knowing why. This practice has turned statistics into a routine that invites much misinterpretations and misuses. The procedure must be learnt with understanding, applying statistical reasoning and informed judgment. To achieve that, students need to be exposed to different approaches, methods and media as there is no one technique that can address completely the problems with the teaching and learning of statistics.

CHAPTER 3 : METHODOLOGY

3.1 Introduction

This chapter describes the methods, procedures and data analysis techniques that were designed to answer the primary research purpose i.e. to investigate the structural relationships of selected cognitive determinants on statistical achievement. This chapter also explains the rationale behind the choice of the research design. The research procedure includes a section about a pilot study to check the validity of the research procedure as well as to refine items in the adapted version of the Statistical Reasoning Assessment (SRA). In addition, a multivariate statistical technique and software, SPSS 18^{th} version were described to justify its use as a data analysis tool for testing the different hypothesized models as suggested in the present study. Following this, the chapter discusses about sample, sampling design, instrument development and data collection. It ends with a short description of the procedure of the statistical data analysis.

3.2 Research Design

The research design and method in any study rest upon the researcher's worldview or in particular research paradigm. A research paradigm can be conveniently categorized as quantitative or qualitative. There are merit and demerit in the choice of either paradigm. The research approach for this present study uses a quantitative design that is elaborated in the next section.

A research design is a researcher's strategy to integrate the different components of the study in a coherent and scientific manner. The current study adopts a quantitative design to capture the evidence needed for answering the research questions effectively and unambiguously. According to Creswell (2009), a quantitative approach would be suitable if the problem is looking into identifying factors that influence outcomes or the utility of an intervention as well as attempts to understand the best predictors of outcomes. This design should be utilized if researchers wish to test a theory/theories or explanation. In addition, this is a cross-sectional study with both primary and secondary data sourced from Diploma students from a public university taking their first introductory statistics course. Multivariate analysis comprising of Principal Component Analysis and Regression modeling are employed. These types of analysis are suitable for social sciences where more often than not the focus is on investigating dependence relationships among variables. Generally, quantitative research design can be categorized into two main types i.e. Observational (correlational) or experimental (MacCallum & Austin, 2000). Cross-sectional design is a 'single-occasion snapshot of a system of variable and constructs' (MacCallum & Austin, 2000) with specifications of directional influences among the variables. Cross-sectional study as opposed to longitudinal study is considered sufficient as this study seeks only to validate the model among variables at a point in time. This design is valid as the selected variables are stable over time. For this study Multiple Linear Regression is employed to identify the relationship between the response variables and the dependent variable. It is hypothesized that the relationships among the variables in the current study are:

$$Y_i = \beta_\circ + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \qquad \textbf{3.1}$$

where Y_i= statistical achievement (SA)

$X_{1=}$ prior mathematical knowledge (PMK)

$X_{2=}$ statistical reasoning (SR)

$X_{3=}$ statistical misconception (MC)

X_4 = English Language (ENG)

X_5 = Gender (GEN)

3.3 Model Testing and Model Adequacy

3.3.1 *R*-squared and Adjusted *R*-squared

The difference between Sum of Squared Total (SST) and Sum of Squared Error (SSE) is the improvement in prediction from the regression model, compared to the mean model. Dividing that difference by SST gives *R*-squared. It is the proportional improvement in prediction from the regression model, compared to the mean model. It indicates the goodness of fit of the model.

R-squared has the useful property that its scale is intuitive: it ranges from zero to one, with zero indicating that the proposed model does not improve prediction over the mean model and one indicating perfect prediction. Improvement in the regression model results in proportional increases in *R*-squared.

One pitfall of *R*-squared is that it can only increase as predictors are added to the regression model. This increase is artificial when predictors are not actually improving the model's fit. To remedy this, a related statistic, Adjusted *R*-squared, incorporates the model's degree of freedom. Adjusted *R*-squared will decrease as predictors are added if the increase in model fit does not make up for the loss of degree of freedom. Likewise, it will increase as predictors are added if the increase in model fit is worthwhile. Adjusted *R*-squared should always be used with models with more than one predictor variable. It is interpreted as the proportion of total variance that is explained by the model. In addition, adjusted *R*-squared can help to determine if outliers exist in the data set.

3.3.2 The *F*-test

The F-test assesses the null hypothesis to see if the regression coefficients are all zero. A significant *F*-test would mean that the observed *R* squared is significant and reliable and not a random effect. In SPSS output this is the generated ANOVA table.

3.3.3 Survey Design

This study uses a survey approach to collect data on the exogenous and endogenous variables in the self-constructed model. Babbie (1990) stated that survey research provides quantitative or numeric description of trends, attitudes or opinions of a population by studying a sample of that population. Creswell (2009) suggested an eight-step survey procedure: 1) decide if surveys are the best designs to use; 2) identify research questions and hypotheses; 3) identify population, sample and sampling design; 4) determine the survey design and data collection procedures; 5) determine the instruments used to collect data; 6) administer the instruments to the targeted respondents; 7) clean up data and analyze; 8) write out the report.

Babbie (1990) suggests that a survey research has definite advantages such as a) providing for 'making refined descriptive assertions'; b) ability to collect data from a large sample; c) ability to provide the researcher to ask many questions; d) provide the researcher considerable flexibility in analysis later on. On the other hand, Babbie (1990) said that survey has its inbuilt disadvantages too, one of which was that the process of standardizing items in the survey can result in forcing the researcher to interpret incorrectly. Furthermore, the survey instrument cannot provide for capturing the feelings and emotions of respondents effectively. There is no possibility of making changes to a constructed survey form as the data collection process progresses. In the event of problems arising from this instrument, changes need to be made and the form administered again resulting in the loss of precious time, effort and finance. Due to the nature of items in a survey there is a certain degree of 'artificiality' in terms of context and suitability, thus compromising the validity of the instrument.

3.4 Sampling

The respondents were sourced from a large public institution of higher learning with the population of Diploma students spread over the 14 states of Malaysia. An initial sample of 381 Diploma students was drawn from students coming from two different states of the country that were chosen through non-random sampling. As the samples are non-randomly selected from intact classes, generalization of the findings is obviously limited but still informative in validating the model. This sample size was reduced to 374 after screening for incomplete and unusable survey forms.

3.4.1 Rationale for Sampled Population

In this study, the research questions were addressed using the findings from the data collected from a large sample of university students doing their first course in introductory statistics. These students are Science-based Diploma students who recently graduated from the Malaysian O-level examination (*Sijil Pelajaran Malaysia*). The respondent selection criteria include their demographics, academic background and conceptual understanding and exposure to the different level of reasoning abilities. Science students who graduated from the *SPM* examination level have two years of additional mathematics that cover ten hours of learning basic statistics in their upper secondary school life. These students enter Diploma courses in this university at the age of 18. Their exposure to formal statistical reasoning and misconceptions can be considered low except for some informal statistical knowledge from mathematics courses in the earlier part of this Diploma courses. This study hopes to contribute to the body of statistical knowledge concerning factors and their interactions among Diploma Science students in a public university in Malaysia.

3.4.2 Descriptions of sample and sample size

Initial number of respondents was 381 sourced from two campuses. After data cleaning, the final sample comprises of 374 usable forms. The students enrolled in a first course in introductory statistics course came from two states in Malaysia. These two states out of the 14 states were selected using a non random sampling technique. The Diploma students come from different Science programs from the Faculty of Applied Sciences of the same university. The course is accredited 3 credit hours by the faculty and is undertaken by students. The classes are taught for 4 hours per week across 14 weeks Each week, the lesson comprises of three parts; lectures, tutorials and lab work using SPSS software.

The students are all Indigenous students (*Bumiputeras*) where the mother tongue is the Malay language. All the students are educated using the primary language of the Malay language and English as the second language. After selecting the states, permissions were sought to collect data from selected classes identified by the lecturers teaching those courses. The research used purposive sampling in the selection of the classes due to the constraints of the need to monitor evaluation grading and standardization of teaching throughout the semester as there were three different Statistics lecturers handling those classes. Thus random sampling was difficult with such a large population. The sample was tested and data continuously collected over one semester taught by the said-lecturers including the researcher.

One of the critical factors to consider in a quantitative design like MLR is the question of sample size. According to Hair, Anderson, Tatham and Black (1999), the desired ratio of sample to independent variables is 20 to 1 but 15 to 1 is sufficient. As the popularity of multiple linear regression (MLR) increased, the question of how large a sample is important to produce reliable results especially for prediction purposes. Maxwell (2000) states that ‘‘sample size will almost certainly have to be much larger for obtaining a useful prediction equation than for testing the statistical significance of the multiple correlation coefficient'' (p. 435). In a study carried out by Knofczynski and Mundfrom (2008), 'a definite relationship, similar to a negative exponential relationship, was found between the squared multiple correlation coefficient and the minimum sample size'. They stated that this relation is directly related to the ability of the MLR to make good predictions.

3.5 Data Collection Instruments

The variables in the model used for this investigation are represented by Prior Mathematical Knowledge (PMK), Statistical Reasoning (SR), Statistical Misconception (MC) and Statistical Achievement (SA). Both primary and secondary data were collected over a period of one semester. Secondary data consist of scores to calculate Prior Mathematical Knowledge and Statistical Achievement. Prior Mathematical knowledge comprises of aggregated score based on grades from General Mathematics and Additional Mathematics taken in their *Sijil Peperiksaan Malaysia (SPM*), an O-level equivalent examination at the end of 11 years of compulsory schooling plus some mathematics courses taken in the first three semesters of their Diploma program. As for the Statistical Achievement score, it is a composite score consisting of their semester test scores and final examination results. The instruments to collect these scores are standard examination papers set by the Examination Council of Malaysia as well as carefully vetted examination and test papers set for all students in this university. (See Appendix C for the methods used to calculate the aggregated scores of the cognitive factors.)

Demographic profile of participants and scores for Statistical Reasoning and Misconception variables were collected through the use of the Statistical Reasoning Assessment Instrument (SRA) adapted from the version by Garfield (2003). A cover letter accompanied the instrument informing the respondents about the purpose and importance of this study, confidentiality of the information provided and instructions on how to answer. All answers given were collected by the lecturers in charge on the same day of its administration. A five-page survey was designed and piloted based on items from SRA (Garfield, 2003). (See Appendix A1).

The final version is given in Appendix A2 where some of the items were rewritten to suit the local context. The main purpose of the pilot studies was to improve the low reliability of the SRA. This was done through the two pilot studies carried out before the real study. In the pilots, the focus group comprises of students and the statistics lecturers went through the items in the original SRA instrument and revised SRA instrument to weed out unsuitable items. The 15-item multiple-choice instrument comprised of two sections: Section A consisted of five open-ended questions to collect information on gender, highest academic qualification, language mastery, prior mathematical knowledge, faculties and statistics courses attended. Section B contained 15 items asking for the respondents' reasoning abilities in 5 main topics taught in this introductory statistics course covering data, distribution, averages, variation and probability. Each multiple-choice item has between 3-6 options depending on the complexity of the items constructed to gauge the reasoning skill of the respondents. Respondents were only required to choose the best option. Each correct answer contributes to an aggregated score for statistical reasoning. The other incorrect options in each item are specially designed to identify the kind of misconceptions carried over from previous statistics courses. The estimated time required to complete the questionnaire based on pilot study was 40 ± 5 minutes.

Item scoring depends on two scoring rubrics designed to measure the respondents' reasoning and misconception (see Appendix B). The method used for calculating the aggregated scores of some of the variables. Briefly the aggregated score for language mastery is measured by combining the grades using Grade Point Aggregate (GPA) scoring as practised by this university. Students' grades in their *SPM* examination and the grades achieved in their compulsory Basic English courses for three semesters were utilized to calculate this score. The PMK score is sourced from the reported grades by each respondent based on their mathematical achievement during his/her *SPM* examination and the grades of the finals for three consecutive semesters. The grades are converted to GPA points and averaged out. The SR score is calculated by adding up all the number of correct answers and divide it by 15. The MC score is calculated by adding up all the number of incorrect answers and divide it by 15. This score is calculated by adding up all the number of incorrect answers and divide it by 15. Finally the SA score is calculated by using the marks achieved by each respondent in his/her final examination statistics paper "Introduction to Statistics". (Language mastery, prior mathematical knowledge, statistical reasoning, statistical misconception and statistical achievement are described in details in Appendix C).

3.6 Procedures for Implementation of Study

The main instrument, the SRA, is responsible for collecting data on exogenous and moderating variables used for building a few regression models. The endogenous variable and exogenous variables were measured using indicators from assessments like quizzes, tests and examination results from the respondents' secondary school final year and compulsory courses from their diploma program in this university.

3.6.1 Preliminary study

Before the study was carried out, permission to run the study in the university concerned was sought and approval by the relevant authorities was secured before the actual study. A pilot study is important to simulate the proposed procedure used in the actual study. This mini study is a feasibility study to determine the suitability of the following: a) the estimated period of time to carry out the study, b) the instructions for administrating the multiple-choice SRA instrument, c) the choice of the participants, d) the sequencing of the research procedure, e) finance, and f) choice of assistant researchers who will be administering the SRA instrument. Within the preliminary study, a pilot test was run to gauge the suitability, reliability and validity of the SRA instrument.

3.6.2 Pilot testing

The main purpose of doing a pilot study was to check on comprehension issues with the SRA instrument. This is intended to improve the reliability of the instruments. It is important to ensure diploma students understood the instructions, clarity of content and context, missing items, suitability of options. Both individual testing and focus group interview were carried out to improve its reliability and validity. Additionally this piloting was to evaluate the time, cost, unforeseen events, and sample size requirement with the aim of improving upon the study design prior to the actual study. The SRA started with the analysis of the SRA used in studies by Garfield (2002), Liu (1998) and Tempelaar et al. (2007). Both the content and context of the items were categorized and compared to the SRA instrument used by Zuraida et al. (2012). After reviewing both the instruments, a new version was drafted and sent for face validation. This procedure was carried out by two senior statistics lecturers teaching in the university where the main study will take place. The final version of this SRA instrument consisted of 15 items and was readied for pilot testing to a group of 58 Diploma students who were not involved in the real study.

The first assessment of this version was carried out at the beginning of March, 2014. Specific instructions were given to students to take note of items they found to be difficult to understand in terms of language or concept or both. Following that, an item analysis was done to determine item difficulty and item discrimination for improvement of the SRA instrument. This helps in determining the validity and reliability of the items constructed.

3.6.3 Item Analysis

Item analysis is a procedure meant to examine collectively student responses to the individual items comprising the SRA instrument. This process functions as a tool to assess the quality of the items and consequently the quality of the instrument itself. This approach can help to improve items in subsequent testing of the items as well as eliminate ambiguous items or bad items. Ultimately with this approach, it is possible to improve the reliability of the SRA. The analysis provides the user with two important indices – difficulty index and discrimination index.

Difficulty index measures the proportion of students who could answer a particular item correctly. It ranges from 0 to 1. A zero score means that none of the students can answer that item while a score of 1.0 represents all students answered correctly. A general rule of thumb is that an item difficulty should be between 0.6 to 0.8 where items with an index of less than 0.6 mean that they are either too difficult, not well written or there may even be more than one answer.

On the other hand, items with 0.8 and above are probably too easy and need to be substituted with an item that is usable i.e item with item difficulty between 0.6 to 0.8.

Item discrimination explains how well an item can differentiate between a 'high achiever' and a 'low achiever' It is actually a point-biserial correlation measures with a range of -1.0 to +1.0 like any correlation index. A positive index means a positive correlation between the different levels of achievement among the students while a negative index indicates an inversed relationship where 'good' students answer incorrectly more frequently than 'bad' students. The items should be positively correlated and index nearer to 1.0 is preferred.

A rule of thumb suggests that 0.2 and above is to be desired.

As seen in Table 3.1, a preliminary analysis of the difficulty level and discriminatory ability of some of the SRA items indicates that item 1, 2, 4, 11, 13 and 14 top the list as most difficult to answer and does not seem to be able to discriminate the good from the poor. Based on the appropriateness of index as discussed in the previous section, the following items can be revised to increase the validity and reliability of the instrument i.e. items 1, 2, 4, 11, 13 and 14. In the next stage of pilot testing, these items as identified above went through another round of item review to produce better items.

To assist further this continual process of refinement and improvement a focus group interview was conducted in phases.

Phase 1: Focus Group

The focus group procedure followed the protocol suggested by Eliot et al. (2005). The questions used in the focus group were related to the 15 items where students were asked in particular why they choose a certain option. The purpose is to understand the rationale behind each of the choices. They were encouraged to speak freely and without interruptions as other interviewees in the group can come in to give their opinion. This created a lively discussion with the focus on the items and their suitability in terms of language, content and context. The whole session took over one and half hours with all conversation recorded. The recording was transcribed and themes were identified. These new evidence were utilized to improve the items and instructions in the SRA. With feedback from the first assessment, the new version was developed.

Phase 2: Assessing the SRA instrument

The second assessment of this version was carried out with a sample of 54 Diploma students who were not targeted to be involved in the real study although they took the same course. Two full-time statistics lecturers helped in the data collection for

Table 3.1: Difficulty index and Discrimination Index of SRA instrument

Item	# Correct (Upper group)	# Correct (Lower group)	Index of Difficulty (p)	level of difficulty	Discrimination (D)	Most popular option	% of students choosing this
Question 1(c)	0	1	3.4	high	-0.1	q1b	86.2
Question 2(d)	1	1	10.3	high	0	q2e	51.7
Question 3(d)	8	5	72.4	low	0.3	q3d	72.4
Question 4(a)	0	2	10.3	high	-0.2	q4b	69.0
Question 5(c)	10	6	79.3	low	0.4	q5c	79.3
Question 6(e)	8	6	62.1	low	0.2	q6e	62.1
Question 7(c)	7	2	37.9	moderate	0.5	q7c	37.9
Question 8(e)	7	5	51.7	moderate	0.2	q8e	51.7
Question 9(b)	4	2	32.1	moderate	0.2	q9a	42.9
Question 10(a)	6	0	28.6	high	0.6	q10c	57.1
Question 11(b)	2	0	7.1	high	0.2	q11a	50.0
Question 12(b)	5	0	35.7	moderate	0.5	q12b	35.7
Question 13(b)	2	1	17.9	high	0.1	q13a	39.3
Question 14(a)	2	1	25.0	high	0.1	q14d	53.6
Question 15(b)	8	3	53.6	moderate	0.5	q15b	53.6

this stage. This part is crucial to determine the inter-item reliability and construct validity. The size of n = 54 was used to run a linear multiple regression model. The model was run using scale data from the independent variables (Prior Mathematical Knowledge, Misconception, and Statistical Reasoning) and dependent variable (Statistical Achievement). The dimensions and items for statistical reasoning and misconceptions were reclassified from suggestion using Principal Component Analysis (PCA). With the final improvement of this version (see Appendix A2), the study was considered to be ready for implementation. The various process employed to address the low reliability issue of SRA make it a valid and reliable instrument to collect statistical reasoning and misconceptions.

Phase 3: Principal Component Analysis

Once the new instrument was ready, it was used to collect data from 206 respondents to run a Principal Component Analysis. This sample was part of the respondents from the real study. It was collected from the first campus.

3.6.4 Results of Principal Component Analysis for pilot testing of SRA (n = 206)

Unidimensionality is an important concept in psychometric instruments and its influence on reliability statistics like Cronbach Alpha – the measure of the internal consistency reliability is very significant.

Thus, for an instrument like SRA to have construct validity, the items must be shown to load onto a fixed number of dimensions. To do that SPSS provides a few options to measure construct validity i.e. Principal Component Analysis (PCA) or Factor Analysis (FA). PCA can confirm what dimensions each question in SRA loads on to.

PCA provides the researcher with indices as to the viability of the different dimensions or subscales for both the statistical reasoning and misconception scales. The eigen values determine the number of dimensions of the SRA based on the sample data.

Furthermore its analysis identifies the loadings of the items onto the factors or dimensions already identified as discussed in section 3 previously i.e. loadings of 1.00 or more are chosen. This will serve to re-specify the model if needed and determine the reference indicators that are relevant to the factor structure.

Table 3.2: Dimensions of SRA (Garfield, 2003)

	Correct Reasoning Skills (CC)	Item/Alternative	Max. Score
1	Correctly interprets probabilities	2d, 3d*	2
2	Understands how to select an appropriate average	1d, 4ab, 12c	3
3	Correctly compute probabilities Understand probabilities as ratios Use combinatorial reasoning	5c 10a, 13b, 14a, 15b	5
4	Understand Independence	6e, 7d, 8e	3
5	Understand sampling variability	11b	1
6	Understand the importance of large samples	9b	1

Table 3.3: Dimensions from PCA analysis based on dataset (n=206)

	Correct Reasoning Skills (CC)	Item/Alternative	Max. Score
1	Correctly interprets probabilities	2d, 5c, 11b	3
2	Understands how to select an appropriate average	1d, 4ab	2
3	Correctly compute probabilities Understand probabilities as ratios Use combinatorial reasoning	8e, 14a, 15b	3
4	Understand Independence	3d, 6e, 13b	3
5	Understand sampling variability	7d, 10a, 12b,	3
6	Understand the importance of large samples	9b	1

*3d means item no. 3 in the SRA instrument and the correct answer for that item is d.

As seen in Tables 3.2 and 3.3, the PCA showed six dimensions in the SRA instrument which has been classified similarly as what had been done by Garfield (2003) but the items used to represent each of the dimensions are significantly different. For example in the case of Garfield (2003), the items used to represent the dimension 'correctly interprets probabilities' was represented by items 2 and 3 but in this study, this dimension is represented by items 2, 5 and 11. The difference in classification is expected due to the issue of reliability of the SRA items. Another factor contributing to this low reliability is the small numbers of items constructed for each dimension with some dimensions represented by one or two items! (See Table 3.4 for the distribution of items to dimensions).

Figure 3.1 provides the detailed analysis of the PCA carried out using a sample of 206 respondents.

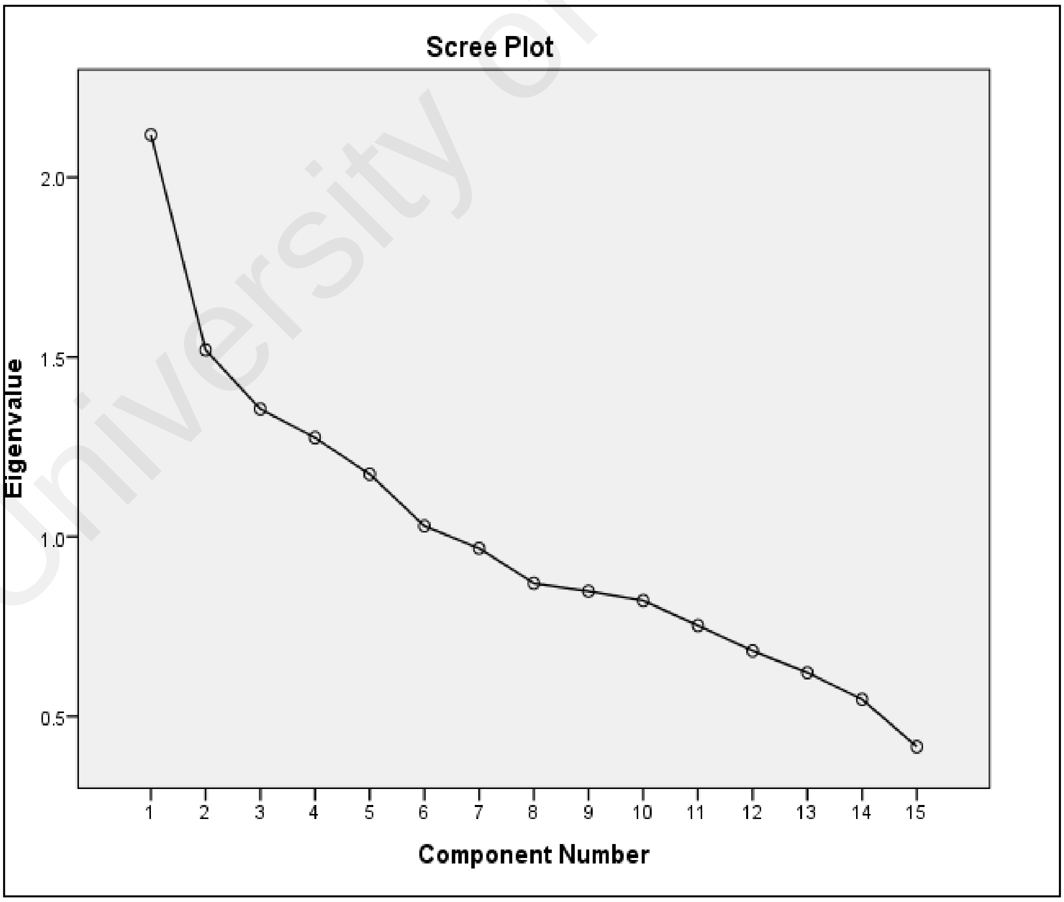

Figure 3.1: Scree Plot showing the six dimensions/components

Table 3.4: The extracted six components after rotation

	Component					
	1	2	3	4	5	6
q1			.740			
q2				.617		
q3		.481				
q4			.691			
q5				.674		
q6		.561				
q7					.748	
q8	.539	.471				
q9						.801
q10					.532	
q11				.594		
q12					-.511	
q13		.684				
q14	.756					
q15	.775					

Rotated Component Matrix[a]
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.[a]
a. Rotation converged in 12 iterations.

Table 3.4 shows the item distribution based on the Rotated Component Matrix. There are some items that had been categorized differently from the one used by Garfield (2003).

In Garfield (2003) research, she used the items to identify students' misconceptions. Table 3.5 explains the different forms of misconceptions that can be evaluated using the SRA. The present study investigates the different levels of misconceptions but the primary interest is to measure the overall misconception level by using the piloted SRA instrument. As there are many different forms of misconceptions, the misconceptions measured in the present study are as listed in Table 3.5.

Table 3.5: Misconceptions in Statistical Reasoning (Garfield, 2003)

	Misconceptions (MC)	Item/Choice	Max. Score
1	Misconceptions involving: Averages are the most common number Fails to take out outliers Confuses mean with median	1a* 1c 12a	3
2	Outcome Orientation misconception	2e, 3ab, 8abd, 9c, 10b	5
3	Law of small numbers	9a, 11c	2
4	Representatives misconception	6abd, 7d, 8c	3
5	Equiprobability bias	10c, 13a, 14d, 15d	4

*1a means the student had misconception involving averages if he had chosen option a.

The next chapter discusses the outcomes of misconceptions identified from the choices of answers given by the respondents during the actual study.

Generally all students suffer from one form of misconception to another form. For this particular set of students it was mainly skewed towards misconception about averages, outcome orientation problem, Law of small numbers misconception and equiprobability bias. Literature as described in Chapter 2 has outlined the underlying causes of these common misconceptions. Please refer to Table 2.7 and 2.8 of Chapter 2.

3.6.5 Validity and Reliability issues of SRA

The main concern of any assessment instrument is the credibility of the results generated. Two key issues in evaluating a test instrument are reliability and validity. To determine the reliability of the test, psychologists refer to an association score known as a correlation coefficient, test-retest reliability, inter-item reliability, parallel form reliability and Cronbach alpha. Equally important when evaluating a test is the issue of validity.

Assessment experts would like to consider three types of validity: construct, internal and external. Validity of the test concerns itself with whether the test measures what it is supposed to measure. Construct validity is about the translation of a concept or construct into a functioning entity that can be studied empirically (Trochim, 2006). A test has construct validity if it can measure the construct of interest by using an operationalized version of this construct. The construct comes from the population while the operationalized version comes from the sample. If the aim is to measure intelligence (construct) through the use of an algebra test, then construct validity will be an issue because a good knowledge about algebra (operationalized construct) is not translated into a measure of intelligence (construct). Construct validity is a very general term. In research this validity can be subdivided into face, content and criterion-related validity like predictive, concurrent, discriminant and convergent validity. Studies reporting the validity and reliability of the SRA instrument are limited to those by Garfield (1998, 2003); Garfield and Chance (2000); Liu (1998); Sundre (2003) and Tempelaar et al., (2007). One of the first studies by Garfield (1998) and a later study by Garfield and Chance (2000) to show criterion validity using aggregated scores indicated extremely low correlations between reasoning and misconception scales on achievement scores. The inter-correlations matrix between the items was generally quite low implying serious problem with internal consistency when using aggregated scores. They had better results using a test-retest reliability approach with r = .7 and r = .75 for the reasoning and misconception scales respectively.

Similarly, Liu (1998) reported a test-retest reliability of r = .70 for statistical reasoning score while she obtained r = 0.75 for the misconception scores. These scores were aggregated based on the calculation of adding the scores for each subscale together to form a composite score. Garfield (2003) reported lower reliabilities for both categories of aggregated scores. Tempelaar et al. (2007) attempted with a similar approach using aggregated scores and found similar reliability indices as Garfield. Their studies showed that Cronbach alpha for both the scales were 0.24 and 0.06 respectively. All these studies yielded unremarkable results even after taking into account items with extremely small p-values and adjusting for subscale effects had little effect on these reliability indices. Analysis of the correlation matrix between all SRA correct reasoning and misconception based on Liu & Garfield (2002) study, showed very low correlation and even negative ones. These negative but significant correlations were identified by Tempelaar et al. (2007) as the cause for the low reliability indices. Tempelaar et al. (2007) suggested the SRA measurement model and the structural model should not use aggregated scores but to model the relationships separately for each of the subscales with the other variables (see Table 3.4 & Table 3.5 for comparison).

Garfield herself admitted that there is much to be done to improve the SRA after studying the results of the reliability and validity indices from the various studies mentioned earlier. Konold & Higgins (2003) concurred on this and commented that the SRA is still an imperfect research and evaluation tool where more work needs to be done. Limitations of the SRA includes problems with the subscales that represent only a small part of the reasoning skills in the introductory course; indicators for the reasoning and misconception latent variables are suspected; and the inappropriate usage of the aggregated scores in the models. Thus, the findings from recent studies had raised new issues and yielded incomplete results prompting new directions and stringent procedures for researchers to carry out better studies to overcome the present weaknesses of the SRA.

3.6.5.1 Checking for Reliability of SRA using Cronbach Alpha

One of the commonly used measures of internal consistency/reliability of an instrument is the ubiquitous Cronbach alpha. The computation of this index relies heavily on the number of items of the instrument and the average inter-item covariance.

Reliability test on SRA instrument with n=206 usable sample.

Table 3.6: Case Processing Summary

		N	%
Cases	Valid	206	96.3
	Excluded[a]	8	3.7
	Total	214	100.0

a. Listwise deletion based on all variables in the procedure.

Table 3.7: Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.497	.492	15

Table 3.8: Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
q1	39.0000	26.702	.097	.109	.493
q2	37.3689	23.161	.185	.102	.479
q3	37.7039	25.468	.093	.071	.499
q4	38.8835	26.201	.091	.090	.495
q5	38.5340	25.840	.155	.099	.484
q6	36.4126	24.322	.369	.250	.447
q7	38.3883	25.263	.147	.109	.485
q8	36.9175	22.222	.342	.259	.431
q9	39.3107	26.703	.038	.051	.504
q10	39.0874	25.875	.090	.121	.497
q11	39.2816	24.037	.285	.123	.455
q12	39.0146	26.327	.061	.118	.502
q13	38.7670	25.028	.089	.082	.504
q14	38.1214	22.019	.308	.279	.439
q15	38.1602	23.969	.237	.298	.464

The reliability analysis shows only a moderately measure of Cronbach alpha of 0.497. One reason for this rather low figure of consistency among the respondents could be the small number of items in the SRA instrument. From literature discussed in Chapter 2, the alpha found in many of the studies similar to this one, is found to be consistently low to moderate (Garfield, 2003; Tempelaar et al., 2007).

3.7 Actual study

Prospective participants were purposively selected from 6 classes of Introductory Statistics Course in a large Malaysian university. A sample of n=381 was selected. The criteria for selection were mainly based on the availability of the classes to take the SRA test and most importantly the willingness of the statistics lecturers and the students who volunteered for this study. After briefing the lecturers on the purpose and conduct of this study, the content and answers of the SRA instrument was discussed in detail as well as the instructions and procedure for the administration of this instrument. The confidentiality of the name and responses of participants were assured. Each lecturer and the main researcher gave out the instrument according to the time table agreed upon by the lecturers concerned and each test took an estimated time of 40 ± 5 minutes to complete. All scripts were collected and handed over to the main researcher who subsequently entered the data. The answers to the test were discussed with the lecturers who taught the participants to ensure that these answers were acceptable. The scoring rubrics for both the reasoning scales and misconception scales were also adjusted from the feedback of these lecturers (see Appendix B).

3.8 Data Analysis Techniques

Data cleaning and data screening were done to filter out data that were considered unusable or incomplete. An exploratory data analysis (EDA) was carried out to get a feel of the data; check for normality of variables; linearity and homoscedasticity of the data set as well as looking out for outliers. Some of the outliers were deleted while some were checked against the original answer scripts to ensure correct data entry. Any outliers that were 3 standard deviations away from the cell means and also discontinuous from the trend observed were deleted to prevent them from influencing model evaluation (Bollen, 1989). Missing values were treated as suggested by SPSS using data imputation where mean values for variables where substituted on the condition that the data set had less than 10% missing values (Kline, 1998).

As illustrated by Byrne (2001), two critical assumptions in MLR are the requirements for the data to be continuous and possessing a multivariate normal distribution. Ignoring the requirement of normality especially when the data appears to be significantly skewed will cause the χ^2 value to be inflated. When sample size is small and non-normality increases, Boomsma (1985) indicated that an increased incidence of non-convergence of analysis and improper solutions will affect the output. Furthermore, fit indices may be modestly underestimated (Marsh, Balla & McDonald, 1988). Ultimately there is an underestimation of standard errors causing 'the regression paths and factor/error covariance to be statistically significant when they are not so in the population' (Byrne, 2001). Multivariate normality can be assessed using MLE approach by examining skewness, kurtosis and univariate normality of the set of variables. If the data is found to be non-normal, z transformation is recommended to be used.

Much of the analytic procedure used in this study followed the suggestions from Field (2013) and Randolph and Myers (2013). In summary the procedure involved:

Step 1: Recode categorical variable into new dichotomous variable called Dummy variable (i.e. Gender, Language Mastery… etc.)

Step 2: Conduct preliminary analyses

a. Examine descriptive statistics of the continuous variables
b. Check the normality assumption by examining histograms of the continuous variables
c. Check the linearity assumption by examining correlations between continuous variables and scatter diagrams of the dependent variable versus independent variables.

Step 3: Conduct multiple linear regression analysis

a. Run model with dependent and independent variables

b. Model check

Step 4: Examine collinearity diagnostics to check for multicollinearity

a) Examine residual plots to check error variance assumptions (i.e., normality and homogeneity of variance)

b) Examine influence diagnostics (residuals, dfbetas, dffits) to check for outliers

c) Examine significance of coefficient estimates to trim the model

Step 5: Revise the model and rerun the analyses based on the results of steps 1-4.

Step 6: Write the final regression equation and interpret the coefficient estimates.

3.8.1 Statistical Software

One statistical software were used in this study namely SPSS version 18. The rationales for the choice of this software had been discussed in Chapter 2. The statistical analysis of the data was first carried out in the preliminary study and also in the actual study.

3.8.2 Preliminary Analysis

The preliminary study only used SPSS to generate multiple regression output to shed light on the significance of the relationships between the exogenous and the endogenous variables. Then the reliability index using Cronbach Alpha was calculated. The multiple regression analysis looked at the relation between a DV with several selected IV under the relevant assumptions. In this study the DV is Statistical achievement, while the IVs are: Statistical Reasoning (SR), Misconception (MC) and Prior Mathematics Knowledge (PMK). Multivariate methods require the assumption of normality i.e. data has a multivariate normal distribution. Shapiro-Wilks test and Chi-square plot can be used to check this assumption. Usually the *p*-value for Shapiro-Wilks must be more than 0.05 and the skewness index at ±1. Two other tests are used to assess the overall sufficiency of the model, R^2 and the adjusted R^2. If the value of R^2 is close to 1 imply that most of the variability in dependent variable is explained by the independent variables.

ANOVA table in SPSS is useful to determine which regression coefficients are significant. If F value is large, then one knows that at least one IV differs. Once it has been determined that at least one of the variables was important, one proceeds to test on individual regression coefficients. If p-value is less than 0.05, the correlation is significant.

3.8.3 Missing values

Missing values or incomplete data are common occurrences in data collection. Incomplete data set has implication on the analysis. Kline (1998) suggested that for missing data that were less than 10% of the total cases, mean imputation can be used to replace them. On the other hand, missing data may be due to certain reasons that will cause what is termed as pattern of missing data. However, the approaches to replacing the missing data or deleting them altogether are much more complicated. The approaches generally depend on three well-established patterns (Little & Rubin, 1987) - MCAR (Missing Completely At Random), MAR (Missing At Random) and NMAR (Nonignorable Missing At Random). For SEM models, by far the commonest method is to use listwise deletion (Boomsma, 1985) and sometimes mean imputations under certain constraints (Kline, 1998). For MCAR cases, Arbuckle (1996) suggested the use of listwise deletion approach. When using pairwise deletion for MCAR cases, it differs from listwise deletion in that 'only cases having missing values on variables tagged for a particular computation are excluded from the analysis'. This approach has the advantage of preserving less deletion of cases which in turn provides for a higher sample size. This means that different computations of selected variables can have varying sample sizes.

3.8.4 Methodological issues on the use of multiple regression analysis

With the objectives of this study in mind, the choice of statistical analysis techniques to achieve them effectively is of prime concern. Although the model can be broken into separate individual multiple regression equations to see the interactions among the variables, due to many constraints (e.g. inflated p-values, measurement errors, unreliable chi-squares statistics among others) this would be a poor approach to choose. Many variables in psychology and education are constructs that are not observable directly. Variables like achievement, reasoning, misconceptions and prior knowledge here are assumed that the errors are considered non-existent. Although Goldberger and Duncan (1973) noted the advantages of structural equations like Structural Equation Model (SEM) over regression parameters under the following circumstances - a) when the observed measures contain measurement errors especially when the variables of interest are among the true effects; b) when there is interdependence or simultaneous causation among the observed variables and c) when important explanatory variables had been omitted unknowingly, it was found the MLR is adequate for the variables in this study. One of the strengths of multiple linear regressions is that one can include factors that can control for spurious effects. However, there always remains the possibility that a spurious factor remains untested as opposed to using SEM. Even though multiple variables may be included in the statistical model, it is still possible to have spurious relationships of which extra care must be taken by the researchers. In addition regression models take into account less complex relationships involving many variables which are observable.

The MLR does have some inherent weaknesses like 1) able to only account for one dependent variable and 2) variables can only be either independent or dependent. In real situations, it is more probable that the analysis involves two or more dependent variable interactions. Furthermore it is normal to be a dependent variable under one scenario and may well be an independent variable in another.

Though these are some of the weaknesses to be aware of, this study does not suffer from such weaknesses as it is only interested in investigating one dependent variable i.e. statistical achievement. In addition, the independent variables are pre-determined from literature review.

3.8.5 The Choice of Software for Analysis

The analysis for the actual study utilizes a well-known software i.e. SPSS. All data used SPSS data file format and analysis of regression models can be carried out within SPSS environment. The choice of SPSS is due to its easy availability of software in public universities all over Malaysia and the researcher's exposure and experience with this software. SPSS is adequate for social science studies of which this study is about.

Descriptive statistics like group sample sizes, mean, standard deviation, standard error, confidence intervals, maximum and minimum were first generated and presented in tabular and graphic format. Demographic profile like gender, highest academic qualification, schooling background, language spoken at home, and statistical experience of the sample were presented and checked to ensure completeness of data. Exploratory data analysis was routinely carried out to look out for outliers and the percentage of missing values in each variable of interest in addition to identifying suspicious data. Data cleaning assures a better and reliable result.

3.8.6 Screening for assumptions of multiple regression

The data must be screened before analysis for univariate and multivariate normality by way of appropriate statistical tests, skewness, kurtosis or other visual techniques like score distribution . One good way to check this is by studying the skew and kurtosis of the individual score distribution of the variables in the model. An absolute index of less than 1.0 shows univariate normality while anything above 2.0 is considered moderately non-normal (Finch, West and MacKinnon, 1997). They noted that for non-normal data the researcher will see an inflated chi-square statistics. Similarly the output holds for multiple regression or correlation when the data is assumed to be linear and the variances of comparing variables are roughly equal. When sample size is large these two assumptions do not have significant impact on the results. It is good practice to check for them in all cases.

3.9 Selecting the best regression model

In constructing a complex model, the critical question to ask about how predictors are selected. This is very important as the regression coefficients depend on these variables. Furthermore the way in which they are entered too can have a great impact on these coefficients (Field, 2013). In normal circumstances, the variables to enter comes from past research but if new predictors are to be inserted, then it is important to note that an exploration of how strongly correlated to the variables identified through past research can be used.

The selection of the variables to be included in the best regression model can be carried out by studying the correlation matrix. The Pearson r for these variables can give an indication of the manner of entry of a particular variable when the stepwise forward technique is being employed as this is based on purely mathematical criterion (Field, 2013).

Deciding on order of entry of variables into model

This is very important as the values of the regression coefficients are partly influenced by the mode of entry of the variables. The way in which variables are entered too can have a great impact on these coefficients as had been clearly explained by Field (2013).

According to Tabachnick and Fidell, (2001) three main options in multiple regression can be chosen i.e. standard multiple regression, hierarchical multiple regression, and stepwise regression. If the standard multiple regression is used, the independent variables are included into the equation simultaneously. This technique is useful for assessing the relations among small number of variables. For the hierarchical multiple regression, the order of entry of variables is important and must be determined before the analysis. The order is normally determined based on past research. The third approach is known as stepwise regressions. As opposed to the other options, decisions about inclusion or omission of the variables from the equation rest upon chance and statistics. 'The stepwise regression also looks like over fitting data because the equation derived from a single sample is too close to the sample, and may not generalise well to the population" (Tabachnick & Fidell, 2001).

The current study employs the stepwise estimation method as it is a better approach of selecting the best predictors for inclusion in the model to be fitted. Each variable is included based on an 'incremental explanatory power they can add to the regression model' (Hair, Anderson, Tatham & Black, 1999). The concept of this technique is to select those IVs with significant partial correlation coefficients. According to Hair, et al. (1999) additional variables may not necessary increase the predictive power of the model but could be counterproductive by reducing it. Strong bivariate correlations among the various variables do not indicate their predictive power. In a multivariate context, some of these bivariate correlations may well be redundant and not needed at all in the regression model if another set of variables could explain this variance better.

The selection and order of entry of the variables for this study requires certain regression technique that involves partial correlation matrix and partial *F*-test. In addition, the stepwise forward technique would be suitable to use (Field, 2013).

The procedure to determine the order of entry

a) Select variables in order of priority when entering into the model

b) Run a partial correlation procedure to find the next important variable by inspecting which variable has the strongest correlation with SA after taking out the variance due to the first variable. This step is repeated until all the variables are assessed.

c) Determine the variables that do not contribute to this variance. Thus these will be eliminated from the model.

d) Run a partial *F*-statistics test to determine if that variable contributes significantly to the variance measured. If the test is significant, retain that variable

e) Once the order of entry for the important predictors is determined, enter the selected variables accordingly.

f) Generate the regression model. The outputs include the model summary, correlation matrix, partial correlation matrices, scatterplots and partial scatterplots and histogram.

3.9.1 Deciding on the best model

The following procedure was employed to answer research questions (i), (ii), (iii) and (iv) that include determining the best fit models and identifying the cognitive determinants of significance. The stringent procedure known as model diagnostics is reported here before it can be concluded about the best model to select (Li, 2007). These steps include:

Step 1: Recode categorical variables into new dummy variables

Step 2: Conduct preliminary analyses using descriptive statistics of the continuous variables. Check the normality assumption by examining histograms of the continuous variables. Check the linearity assumption by examining correlations between continuous variables and scatter diagrams of the dependent variable versus independent variables.

Step 3: Conduct initial multiple linear regression analysis by running the model with dependent and independent variables

Step 4: Model Assumptions to look out for:

-collinearity diagnostics to check for multicollinearity

-residual plots to check error variance assumptions (i.e., normality and homogeneity of variance)

-diagnostics (residuals, dfbetas) to check for outliers (Li, 2007)

Step 6: Examine significance of coefficient estimates to trim the model

Step 7: Select important variables to be entered into the model where priority of entry depends on the strength of that variable with the dependent variable, SA

Step 8: Run a partial F-statistics test to determine if that variable contributes significantly to the variance measured. If the test is significant, retain that variable

Step 9: Run a partial correlation procedure to find the next important variable by inspecting which variable has the strongest correlation with SA after taking out the variance due to the first variable.

Step 10: Determine the variables that do not contribute to this variance. Thus these will be eliminated from the model.

Step 11: Run a partial *F*- statistics test again to determine if the variable contributes significantly to the variance accounted for.

Step 12: Enter the selected variables in sequence into the model according to their importance

Step 13: Generate the regression model.

Step 14: Assess the accuracy of the regression model – 1) assess whether the model fit the observed data and 2) assess whether the model can be generalized to other samples (Field, 2013).

Step 15: For assessing model fit, check if the outliers influence the outcomes of the hypothesized model by studying the residuals. By inspecting the influential cases one can determine if certain cases exert undue influence over the parameters of the model.

Step 16: To evaluate if the model can be generalized, this involves checking assumptions and cross validation. If the assumptions of multiple regression are met: Normality of residuals, linearity, homoscedasticity, independence of error, equality of variance, autocorrelation and multi-collinearity, there is some good evidence to conclude that the model is generalizable.

Another approach to determine generalizability, is to cross validate (Field, 2013). In SPSS, one can get some statistics that give supports to generalization of model – adjusted R^2, and data splitting.

Step 17: Run scatter plots or partial plots to identify these outliers. Then run the model again with and without those outliers. Compare the R, R^2, beta to see if there are significant differences, If none, then the outliers can be kept as they do not have much influence on the outcomes.

Now check to see if most of the critical assumptions are met. Only when the assumptions are met can one be sure that the regression model identified is considered accurate and generalizable. If some of the critical assumptions are not met, do a transformation of the data set and rerun the procedure as described above till all critical assumptions are met.

If this transformed data set does not significantly contribute a higher variance to the model, keep the original model.

3.10 Procedure for testing moderation effect

Similarly, a moderator effect procedure was developed to answer research questions (v) and (vi) about the interaction effects of gender and language mastery.

3.10.1 General Guideline to assess a moderator effect in a causal relationship

Dawson (2014) described one approach to test for moderation effect using an Ordinary Least Square Regression model. Given the equation,

$$Y = \beta_1 + \beta_2 X + \beta_3 Z + \beta_4 XZ + \mathcal{E} \qquad \textbf{3.2}$$

where Y is the outcome, X the predictor, Z the moderator and XZ the interaction between X and Z. To test this two-way interaction, one only needs to check if the product effect i.e. XZ is statistically significant.

The following steps are recommended by Field (2013)

Step 1: Using a survey of the relevant literature, identify predictor (IV_1), the moderator known as IV_2, and of course the outcome variable (DV). Here the IVs can be discrete or continuous.

Step 2: Centered the IV but not the DV. Create a new variable to test the interaction effect by multiplying the selected centered IV with the centered moderator.

Step 3: Run the regression analysis again but this time with an added interaction term. Put in the centered IVs and centered moderator like normal and then put in the interaction variable in a separate block. If the *p*- value is less than .05 then there is a moderation effect.

This procedure can be translated into an easier format if the test of moderation is carried out using the SPSS software. These steps have been suggested by Wu and Zumbo (2008) after the data had been standardized and mean-centered.

3.11 Summary

This chapter described the methods, procedures and data analysis techniques designed to answer the primary research purpose i.e. to determine the relationships of selected cognitive determinants on statistical achievement as well as answering the proposed secondary objectives. It explained the rationale behind the choice of research design using a multivariate linear model. The research procedure includes a section on a pilot study to refine an adapted version of Statistical Reasoning Assessment Instrument for the main study and determine its internal consistency. A detailed account of how the equation modeling with SPSS is used as the main data analysis method for testing the different hypothesized models was described. Finally this chapter closed with a discussion on the procedure of statistical analyses of the data that are recommended to use to answer the objectives of this study.

CHAPTER 4 : RESULTS

4.1 Introduction

The main purpose of this study was to assess the relations between students' statistical achievement and cognitive determinants like prior mathematical knowledge, statistical reasoning, misconceptions concerning statistics and the influence of two other factors i.e. language mastery and gender on the reported relationships. To accomplish this task, a survey form was used to collect both primary and secondary data. The data analysis is aimed at gauging the students' competency in mathematics, reasoning and statistics achievement. These analyses were guided by four major research questions. This chapter is divided into five parts covering a section on descriptive analysis and four major sections that will answer the objectives of this study.

1) Descriptive analysis
2) The relationships between statistical achievement and the predictors (i.e. prior mathematical knowledge, statistical reasoning and statistical misconception)
3) The effect of gender and language mastery on the relationships in objective (2)
4) The relationships between statistical reasoning and the predictors (i.e. prior mathematical knowledge, statistical misconception)
5) The influence of gender and language mastery on the relationships in objective (4).

4.2 Descriptive Analysis

4.2.1 Description of Sample and Population

The respondents were sourced from a Malaysian public institution of higher learning.

The sample for this investigation comprises initially of 381 Diploma Science students enrolled in an introductory statistics course that comes from a total of N=900 students. They took different science programs in the Faculty of Applied Sciences. Students took this course in their fourth semester. The course is worth 3 credit hours. Statistics classes were conducted for 14 weeks where they are taught statistics for 4 hours each week. After cleaning the data, the sample was reduced to 374 usable cases. The gender composition of the sample comprises of male 20.6% and female 79.4%. An obvious disparity is the gender distribution where the majority consisted of female.

The students were all indigenous students (*Bumiputeras)* where the mother tongue was the Malay language. In the university where the current study was carried out, the students were instructed in English for all their core courses. Generally students' English Language mastery was considered good with 62.8% of the sample scoring good grades while 26.2% getting decent grades (see Table 4.1 for details).

Table 4.1: Language Mastery Distribution of Sample

		English Language			
	Aggregated score*	**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Weak	≤ 2 00	4	11.0	11.0	11.0
Average	Between 2.00 and 3.00	98	26.2	26.2	37.2
Good	≥ 3.00	235	62.8	62.8	100.0
Total		374	100.0	100.0	

*Method of aggregated score calculation is shown in Appendix C

4.2.2 Descriptive results of cognitive variables

Table 4.2 shows the mean, median and the dispersion of scores for the variables, statistical achievement (SA) – prior mathematical knowledge (PMK), statistical reasoning (SR) and misconception (MC).

Table 4.2: *Aggregated scores for independent and dependent variables

		Prior Mathematical Knowledge*	Statistical Achievement*	Statistical Reasoning*	Misconception*
N	Valid	374	374	374	374
	Missing	0	0	0	0
Mean		78.54	64.63	38.17	34.44
Median		79.75	70.80	37.20	34.70
Mode		70.00	75.00	33.90	34.00
Std. Deviation		11.72	24.78	13.83	11.56
95% CI		[77.35,79.74]	[62.11,67.15]	[36.76,39.57]	[33.27,35.62]
Skewness		-.16	-.67	.27	-.13
Std. Error of Skewness		.13	.13	.13	.13
Kurtosis		-.73	-.31	-.15	.20
Std. Error of Kurtosis		.25	.25	.25	.25

*Methods of aggregated score calculation are shown in Appendix C

As seen from Table 4.2, prior mathematical knowledge (PMK) and statistical achievement (SA) scores were high compared to the other two response variables. At a glance, the students showed quite good mastery of prior mathematical knowledge at the time of the study and their mean statistical achievement measured at the end of study was well above average. The respondents could only garner an average of 38.17 in Statistical Reasoning (SR) and a reasonably high level of Misconception (MC) about statistics (34.44). The low scores for both SR and MC are not surprising as the trend is almost similar in other studies in Malaysia or other parts of the world (Garfield, 2003; Tempelaar, 2006; Zuraida et al, 2012).

4.2.3 Correlational analysis of variables of interest

Before the onset of the regression analysis, a correlation matrix was generated to gauge the strength of the relationships among these variables.

4.2.3.1 Pearson's correlation coefficient

Pearson's correlation requires that data are interval for it to be an accurate measure of the linear relationship between variables. Univariate distributions of the variables under investigation have been found to be normally distributed. The acceptable range for skewness or kurtosis below +1.5 and above -1.5 (Tabachnick & Fidell, 2001). The skewness and kurtosis of all variables range from -0.75 to +0.75 (see Table 4.2). This analysis helps in determining the univariate normality of the variables.

Table 4.3: Analysis of Correlation Matrix

		Statistical Achievement	**Prior Mathematical Knowledge**	**Statistical Reasoning**	**Misconception**	**English Language**
Statistical Achievement	Pearson Correlation	1	.277**	.156**	-.122*	.048
	Sig. (2-tailed)		.000	.002	.019	.355
Prior Mathematical Knowledge	Pearson Correlation	.277**	1	.019	-.025	-.050
	Sig. (2-tailed)	.000		.713	.625	.332
Statistical Reasoning	Pearson Correlation	.156**	.019	1	-.525**	.270**
	Sig. (2-tailed)	.002	.713		.000	.000
Misconception	Pearson Correlation	-.122*	-.025	-.525**	1	-.170**
	Sig. (2-tailed)	.019	.625	.000		.001
English Language	Pearson Correlation	.048	-.050	.270**	-.170**	1
	Sig. (2-tailed)	.355	.332	.000	.001	

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).
N.B. Gender has been deleted from this analysis as it is a dichotomous variable

As seen in Table 4.3 there is a significant correlation between Statistical Achievement (SA) and Prior Mathematical knowledge (PMK) at r=.277, $p < .001$. Achievement also correlates with Statistical Reasoning (SR) r=.156, p=.002 though not as strong as PMK. Similarly SA correlates with Misconception (MC) at r= -.122, p=.019. However SA is not correlated to Language Mastery (EN) where r=.048, p=.355.

SR shows significant relationship with SA as stated earlier. Apart from that, it also correlates negatively and quite strongly with MC (r= -.515, $p < .001$). A negative correlation index indicates an inverse relationship between two variables. In this case, those with high reasoning skills will have lower misconception in statistics. Conversely if a student achieves low reasoning score then he/she is suspected to conceive high level of misconception as specified in the misconception table by Garfield (2003). In addition SR shows significant positive correlation with English Language (r=.270, p <.001).

On the other hand, it can be seen that MC correlates negatively with language mastery (ENG). One would suspect that a student who is good in language probably possesses less misconception about statistics.

4.3 Relationships of Students' statistical achievement with selected variables like reasoning, prior knowledge, misconception, language mastery and gender

The first two research question in this investigation pertained to the structure and relationship of students' statistical achievement with selected variables. To address the second question, the best Multiple Linear Regression Model was hypothesized as:

$$Y_i = \beta_\circ + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \qquad \textbf{4.1}$$

where Y_i= statistical achievement (SA)

$X_{1=}$ prior mathematical knowledge (PMK)

$X_{2=}$ statistical reasoning (SR)

$X_{3=}$ statistical misconception (MC)

X_4 = English Language (ENG)

X_5 = Gender (GEN)

To check for the independent variables that contribute significantly to the variance of the model, a series of diagnostic tests are run. To start off the selection of independent variables to be substituted into the regression model, the correlation matrix was generated as given in Table 4.4.

4.3.1 Diagnostics on the Hypothesized Model

4.3.1.1 Checking for order of entry into the model using Partial Correlation Matrix Results

Table 4.4: Correlation Matrix

		Statistical Achievement	**Prior Mathematical Knowledge**	**Statistical Reasoning**	**Misconception**	**English Language**	**Gender**
Statistical Achievement	Pearson Correlation	1	.277**	.156**	-.122*	.048	-.005
	Sig. (2-tailed)		.000	.002	.019	.355	.926
	N	374	374	374	374	374	374
Prior Mathematical Knowledge	Pearson Correlation	.277**	1	.019	-.025	-.050	.157**
	Sig. (2-tailed)	.000		.713	.625	.332	.002
	N	374	374	374	374	374	374
Statistical Reasoning	Pearson Correlation	.156**	.019	1	-.525**	.270**	-.024
	Sig. (2-tailed)	.002	.713		.000	.000	.645
	N	374	374	374	374	374	374
Misconception	Pearson Correlation	-.122*	-.025	-.525**	1	-.170**	-.047
	Sig. (2-tailed)	.019	.625	.000		.001	.365
	N	374	374	374	374	374	374
English Language	Pearson Correlation	.048	-.050	.270**	-.170**	1	.064
	Sig. (2-tailed)	.355	.332	.000	.001		.219
	N	374	374	374	374	374	374
Gender	Pearson Correlation	-.005	.157**	-.024	-.047	.064	1
	Sig. (2-tailed)	.926	.002	.645	.365	.219	
	N	374	374	374	374	374	374

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

Table 4.4 shows that the independent variable, PMK has the highest correlation index with the dependent variable, SA (Pearson r = .277, p= 0.001).

Once PMK is identified as the first variable to enter the model in the Stepwise forward method, one must know the next variable to enter. This is done through the Partial correlation matrix approach as shown in Table 4.5.

Table 4.5: Correlation matrix controlling for Prior Mathematical Knowledge

Control Variables			Statistical Achievement	Statistical Reasoning	Misconception	English Language	Gender
Prior Mathematical Knowledge	Statistical Achievement	Correlation	1.000	.157	-.119	.065	-.051
		Significance (2-tailed)	.	.002	.021	.214	.327
		df	0	371	371	371	371
	Statistical Reasoning	Correlation	.157	1.000	-.525	.271	-.027
		Significance (2-tailed)	.002	.	.000	.000	.600
		df	371	0	371	371	371
	Misconception	Correlation	-.119	-.525	1.000	-.172	-.044
		Significance (2-tailed)	.021	.000	.	.001	.402
		df	371	371	0	371	371
	English Language	Correlation	.065	.271	-.172	1.000	.073
		Significance (2-tailed)	.214	.000	.001	.	.162
		df	371	371	371	0	371
	Gender	Correlation	-.051	-.027	-.044	.073	1.000
		Significance (2-tailed)	.327	.600	.402	.162	.
		df	371	371	371	371	0

The results of the SPSS output presented in Table 4.5 show that the strongest correlation is between SA and SR (Pearson r=.157, p=.002) after controlling for the earlier variable PMK. Partial correlation is actually the value of a correlation between two variables of interest after taking into account the influence of the third variable upon the correlation. Thus this is important for us to take out the influence of the third variable, PMK in this case.

In effect, the user now knows that the next variable to enter the model is SR after PMK.

To continue this process one goes on to generate other partial correlation matrices as given in Table 4.6-Table 4.9.

Table 4.6: Correlation matrix controlling for Prior Mathematical Knowledge

Correlations

Control Variables			**Statistical Achievement**	**Statistical Reasoning**	**Misconception**	**English Language**	**Gender**
Prior Mathematical Knowledge	Statistical Achievement	Correlation	1.000	.157	-.119	.065	-.051
		Significance (2-tailed)	.	.002	.021	.214	.327
		df	0	371	371	371	371
	Statistical Reasoning	Correlation	.157	1.000	-.525	.271	-.027
		Significance (2-tailed)	.002	.	.000	.000	.600
		df	371	0	371	371	371
	Misconception	Correlation	-.119	-.525	1.000	-.172	-.044
		Significance (2-tailed)	.021	.000	.	.001	.402
		df	371	371	0	371	371
	English Language	Correlation	.065	.271	-.172	1.000	.073
		Significance (2-tailed)	.214	.000	.001	.	.162
		df	371	371	371	0	371
	Gender	Correlation	-.051	-.027	-.044	.073	1.000
		Significance (2-tailed)	.327	.600	.402	.162	.
		df	371	371	371	371	0

Table 4.7: Correlation matrix controlling for PMK, SR and GEN

Correlations					
Control Variables			Statistical Achievement	Misconception	English Language
Prior Mathematical Knowledge & Statistical Reasoning & Gender	Statistical Achievement	Correlation	1.000	-.047	.027
		Significance (2-tailed)	.	.362	.602
		df	0	369	369
	Misconception	Correlation	-.047	1.000	-.031
		Significance (2-tailed)	.362	.	.556
		df	369	0	369
	English Language	Correlation	.027	-.031	1.000
		Significance (2-tailed)	.602	.556	.
		df	369	369	0

Table 4.8: Correlation matrix controlling for PMK, SR, GEN and MC

Correlations				
Control Variables			**Statistical Achievement**	**English Language**
Prior Mathematical Knowledge & Statistical Reasoning & Gender & Misconception	Statistical Achievement	Correlation	1.000	.026
		Significance (2-tailed)	.	.622
		df	0	368
	English Language	Correlation	.026	1.000
		Significance (2-tailed)	.622	.
		df	368	0

The findings, as shown in the Tables 4.7 and 4.8 show that the correlations for MC, ENG and GEN are not statistically significant. This can be taken to mean that they will not contribute any significant marginal variation to the model.

The Choice of Entry is based on the partial correlations of the variables. The strongest was for PMK as can be seen from Table 4.4, next was SR, Gender, Misconception and finally Language Mastery. (Please see Table 4.9)

Table 4.9: Order of entry into the regression model

Variables Entered/Removed[a]			
Model	**Variables Entered**	**Variables Removed**	**Method**
1	Prior Mathematical Knowledge[b]	.	Enter
2	Statistical Reasoning[b]	.	Enter
3	Gender[b]	.	Enter
4	Misconception[b]	.	Enter
5	Dummy variable for good[b]	.	Enter
6	Dummy variable for weak[b]	.	Enter

a. Dependent Variable: Statistical Achievement
b. All requested variables entered.

The next stage is to confirm the significance of these variables in the model.

Partial *F*-test statistics are utilized to determine the order of entry for the selected cognitive determinants. Basically this type of *F*-test is to confirm that a variable that is correlated to the dependent variable do contribute significantly to the total variance of the model given after having taken into account the contribution of variances of the other predictors already in the model. In other word, by studying how much variation the variable PMK explains when the other variables are already in the model, the selection of the variables can then be carried out. This is known as marginal contribution of a variable like PMK given that the variances of the other variables SR, MC, ENG, GEN are already taken into account. The generated output helps to determine if a marginal contribution is significant or not.

Tables 4.10, 4.11 and 4.12 show the results of those factors that significantly impact statistical achievement using the Stepwise estimation method. For a complete regression analysis of all the factors entered/removed/excluded from the model and the residual statistics, refer to Appendix E, F and G.

The prediction model contained only two of the five factors or determinants of statistical achievement. The ANOVA table (Table 4.12) showed that the model was statistically significant, $F_{2,371}$ = 20.536, p<.001 and accounted for approximately 10% of the variance of statistical achievement (R^2 = .100, Adjusted R^2 = .095) as indicated in the output from Table 4.10. Comparing the *R* squared and the Adjusted *R* squared, there is a shrinkage of .100-.095 = .005 or 0.5% which is comparatively small. This is taken to mean that the model is generalizable using this sample (Field, 2013). The effect size (*ES*) for multiple regression is given by $f^2 = R^2 / 1 - R^2$ (Cohen, 1992). This gives an *ES* = .11 which is a medium effect given the sample size is large (n = 374).

Statistical achievement was found to be primarily predicted by Prior Mathematical Knowledge (PMK) and Statistical Reasoning (SR). The unstandardized and standardized regression coefficients of these two variables and the squared semi-partial correlations are given in Table 4.11. Squared semi-partial correlation (sr^2) informs us of the unique variance explained by each of the variable. This index is calculated using the Part column under Correlations list of Table 4.11 for the variables concerned. sr^2 for PMK is given by (.274 x .274 = .075) while SR is calculated by using (.151 x .151 = .023). This is interpreted as PMK and SR uniquely accounted for roughly 7.5% and 2.3% respectively for the variance found in SA. The contributions toward the variance can also be verified by looking at the regression weights of the two variables. PMK provided a much bigger portion of the weightage in the model as compared to SR.

The rest of the factors that included gender, misconception and language mastery were dropped from the model as the contributions to the variance by these factors are minimal and insignificant (see Appendix F where the excluded variables are listed). Although these variables are not significant in this model, it may be significant if combined with a different set of IVs. A point to note is that a variable may possess a low weight in the model or may not contribute significantly to the prediction of the model, it must not be presumed that it is itself a poor predictor (Hair et al., 1999)

Table 4.10: Checking for the best model

Model Summary

Std. Error of the Estimate	**Model**	***R***	***R* Square**	**Adjusted *R* Square**	**Std. Error of the Estimate**	**Change Statistics**					**Durbin-Watson**
						***R* Square Change**	***F* Change**	***df1***	***df2***	**Sig. *F* Change**	
23.84017	1	.277[a]	.077	.074	23.84017	.077	31.006	1	372	.000	
23.57646	2	.316[b]	.100	.095	23.57646	.023	9.368	1	371	.002	1.912

a. Predictors: (Constant), Prior Mathematical Knowledge

b. Predictors: (Constant), Prior Mathematical Knowledge, Statistical Reasoning

c. Dependent Variable: Statistical Achievement

Table 4.10 informs that Prior Mathematical Knowledge and Statistical Reasoning are significant predictors of the outcome variable Statistical Achievement as represented by Model 2. The R square = .100 meaning the two predictors only explain 10% of the variance.

Table 4.11: Identifying the best regression model coefficients[a]

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	***t***	**Sig.**	**95.0% Confidence Interval for *B***		**Correlations**			
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	sr^2
1	(Constant)	18.582	8.362		2.222	.027	2.140	35.024				
	Prior Mathematical Knowledge	.586	.105	.277	5.568	.000	.379	.793	.277	.277	.277	.077
2	(Constant)	8.746	8.872		.986	.325	-8.699	26.191				
	Prior Mathematical Knowledge	.580	.104	.274	5.571	.000	.375	.785	.277	.278	.274	.075
	Statistical Reasoning	.270	.088	.151	3.061	.002	.097	.444	.156	.157	.151	.023

a. Dependent Variable: Statistical Achievement

b. Predictors: (Constant), Prior Mathematical Knowledge

c. Predictors: (Constant), Prior Mathematical Knowledge, Statistical Reasoning

Table 4.12: Significance of the regression model

ANOVA[a]

Model		Sum of Squares	*df*	Mean Square	*F*	Sig.
1	Regression	17622.173	1	17622.173	31.006	.000[b]
	Residual	211427.536	372	568.354		
	Total	229049.709	373			
2	Regression	22829.508	2	11414.754	20.536	.000[c]
	Residual	206220.200	371	555.850		
	Total	229049.709	373			

a. Dependent Variable: Statistical Achievement

b. Predictors: (Constant), Prior Mathematical Knowledge

c. Predictors: (Constant), Prior Mathematical Knowledge, Statistical Reasoning

Table 4.12 shows that the model is significant implying at least one of the variables significantly contributes to the model.

In essence, the model that is suggested here takes the form of:

$$\mathrm{Y} = B_{\circ} + B_1 x_1 + B_2 x_2 \qquad \textbf{4.2}$$

where Y= statistical achievement (SA)

x_1 = prior mathematical knowledge (PMK)

x_2= statistical reasoning (SR)

The final model is given by equation 4.33

$$\mathrm{SA} = 8.75 + .580\ (\mathrm{PMK}) + .270(\mathrm{SR}) \qquad \textbf{4.3}$$

The model tells us that for every increase of one unit of PMK, there is a corresponding increase of 0.580 unit in SA while increasing one unit of SR, sees an increase of 0.270 unit in SA.

The model shows the relationship of the predictors PMK and SR with the outcome variable, SA with PMK showing a stronger effect on SA than SR (See Table 4.11 for the results of the constant and unstandardized coefficients given in Equation 4.3). Looking at the standardized coefficients of .274 and .151 for PMK and SR respectively, it implies that the impact of PMK is roughly twice that of SR on SA. With a R square of .100 (see Table 4.10), the two predictors could only account for 10% of the variance. In conclusion, the model has answered the first research question that clearly identified PMK and SR on the cognitive determinants of SA.

4.3.2 Assumption checks for the Regression Model

This section runs tests to check the all assumptions of multiple regression modeling are fulfilled.

4.3.2.1 Assumption Checks on Normality of dataset

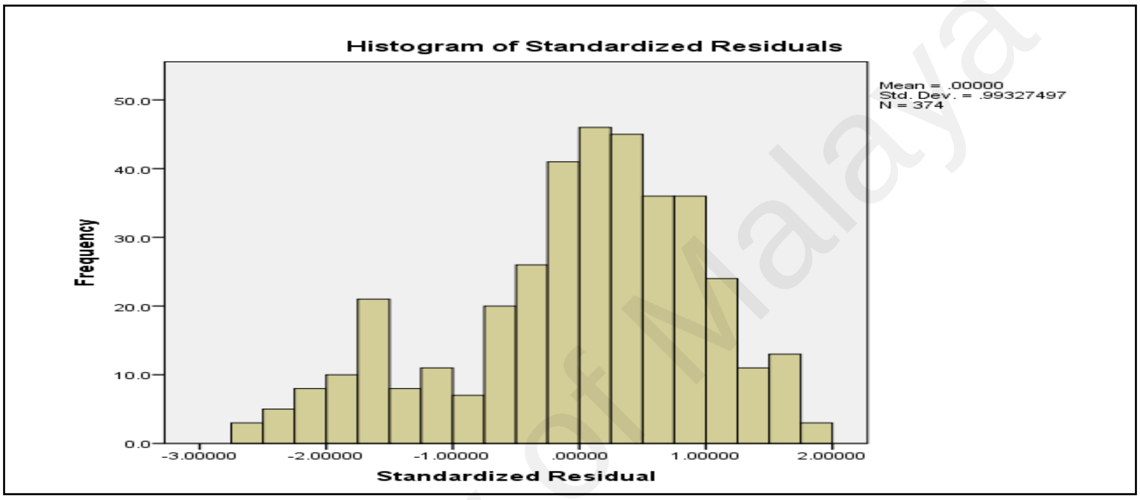

Figure 4.1: Residuals analysis on normality of dataset

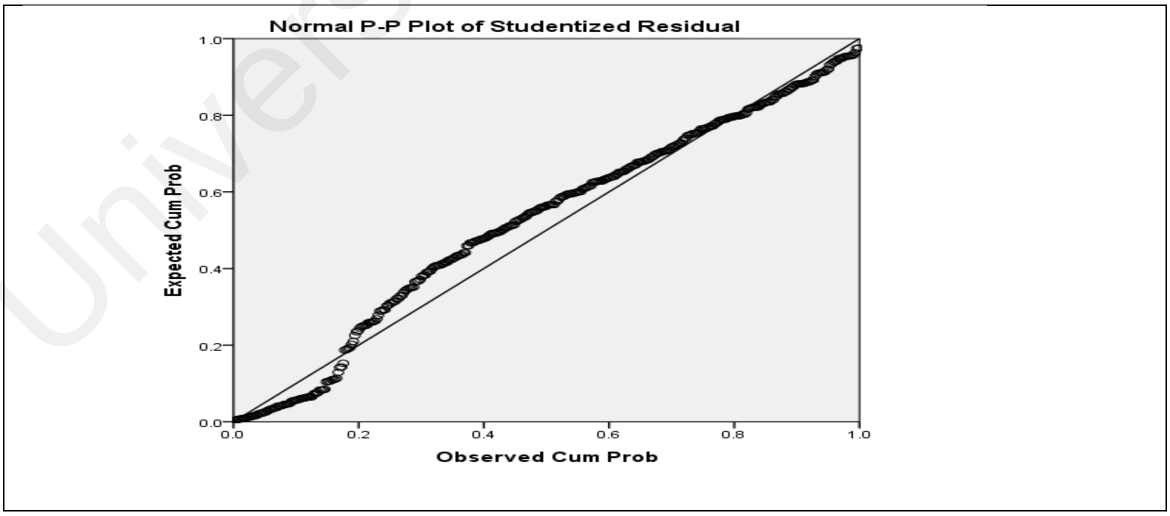

Figure 4.2: Normal P-P plot on normality of dataset

Figure 4.2 and Figure 4.3 show that the standardized residuals are approximately normal.

Table 4.13: Identifying the collinearity measures

Collinearity Diagnostics[a]

Model	Dimension	Eigenvalue	Condition Index	Variance Proportions (Constant)	Prior Mathematical Knowledge	Statistical Reasoning	Gender	Misconception	English Language
1	1	1.989	1.000	.01	.01				
	2	.011	13.492	.99	.99				
2	1	2.907	1.000	.00	.00	.01			
	2	.082	5.943	.03	.05	.94			
	3	.011	16.630	.97	.94	.04			
3	1	3.861	1.000	.00	.00	.01	.00		
	2	.097	6.306	.01	.01	.88	.08		
	3	.032	11.044	.05	.22	.06	.84		
	4	.010	19.664	.95	.77	.05	.07		
4	1	4.747	1.000	.00	.00	.00	.00	.00	
	2	.165	5.357	.00	.00	.27	.00	.21	
	3	.054	9.405	.00	.01	.30	.48	.33	
	4	.026	13.488	.02	.47	.20	.43	.21	
	5	.008	24.589	.98	.52	.22	.10	.26	
5	1	5.704	1.000	.00	.00	.00	.00	.00	.00
	2	.168	5.826	.00	.00	.23	.00	.21	.00
	3	.054	10.310	.00	.01	.28	.48	.32	.00
	4	.041	11.823	.00	.02	.21	.05	.03	.85
	5	.026	14.803	.02	.49	.16	.40	.19	.01
	6	.007	28.630	.98	.48	.12	.07	.24	.14

Dependent Variable: Statistical Achievement

4.3.2.2 Assumption Checks on Multicollinearity of dataset

The collinearity diagnostics like condition index and variance proportions indicate that variables investigated do not show multicollinearity (see Table 4.13).

Table 4.14: Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	40.6094	86.7603	64.6332	8.00263	374
Std. Predicted Value	-3.002	2.765	.000	1.000	374
Standard Error of Predicted Value	1.478	7.352	2.878	.815	374
Adjusted Predicted Value	39.1617	88.6228	64.6438	7.99551	374
Residual	-62.90804	46.17880	.00000	23.45277	374
Std. Residual	-2.664	1.956	.000	.993	374
Stud. Residual	-2.686	1.971	.000	1.001	374
Deleted Residual	-63.93390	46.87767	-.01068	23.81318	374
Stud. Deleted Residual	-2.709	1.978	-.001	1.003	374
Mahal. Distance	.464	35.163	4.987	3.634	374
Cook's Distance	.000	.025	.003	.004	374
Centered Leverage Value	.001	.094	.013	.010	374

a. Dependent Variable: Statistical Achievement

4.3.2.3 Checking for Outliers in the sample

There are various techniques of checking for multivariate outliers. One of more popular method is to use Mahalanobis Distance to identify outliers. The distances as given in Table 4.14 have a minimum of 0.464 and a maximum of 35.163 with a mean of 4.987 (*SD* = 3.634) where generally most of the data points are not less than 1.0. Data points less than 1.0 are considered outliers (Hair et al., 1999)

In addition, studentized deleted residuals do not show obvious outliers that need to pay attention to as the standard deviation is small (see Table 4.14). Figure 4.1 that illustrates the 3-D representation of the three variables, does not show extreme outliers that need to be taken into account in the analysis.

Figure 4.4 shows a scatterplot of zpred versus zresid to check for linearity, homoscedasticity and independent errors (Field, 2013). The random pattern of the points shows that the assumptions of linearity, homoscedasticity and independent errors are satisfied.

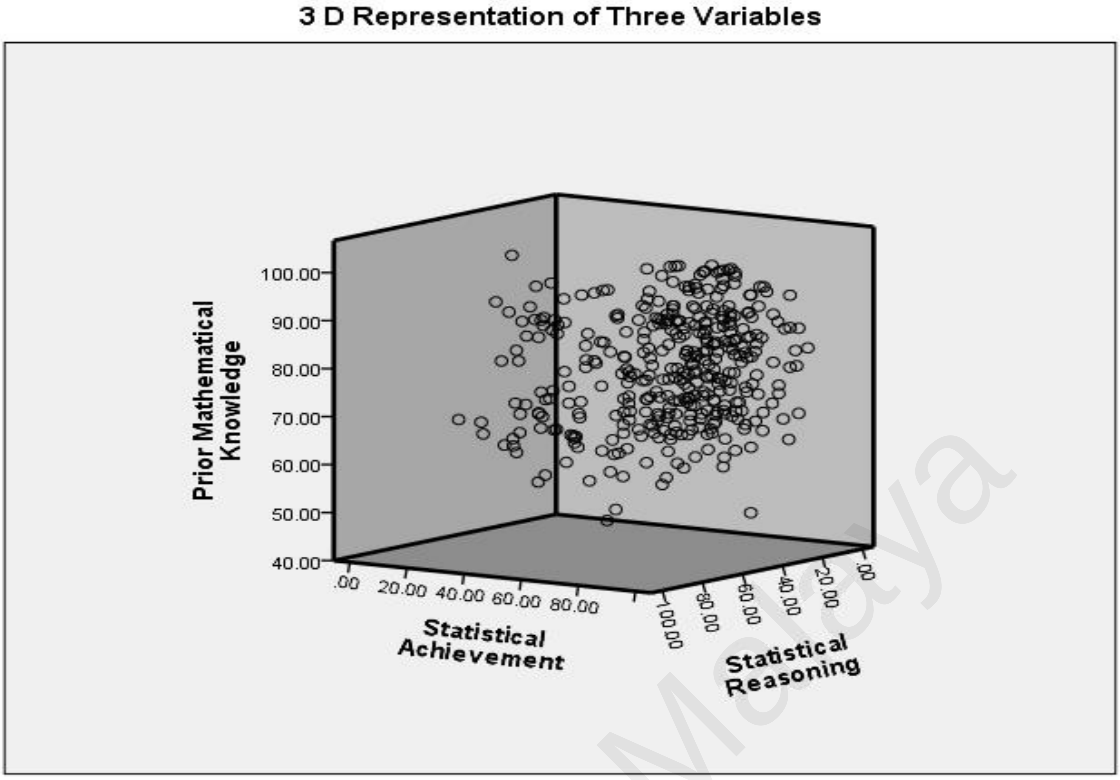

Figure 4.3: Data points distribution in 3D plot to identify outliers

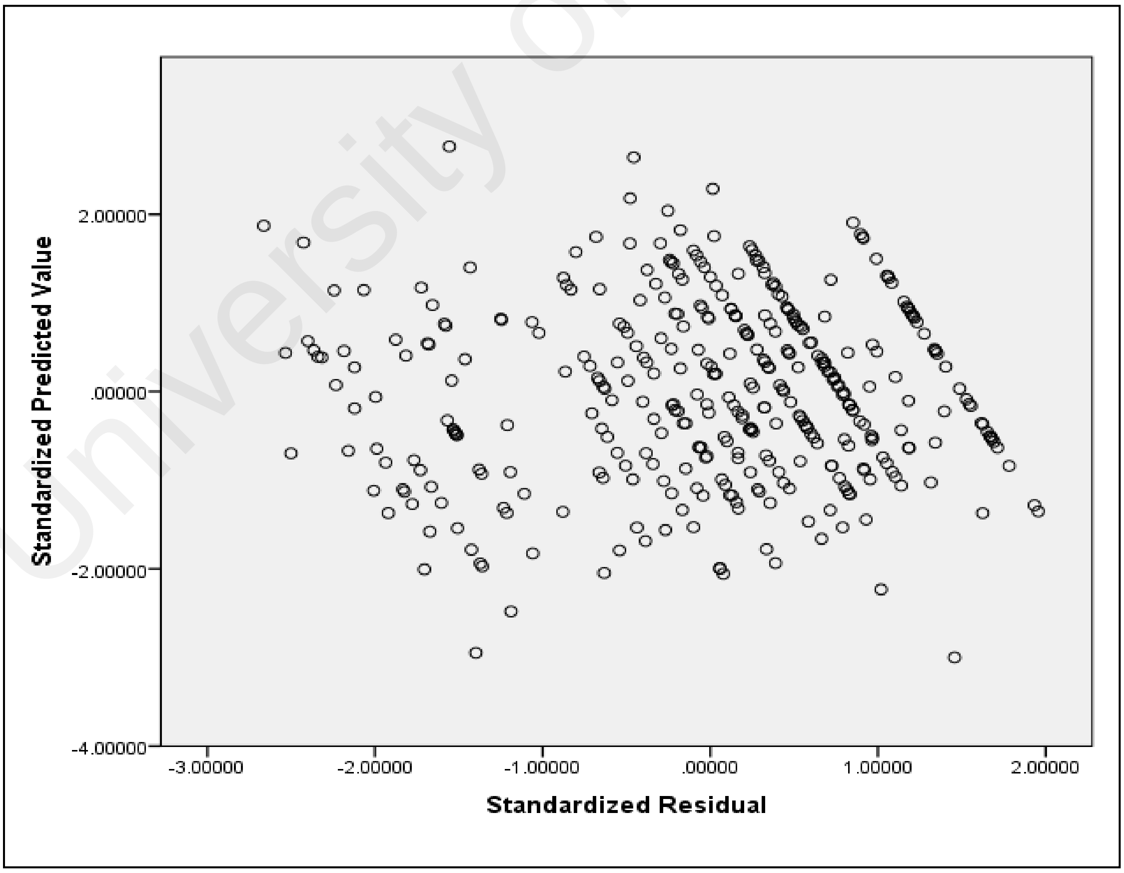

Figure 4.4: Scatterplot on zpred versus zresid to check for linearity, homoscedasticity and independence (Field, 2013)

Checking for Multicollinearity

The VIF and Tolerance Indices show no multicollinearity with VIF < 2.00. The Table 4.15 shows on the average, VIF is around 1.00.

Furthermore, correlation coefficients in Table 4.4 did not show strong correlations among all the variables proving further indication of no multicollinearity effect.

According to StatPac (2010) manual, multicollinearity can also be assessed by generating the collinearity diagnostics as shown in Table 4.13 & Table 4.15. None of the condition indices were between 30–100 and the variance proportion rows do not indicate any variable with more than 2 numbers over 0.5.

4.3.3 Best Model for the regression analysis

In conclusion, the general model takes the form of:

$$\mathrm{Y} = B_{\circ} + B_1 x_1 + B_2 x_2 \qquad \textbf{4.4}$$

where Y= statistical achievement (SA)

x_1 = prior mathematical knowledge (PMK)

x_2= statistical reasoning (SR)

The final model is given by equation 4.5

$$\mathrm{SA} = 8.75 + .58\,(\mathrm{PMK}) + .27(\mathrm{SR}) \qquad \textbf{4.5}$$

Table 4.15: Tolerance and VIF indices for checking multicollinearity

Excluded Variables[a]

Model		Beta In	*t*	Sig.	Partial Correlation	Collinearity Statistics Tolerance	VIF	Minimum Tolerance
1	Statistical Reasoning	.151[b]	3.061	.002	.157	1.000	1.000	1.000
	Gender	-.049[b]	-.981	.327	-.051	.976	1.025	.976
	Dummy variable Gender	.049[b]	.981	.327	.051	.976	1.025	.976
	Misconception	-.115[b]	-2.316	.021	-.119	.999	1.001	.999
	Language transf	.075[b]	1.512	.131	.078	.999	1.001	.999
	Dummy variable for weak	-.057[b]	-1.135	.257	-.059	.995	1.005	.995
	dummy variable for good	.071[b]	1.413	.159	.073	.993	1.007	.993
2	Gender	-.045[c]	-.909	.364	-.047	.975	1.026	.975
	Dummy variable Gender	.045[c]	.909	.364	.047	.975	1.026	.975
	Misconception	-.049[c]	-.849	.397	-.044	.724	1.381	.724
	Language transf	.038[c]	.746	.456	.039	.930	1.075	.930
	Dummy variable for weak	-.028[c]	-.548	.584	-.028	.956	1.046	.956
	dummy variable for good	.036[c]	.697	.486	.036	.934	1.071	.934
3	Dummy variable Gender	.[d]	.	.	.	.000	.	.000
	Misconception	-.053[d]	-.912	.362	-.047	.721	1.387	.721
	Language transf	.044[d]	.854	.394	.044	.918	1.089	.918
	Dummy variable for weak	-.033[d]	-.658	.511	-.034	.943	1.061	.943
	dummy variable for good	.040[d]	.775	.439	.040	.927	1.079	.927
4	Dummy variable Gender	.[e]	.	.	.	.000	.	.000
	Language transf	.044[e]	.855	.393	.045	.918	1.089	.684
	Dummy variable for weak	-.033[e]	-.651	.516	-.034	.943	1.061	.701
	dummy variable for good	.040[e]	.781	.436	.041	.927	1.079	.688
5	Dummy variable Gender	.[f]	.	.	.	.000	.	.000
	Dummy variable for weak	.001[f]	.009	.993	.000	.385	2.595	.375
	dummy variable for good	.001[f]	.009	.993	.000	.161	6.208	.160
6	Dummy variable Gender	.[g]	.	.	.	.000	.	.000
	dummy variable for good	.[g]	.	.	.	.000	.	.000

. Dependent Variable: Statistical Achievement

b. Predictors in the Model: (Constant), Prior Mathematical Knowledge

c. Predictors in the Model: (Constant), Prior Mathematical Knowledge, Statistical Reasoning

d. Predictors in the Model: (Constant), Prior Mathematical Knowledge, Statistical Reasoning, Gender

e. Predictors in the Model: (Constant), Prior Mathematical Knowledge, Statistical Reasoning, Gender, Misconception

f. Predictors in the Model: (Constant), Prior Mathematical Knowledge, Statistical Reasoning, Gender, Misconception, Dummy variable for good

4.4 Moderating effect of language mastery and gender on the relationships between statistical achievement and the predictors

The next section deals with the question of moderation by certain qualitative or quantitative variables. This research only deals with two variables i.e. language mastery and gender. The moderation analysis follows this procedure:

Analyze>descriptives>save as standardized values (select the independent and moderating variable). Transform>compute (calculate the product of the 2 standardized variables). Analyze > regression > linear (select the dependent variable, insert the independent and moderating variable, click next, and add the product. Is the p value of the product or interaction significant? If yes, there is moderation.

4.4.1 The moderating effect of the variables language mastery (ENG) and gender (GEN) on the relationships of the other response variables like SA, SR, PMK and MC

The last two research questions seek to identify if ENG and GEN have indirect effect on the relationships formed among the variables SA, SR, PMK and MC. The first research question has found that only SR and PMK have significant effect on SA. Thus the moderation analysis is done based on this fact:

4.4.1.1 Does English language mastery moderate the influence of statistical reasoning on statistical achievement?

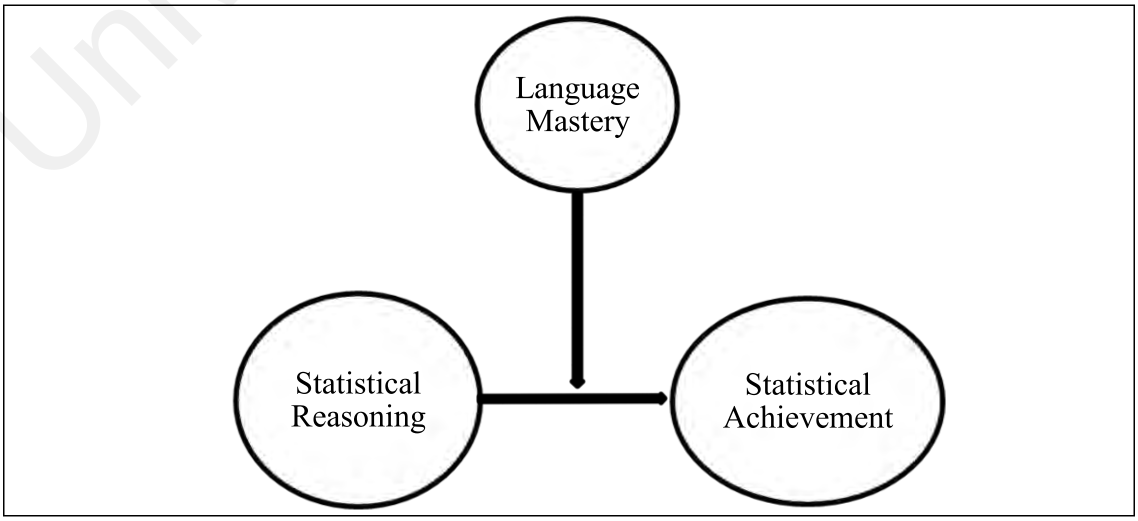

Figure 4.5: Moderating effect of ENG on the relationship between SR and SA

Regression analysis for SA, SR and ENG

To confirm the moderating effect of ENG, the procedure explained in chapter 3 will be used to study this effect as portrayed in Figure 4.5.

Below is the analysis as outlined by the procedure.

Table 4.16: Influence of ENG on SR and SA

Model Summary

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	*F* Change	*df1*	*df2*	Sig. *F* Change
1	.156[a]	.024	.017	24.57515	.024	3.087	3	370	.027

a. Predictors: (Constant), zSR_zENG, Zscore: Statistical Reasoning, Zscore: English Language
b. Dependent Variable: SA

Table 4.17: Regression Coefficients

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	*t*	Sig.
		B	Std. Error	Beta		
1	(Constant)	64.670	1.322		48.928	.000
	Zscore: Statistical Reasoning	3.839	1.329	.155	2.889	.004
	Zscore: English Language	.127	1.353	.005	.094	.925
	zSR_zENG	-.138	1.351	-.005	-.102	.919

a. Dependent Variable: SA

Figure 4.5 represents a multiple regression model that has been designed to investigate whether the association between SA and SR depends on Language mastery (ENG). After centering SA and SR and computing the zSR_zENG interaction term (Dawson, 2014), the two predictors and the interaction were entered into a simultaneous regression model. Results given in Table 4.16 and Table 4.17 indicate that SR (b = 3.839, SE_b = 1.329, β = .155, p = .004) was associated with SA but ENG (b = .127, SE_b = 1.353, β = .005, p = .925) was not. In addition the interaction between SR and ENG was not significant (b = -.138, SE_b = 1.351, β =. -.005, p = .919), suggesting that SR does not depend on ENG.

As such it confirms that gender does not act as a moderator in the relationship between SA and SR.

4.4.1.2 Does English language mastery moderate the influence of prior mathematical knowledge on statistical achievement?

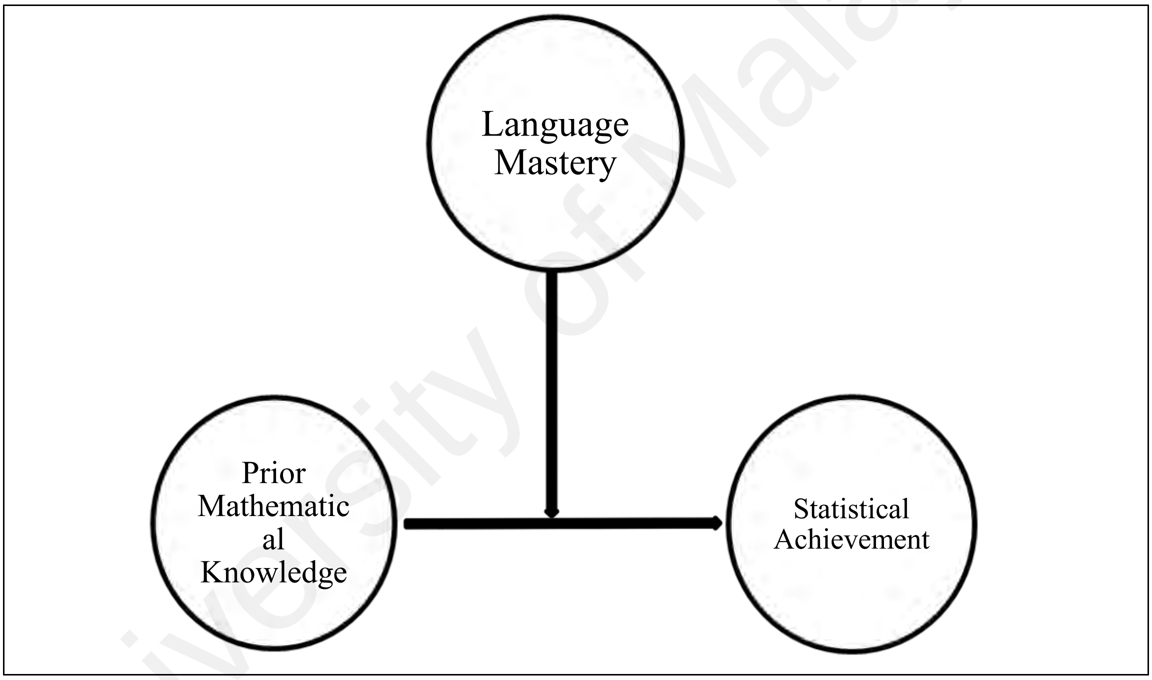

Figure 4.6: Moderating effect of ENG on the relationship between PMK and SA

Regression analysis for SA, PMK and ENG

Table 4.18: Influence of ENG on PMK and SA

Model Summary

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate	Change Statistics: *R* Square Change	*F* Change	*df1*	*df2*	Sig. *F* Change
1	.297[a]	.088	.081	23.75935	.088	11.917	3	370	.000

a. Predictors: (Constant), zPMK_zENG, Zscore: English Language, Zscore: Prior Mathematical Knowledge

b. Dependent Variable: SA

Table 4.19: Regression Coefficients

Coefficients[a]

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	***t***	**Sig.**
		B	Std. Error	Beta		
1	(Constant)	64.530	1.230		52.462	.000
	Zscore: English Language	1.422	1.234	.057	1.153	.250
	Zscore: Prior Mathematical Knowledge	6.683	1.242	.270	5.382	.000
	zPMK_zENG	-2.064	1.196	-.086	-1.725	.085

a. Dependent Variable: SA

A multiple regression model (Figure 4.6) was tested to investigate whether the association between SA and PMK depends on Language mastery (ENG). After centering SA and PMK and computing the zPMK_zENG interaction term (Dawson, 2014), the two predictors and the interaction were entered into a simultaneous regression model. Results as seen in Table 4.19 indicate that PMK (b = 6.683, SE_b = 1.242, β = .270, p < .001) was associated with SA but ENG (b = 1.422, SE_b = 1.234, β = .057, p = .250) was not. In addition the interaction between PMK and ENG was not significant (b = -2.064, SE_b = 1.196, β =-.086, p = .085), suggesting that PMK does not depend on ENG.

As such it confirms that ENG does not act as a moderator in the relationship between SA and PMK.

4.4.1.3 Does gender moderate the influence of statistical reasoning on statistical achievement?

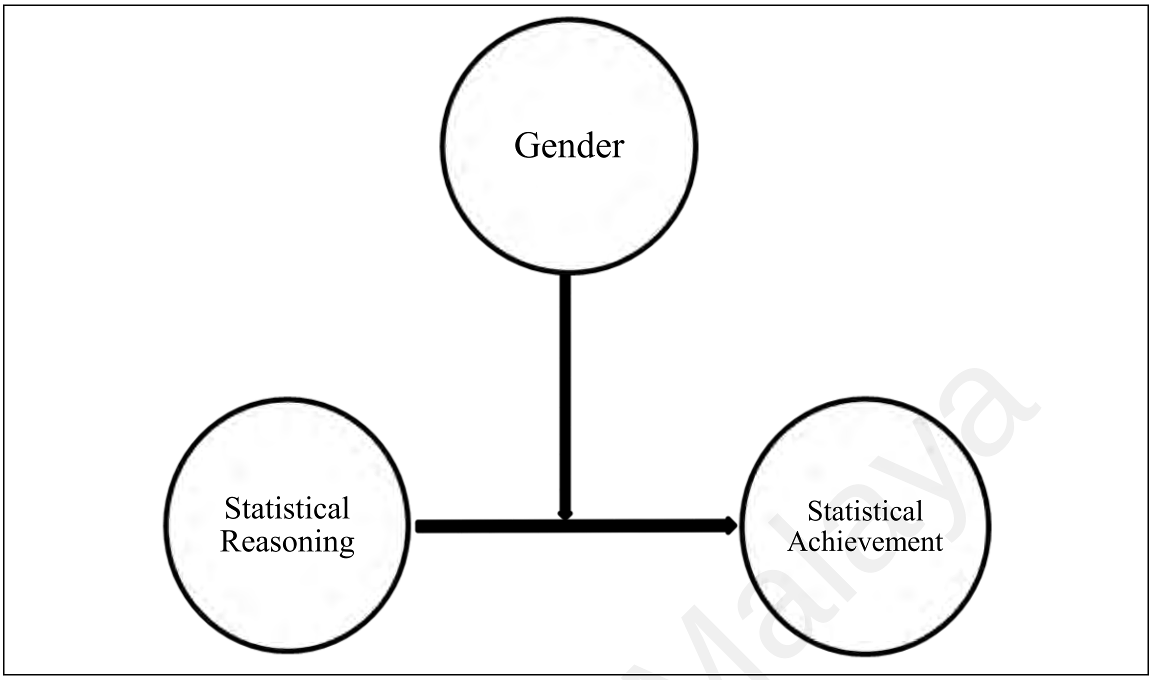

Figure 4.7: Moderating effect of ENG on the relationship between SR and SA

Regression analysis for SA, SR and GENDER

Table 4.20: Influence of GEN on SR and SA

Model Summary

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	*F* Change	*df1*	*df2*	Sig. *F* Change
1	.171[a]	.029	.021	24.51346	.029	3.724	3	370	.012

a. Predictors: (Constant), zSR_zGEN, Zscore: Gender, Zscore: Statistical Reasoning
a. Dependent Variable: SA

Table 4.21: Regression Coefficients

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	*t*	Sig.
		B	Std. Error	Beta		
1	(Constant)	64.593	1.268		50.945	.000
	Zscore: Statistical Reasoning	3.754	1.272	.151	2.951	.003
	Zscore: Gender	.028	1.270	.001	.022	.983
	zSR_zGEN	-1.671	1.216	-.071	-1.374	.170

a. Dependent Variable: SA

Figure 4.7 represents a multiple regression model designed to investigate whether the association between SA and SR depends on Gender (GEN). After centering SA and SR and computing the zSR_zGEN interaction term (Dawson, 2014), the two predictors and the interaction were entered into a simultaneous regression model. Results given in Table 4.21 show that SR (b = 3.754, SE_b = 1.272, β = .151, p = .003) was associated with SA but GEN (b = .028, SE_b = 1.270, β = .001, p = .983) was not. In addition the interaction between SR and GEN was not significant (b = -1.671, SE_b = 1.216, β = -.071, p = .170), suggesting that SR does not depend on GEN.

As such it confirms that GEN does not act as a moderator in the relationship between SA and SR.

4.4.1.4 Does gender moderate the influence of prior mathematical knowledge on statistical achievement?

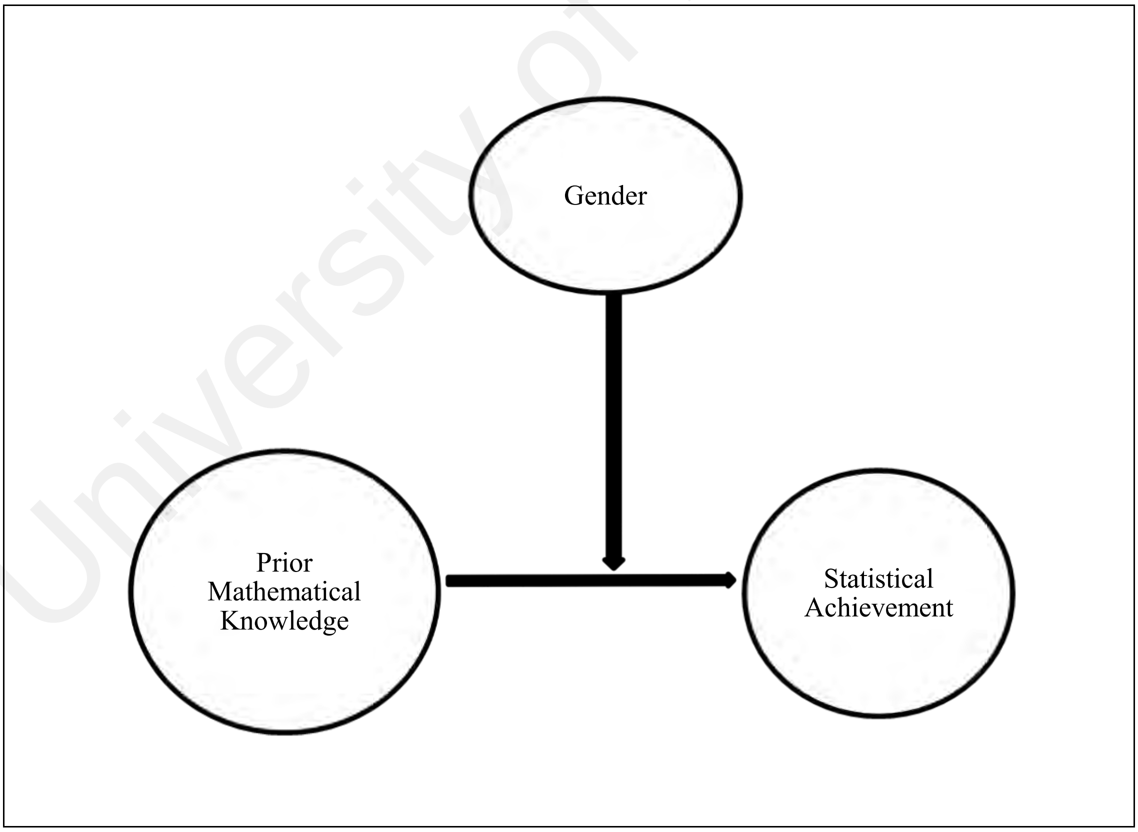

Figure 4.8: Moderating effect of GEN on the relationship between PMK and SA

Regression analysis for SA, PMK and GENDER

Table 4.22: Influence of GEN on PMK and SA

Model Summary

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	*F* Change	*df1*	*df2*	Sig. *F* Change
1	.289[a]	.083	.076	23.82238	.083	11.203	3	370	.000

a. Predictors: (Constant), zPMK_zGEN, Zscore: Prior Mathematical Knowledge, Zscore: Gender
b. Dependent Variable: SA

Table 4.23: Regression Coefficients

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	*t*	Sig.
		B	Std. Error	Beta		
1	(Constant)	64.877	1.247		52.031	.000
	Zscore: Prior Mathematical Knowledge	7.044	1.249	.284	5.640	.000
	Zscore: Gender	-1.578	1.280	-.064	-1.233	.218
	zPMK_zGEN	-1.562	1.238	-.064	-1.262	.208

a. Dependent Variable: SA

A multiple regression model (Figure 4.8) was tested to investigate whether the association between SA and PMK depends on Gender (GEN). After centering SA and SR and computing the zPMK_zGEN interaction term (Dawson, 2014), the two predictors and the interaction were entered into a simultaneous regression model. Results shown in Table 4.23 indicate that PMK (b = 7.044, SE_b = 1.249, β = .284, p < .001) was associated with SA but GEN (b = -1.578, SE_b = 1.280, β = -.064, p = .218) was not. In addition the interaction between PMK and GEN was not significant (b = -1.562, SE_b = 1.238, β = -.064, p = .208), suggesting that PMK does not depend on GEN.

As such it confirms that GEN does not act as a moderator in the relationship between SA and PMK.

4.5 Relationships of Students' statistical reasoning with selected variables like prior knowledge, misconception, language mastery and gender

The third and fourth research questions in this investigation pertained to the structure and relationship of students' statistical reasoning with selected variables. To address the fourth question, the best Multiple Linear Regression Model was hypothesized as:

$$\mathrm{Y}_i = \beta_\circ + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \qquad \textbf{4.6}$$

where Y_i= statistical reasoning (SA)

$x_{1=}$ prior mathematical knowledge (PMK)

$x_{2=}$ statistical achievement (SR)

$x_{3=}$ statistical misconception (MC)

x_4 = English Language (ENG)

x_5 = Gender (GEN)

The procedure for selecting of order of entry is the same as that of the previous Multiple Linear Regression on Statistical Achievement. Results of the analysis on statistical reasoning are discussed next.

The first step in the procedure is to study the correlation matrix generated. Notice from the correlation table (Table 4.24), the independent variable, Misconception (MC) has the highest correlation index with the dependent variable, Statistical Reasoning (SR) (Pearson r = -.525, p< 0.001) with English Language (ENG) (Pearson r = .270, p< 0.001) and Statistical Achievement (SA) (Pearson r = .156, p= 0.002) following suit.

Once MC is identified as the first variable to enter the model in the Stepwise forward method, one needs to know the next variable to enter. This is done through the Partial correlation matrix approach. Based on the results of the correlation matrix (Table 4.24), probable factors that are significant to the model are misconception, language mastery, and statistical achievement. This has been shown to be true from Table 4.25.

Using partial F statistics, the order of entry has been identified as in Table 4.25

Table 4.24: Order of Entry of variables

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Misconception	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	English Language	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	Statistical Achievement	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Statistical Reasoning

Table 4.26 and 4.27 show the results of those factors that significantly impact statistical reasoning using the Stepwise estimation method. For a complete regression analysis of all the factors excluded from the model and the residual statistics, refer to Appendix H and I.

Table 4.26 summarized the variances as represented by *R* Square and Adjusted *R* Square. Three models are generated as additional variable is added to the analysis in a stepwise manner. *R*-square is computed to measure the amount of the variation in the DV explained by the IV for a linear regression model while adjusted *R*-square although serves the same function but make adjustments to the statistic after taking into account the number of independent variables entered into the model and the strength of the correlation values. *R* square change is a measure of the difference between the *R* square if the first model and that of the second model.

Table 4.25: Correlation Matrix for the selected factors

Correlations

		English Language	**Gender**	**Prior Mathematical Knowledge**	**Statistical Achievement**	**Statistical Reasoning**	**Misconception**
English Language	Pearson Correlation	1	.064	-.050	.048	.270**	-.170**
	Sig. (2-tailed)		.219	.332	.355	.000	.001
	N	374	374	374	374	374	374
Gender	Pearson Correlation	.064	1	.157**	-.005	-.024	-.047
	Sig. (2-tailed)	.219		.002	.926	.645	.365
	N	374	374	374	374	374	374
Prior Mathematical Knowledge	Pearson Correlation	-.050	.157**	1	.277**	.019	-.025
	Sig. (2-tailed)	.332	.002		.000	.713	.625
	N	374	374	374	374	374	374
Statistical Achievement	Pearson Correlation	.048	-.005	.277**	1	.156**	-.122*
	Sig. (2-tailed)	.355	.926	.000		.002	.019
	N	374	374	374	374	374	374
Statistical Reasoning	Pearson Correlation	.270**	-.024	.019	.156**	1	-.525**
	Sig. (2-tailed)	.000	.645	.713	.002		.000
	N	374	374	374	374	374	374
Misconception	Pearson Correlation	-.170**	-.047	-.025	-.122*	-.525**	1
	Sig. (2-tailed)	.001	.365	.625	.019	.000	
	N	374	374	374	374	374	374

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 4.26: Summary statistics

Model Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.525[a]	.276	.274	11.78640	.276	141.471	1	372	.000
2	.556[b]	.309	.305	11.52620	.033	17.985	1	371	.000
3	.563[c]	.317	.311	11.47721	.008	4.174	1	370	.042

a. Predictors: (Constant), Misconception

b. Predictors: (Constant), Misconception, English Language

c. Predictors: (Constant), Misconception, English Language, Statistical Achievement

d. Dependent Variable: Statistical Reasoning

The model summary indicates that R-square is .317. This indicates that 31.7% of the variance in statistical reasoning can be explained by sum of all the factors above. However the contributions to the variance by some of these factors are minimal and insignificant. Comparing the *R* square and the Adjusted *R* square, there is a shrinkage of .317-.309 = .008 or 2.52% which is rather small. This is taken to mean that the model is generalizable using this sample.

The prediction model contained only three of the five factors affecting statistical reasoning. The ANOVA table (Appendix H) showed that the model was statistically significant, $F_{3,370}$ = 57.169, p<.001 and accounted for approximately 31% of the variance of statistical reasoning (R^2 = .317, Adjusted R^2 = .311) as indicated in the output from Table 4.26. Comparing the R square and the Adjusted R square, there is a shrinkage of .317-.311 = .006 or 0.6% which is comparatively small. This is taken to mean that the model is generalizable using this sample. The effect size (*ES*) for multiple regression is given by $f^2 = R^2/ 1- R^2$ (Cohen, 1992). This gives an *ES* = .46 which is a large effect.

Statistical reasoning was found to be primarily predicted by Misconception (MC), English Language (ENG) and Statistical Achievement (SA). The unstandardized and standardized regression coefficients of these two variables and the squared semi-partial correlations are given in Table 4.27. Squared semi-partial correlation (sr^2) informs that the unique variance explained by each of the variable. This index is calculated using the Part column under Correlations list of Table 4.27 for the variables concerned. sr^2 for MC is given by (-.473 x -.473 = .224) while ENG is calculated by using (.181 x .181 = .033) and SA is (.088 x .088 = .008). This is interpreted as MC, ENG and SA uniquely accounted for roughly 22.4%, 3.3% and .8% respectively for the variance of SR. MC has the greatest effect on SR while ENG was essentially moderate and SA has small but significant effect. These results can also be verified by looking at the regression weights of the three variables. MC provided a much bigger portion of the weightage in the model as compared to ENG and SA (-.483 for MC while ENG and SA are merely .183 and .088 respectively). These values can be found from Table 4.27 under the Standardized Coefficients column.

The rest of the factors that included gender and Prior Mathematical Knowledge were dropped from the model as the contributions to the variance by these factors are minimal and insignificant (see Appendix I where the excluded variables are listed). Although these variables are not significant in this model, it may be significant if combined with a different set of IVs. (Hair et al., 1999).

Table 4.27 shows that only misconception, language mastery, and statistical achievement have significant influence on statistical reasoning.

Table 4.27: Coefficients of the regression model

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	59.792	1.918		31.181	.000	56.021	63.562
	Misconception	-.628	.053	-.525	-11.894	.000	-.732	-.524
2	(Constant)	47.072	3.537		13.308	.000	40.117	54.028
	Misconception	-.590	.052	-.493	-11.263	.000	-.693	-.487
	English Language	3.497	.825	.186	4.241	.000	1.876	5.119
3	(Constant)	43.607	3.909		11.155	.000	35.920	51.294
	Misconception	-.578	.053	-.483	-10.999	.000	-.681	-.474
	English Language	3.451	.822	.183	4.200	.000	1.835	5.066
	Statistical Achievement	.049	.024	.088	2.043	.042	.002	.097

a. Dependent Variable: Statistical Reasoning

The hypothesized model suggests:

$$\mathrm{Y_i} = \beta_\circ + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 \qquad \textbf{4.7}$$

where Y_i= statistical reasoning (SR)

x_1= prior mathematical knowledge (PMK)

x_2= statistical achievement (SA)

x_3= statistical misconception (MC)

x_4 = English Language (ENG)

x_5 = Gender (GEN)

The best model is:

$$Y = 43.61 + 0.05x_2 - 0.58x_3 + 3.45x_4 \quad \textbf{4.8}$$

In physical unit, for every increase of one unit of SA, there is an increase of only 0.05 unit of SR while an increase of one unit of MC sees a decrease of 0.58 unit of SR. The greatest effect can be seen from ENG. For an increase of one unit of ENG, there is a corresponding increase of 3.45 units of SR.

Based on this model, only SA, MC and ENG were significant cognitive determinants affecting SR, thus successfully answered the third research question.

The ability of the students in reasoning very much depends on the level of misconception and their language mastery over other factors. This is logical as reasoning requires a good degree of understanding of the grammatical structure of the items and the technical terms involved. It should be noted that the SRA items are long and contains underlying concepts that can only be explicated by reading the questions carefully and attentively. It can be seen the regression coefficients for misconception variable are negative signalling an inverse relationship between SR and MC. Students with high level of misconceptions and low degree of language mastery in English generally fare badly in the statistical reasoning ability as measured using the SRA instrument. Though statistical achievement has some positive influence, it is rather small as compared to the other two variables.

The speed of the students answering the items in SRA seems to indicate that the majority took less than an hour to finish the questions whereby the administration of this instrument did not specify a timed prerequisite.

SRA has an intrinsic weakness as an instrument to measure the students' reasoning skill as it is dependent on student's mastery in the language.

4.5.1 Assumption checks for Regression Model

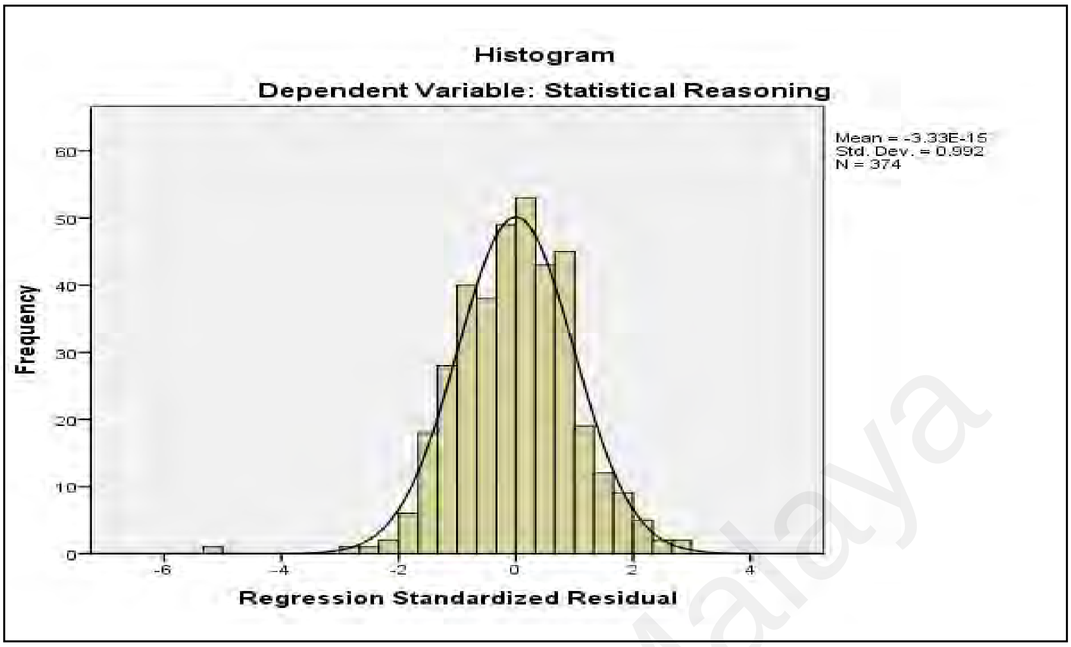

Figure 4.9: Scatterplot on distribution of SA versus MC

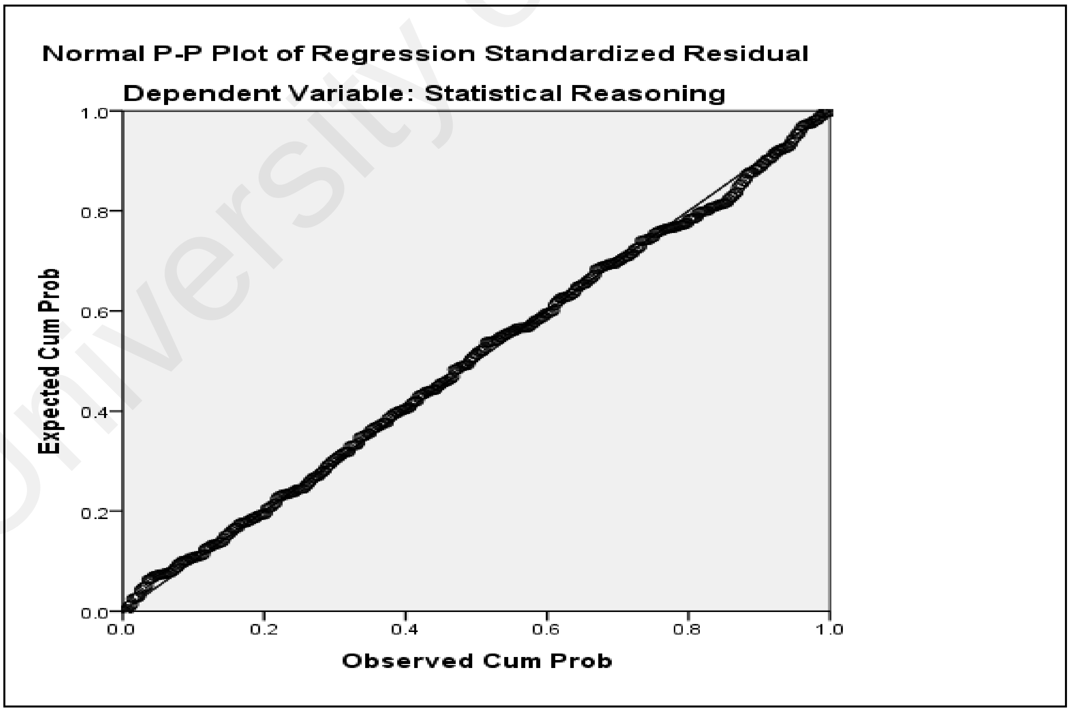

Figure 4.10: Scatterplot on distribution of statistical reasoning normality check

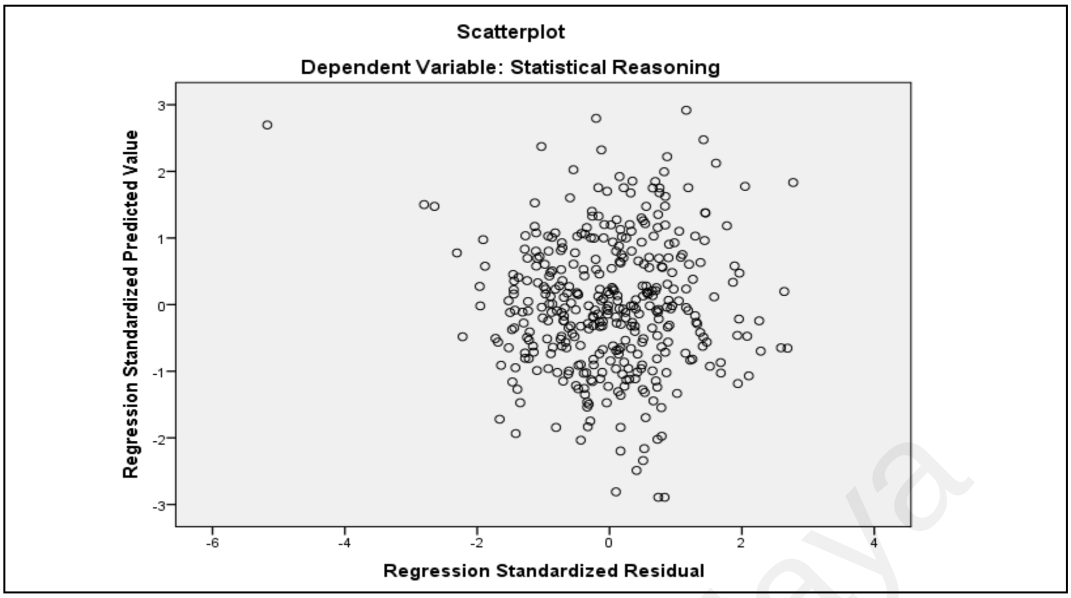

Figure 4.11: Scatterplot on distribution of standardized residual showing, linearity, homoscedasticity and independence (Field, 2013)

The normality checks for statistical reasoning were done as shown in Figure 4.9 and Figure 4.10 whereas Figure 4.11 shows the scatterplot that indicating linearity, homoscedasticity and independence of errors.

Table 4.28: Residuals Checks

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	*N*
Predicted Value	15.4425	61.0940	38.1666	7.85541	374
Residual	-59.34327	31.83734	.00000	11.38105	374
Std. Predicted Value	-2.893	2.919	.000	1.000	374
Std. Residual	-5.172	2.775	.000	.992	374

a. Dependent Variable: Statistical Reasoning

4.5.2 Best model for regression of cognitive determinants on Statistical Reasoning

The model

$$Y = 43.61 + 0.05x_2 - 0.58x_3 + 3.45x_4 \quad \textbf{4.9}$$

where Y= statistical reasoning (SR)

$x_{2=}$ statistical achievement (SA)

$x_{3=}$ misconception (MC)

x_4 = English Language (ENG)

In the standardized unit by employing the Standardized Coefficients, one can say that statistical reasoning has an inverse relation with misconception whereby an increase of approximately half a unit of misconception score will see a decrease of about one unit of SR score. Language mastery shows a strong positive effect on statistical reasoning. This highlighted the case that language plays a role in determining the students' reasoning skills. This is a logical conclusion as one can see that the SRA instrument requires a substantial language mastery to understand the items! SA does not have much impact on SR though significant.

4.6 Moderating effect of language mastery and gender on the relationships between statistical reasoning and the predictors

The next section deals with the question of moderation by certain qualitative or quantitative variables. This research only deals with two variables i.e. language mastery and gender. The moderation analysis follows this procedure:

Step 1: Using a survey of the relevant literature, identify predictor (IV_1), the moderator known as IV_2, and of course the outcome variable (DV). Here the IVs can be discrete or continuous.

Step 2: Centered the IV but not the DV. Create a new variable to test the interaction effect by multiplying the selected centered IV with the centered moderator.

Step 3: Run the regression analysis again but this time with an added interaction term. Put in the centered IVs and centered moderator like normal and then put in the interaction variable in a separate block. If the *p*- value is less than .05 then there is a moderation effect.

The moderating effect of the variables language mastery (ENG) and gender (GEN) on the relationships of the response variables like SR, PMK and MC

The next research question seeks to inquire if GEN and ENG have indirect effect on the relationships form among the variables SR, PMK and MC. The previous research question has found that only MC and ENG have significant effect on SR. Thus the moderation analysis is done based on this fact:

4.6.1.1 Does language mastery moderate the influence of misconception on statistical reasoning?

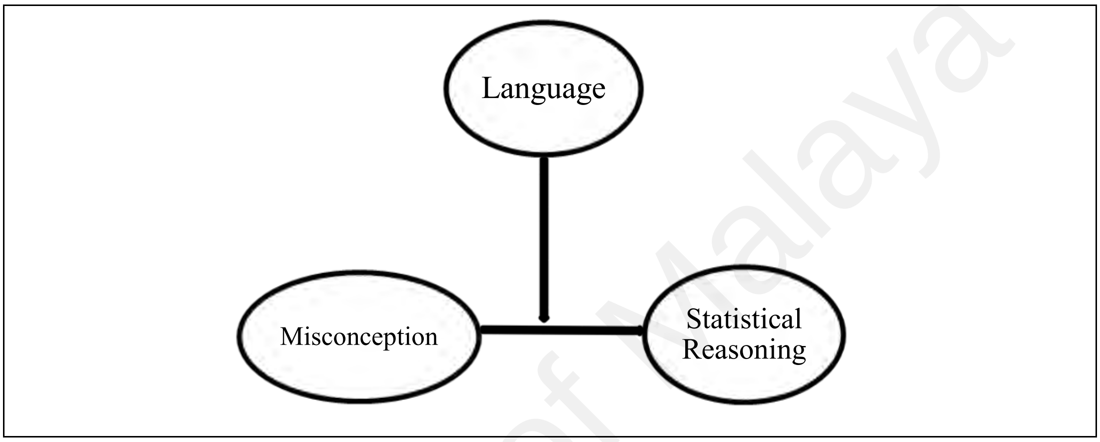

Figure 4.12: Moderating effect of ENG on the relationship between MC and SR

To confirm the moderating effect and identify which is the moderator, MC or ENG, the procedure explained in chapter 3 will be used to study this effect.

The following is the analysis as outlined by the procedure.

Regression analysis for MC, SR and ENG

Table 4.29: Moderator Effect on language mastery on the said relationship

Model Summary

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate	Change Statistics: *R* Square Change	*F* Change	*df1*	*df2*	Sig. *F* Change	Durbin-Watson
1	.525[a]	.276	.274	11.78640	.276	141.471	1	372	.000	
2	.556[b]	.309	.305	11.52620	.033	17.985	1	371	.000	
3	.562[c]	.316	.310	11.48490	.007	3.673	1	370	.056	1.789

a. Predictors: (Constant), Misconception
b. Predictors: (Constant), Misconception, English Language
c. Predictors: (Constant), Misconception, English Language, zMC_zENG
d. Dependent Variable: Statistical Reasoning

Table 4.30: Regression Coefficient

Coefficients[a]

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	***t***	**Sig.**
		B	Std. Error	Beta		
1	(Constant)	-.004	.043		-.082	.934
	Zscore(MC)	-.493	.044	-.493	-11.263	.000
	Zscore(Language)	.187	.044	.186	4.241	.000
2	(Constant)	-.019	.044		-.445	.657
	Zscore(MC)	-.484	.044	-.484	-11.039	.000
	Zscore(Language)	.197	.044	.196	4.451	.000
	zMC_zENG	-.094	.049	-.083	-1.917	.056

a. Dependent Variable: Zscore(SR)

Table 4.31: ANOVA

	Model	**Sum of Squares**	***df***	**Mean Square**	***F***	**Sig.**
1	Regression	19652.994	1	19652.994	141.471	.000[b]
	Residual	51677.938	372	138.919		
	Total	71330.932	373			
2	Regression	22042.338	2	11021.169	82.957	.000[c]
	Residual	49288.594	371	132.853		
	Total	71330.932	373			
3	Regression	22526.833	3	7508.944	56.928	.000[d]
	Residual	48804.100	370	131.903		

a. Dependent Variable: Statistical Reasoning
b. Predictors: (Constant), Misconception
c. Predictors: (Constant), Misconception, English Language
d. Predictors: (Constant), Misconception, English Language, zMC_zENG

A multiple regression model (Figure 4.12) was tested to investigate whether the association between MC and SR depends on Language mastery (ENG). After centering MC and SR and computing the zMC x zENG interaction term (Dawson, 2014), the two predictors and the interaction were entered into a simultaneous regression model. Results from Table 4.29 indicate that MC (b = -.493, SE_b = .044, β = -.493, p < .001) and ENG (b = .187, SE_b = .044, β = .186, p < .001) were both associated with SR. However the interaction between MC and ENG (ZMC_ZENG) was not significant (b = -.094, SE_b = .049, β = -.083, p < .001 while ZMC_ZENG, p > 0.05), suggesting that MC does not depend on ENG. Table 4.30 shows that both the generated models are significant.

As such it confirms that English language does not act as a moderator in the relationship between MC and SR.

4.6.1.2 Does gender moderate the influence of misconception on statistical reasoning?

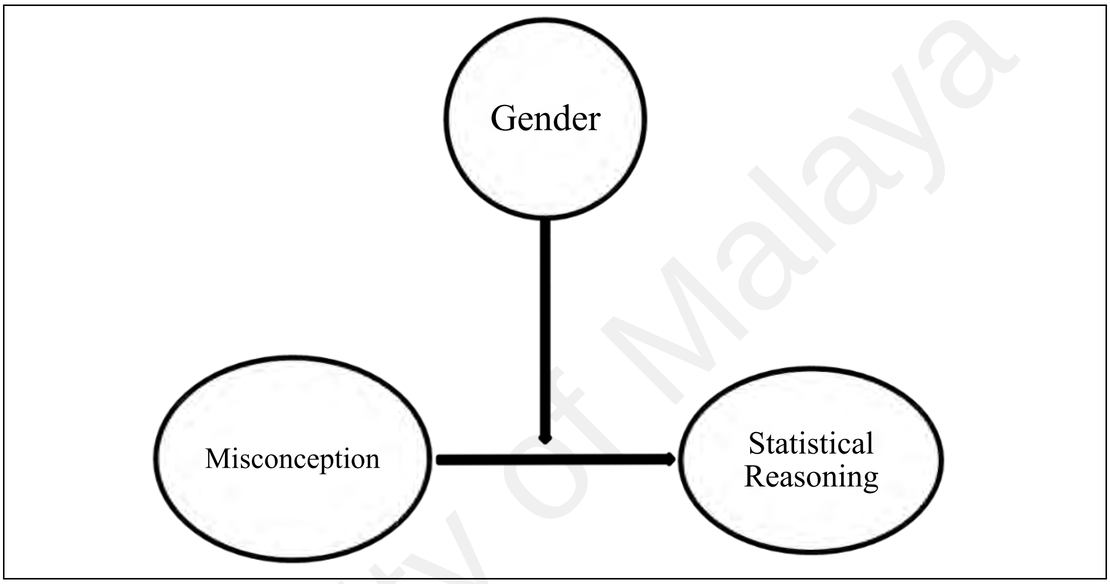

Figure 4.13: Moderating effect of GEN on the relationship between MC and SR

The procedure for testing the existence of a moderating effect of GEN on the relationship between SR and MC is described in the next section.

Regression analysis for MC, SR and GEN

Table 4.30: Regression analysis to test for moderating effect of GEN on SR and MC.

Model Summary[c]

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate	Change Statistics: *R* Square Change	*F* Change	*df1*	*df2*	Sig. *F* Change	Durbin-Watson
1	.527[a]	.278	.274	11.78300	.278	71.384	2	371	.000	
2	.530[b]	.281	.275	11.77698	.003	1.379	1	370	.241	1.869

a. Predictors: (Constant), Dummy_GEN, MC
b. Predictors: (Constant), Dummy_GEN, MC, zMC_zDummy_GEN
c. Dependent Variable: SR

Table 4.31: Regression coefficients

Coefficients[a]

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	***t***	**Sig.**
		B	Std. Error	Beta		
1	(Constant)	61.206	2.307		26.531	.000
	MC	-.631	.053	-.527	-11.936	.000
	Dummy_GEN	-1.663	1.509	-.049	-1.102	.271
2	(Constant)	65.758	4.510		14.580	.000
	MC	-.759	.121	-.634	-6.257	.000
	Dummy_GEN	-1.791	1.512	-.052	-1.185	.237
	zMC_zDummy_GEN	1.827	1.555	.119	1.174	.241

Table 4.32: ANOVA table

Model		**Sum of Squares**	***df***	**Mean Square**	***F***	**Sig.**
1	Regression	**19821.645**	**2**	**9910.822**	**71.384**	**.000[b]**
	Residual	51509.288	371	138.839		
	Total	71330.932	373			
2	Regression	20012.927	3	6670.976	48.097	.000[c]
	Residual	51318.005	370	138.697		
	Total	71330.932	373			

a. Dependent Variable: SR
b. Predictors: (Constant), Dummy_GEN, MC
c. Predictors: (Constant), Dummy_GEN, MC, zMC_zDummy_GEN

Figure 4.13 represents a multiple regression model to investigate whether the association between MC and SR depends on Gender (GEN). After centering MC and SR and computing the zMC_zDummy_GEN interaction term (Dawson, 2014), the two predictors and the interaction were entered into a simultaneous regression model. Results from Table 4.31 indicate that MC (b = -.759, SE_b = .121, β = -.634, p < .001) and Dummy_GEN (b = -1.791, SE_b = 1.512, β = -.052, p < .001) were both associated with SR. However the interaction between MC and Dummy_GEN ZMC_ZGEN was not significant (b = 1.827, SE_b = 1.555, β = .119, p < .001 while ZMC_ZGEN, p > 0.05), suggesting that MC does not depend on GEN. Table 4.32 shows that both the models are significant.

As such it confirms that gender does not act as a moderator in the relationship between MC and SR.

4.7 Summary

The extensive amount of findings elicited from the chapter can be summarised according to each of the research questions

I. Descriptive analysis

PMK (*M* = 78.54, *SD* = 11.72) and SA (*M* = 64.63, *SD* = 24.78) as compared to SR (*M*= 38.17, *SD* = 13.83) and MC (*M* = 34.44, *SD* = 11.56). On average, students showed quite good mastery of prior mathematical knowledge and their mean statistical achievement was well above average. Unfortunately, they did not do well in Statistical Reasoning (SR) and had a substantially high level of Misconception (MC) about statistics

II. The relationships between statistical achievement and the predictors (i.e. prior mathematical knowledge, statistical reasoning and statistical misconception)

Regression Model was hypothesized as:

$$\mathrm{Y_i} = \beta_\circ + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \qquad \textbf{4.10}$$

where $\mathrm{Y_i}$= statistical achievement (SA)

X_1= prior mathematical knowledge (PMK)

X_2= statistical reasoning (SR)

X_3= statistical misconception (MC)

X_4 = English Language (ENG)

X_5 = Gender (GEN)

The model takes the form of:

$$Y = B_\circ + B_1 x_1 + B_2 x_2 \qquad \textbf{4.11}$$

where Y= statistical achievement (SA)

x_1 = prior mathematical knowledge (PMK)

x_2= statistical reasoning (SR)

The final model with unstandardized coefficients is given by equation 4.313

$$Y = 8.75 + .58\, x_1 + .271 x_2 \qquad \textbf{4.12}$$

or

$$\text{SA} = 8.75 + .58\ (\text{PMK}) + .27(\text{SR}) \qquad \textbf{4.13}$$

The final model only consists of prior mathematical knowledge and statistical reasoning as significant contributors. PMK contributes almost twice as much as compared to SR (see Table 4.11 for Standardized Coefficients in making this comparison). However both of them only contributed 10% of the variance in Statistical Achievement, raising the question: What other factors are influencing SA? Literature has pointed to a whole range of cognitive and non-cognitive determinants not studied in this research.

III. The moderating effect of the variables language mastery (ENG) and gender (GEN) on the relationships of the other response variables like SA, SR, PMK and MC.

The analysis using the recommended moderation technique shows that neither language mastery nor gender has any indirect effect on the different relationships among SR, PMK and MC on SA.

IV. The relationships between statistical reasoning and the predictors (i.e. prior mathematical knowledge, statistical misconception)

The hypothesized model suggests:

$$y_i = \beta_\circ + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \varepsilon \quad \textbf{4.14}$$

where y_i= statistical reasoning (SR)

x_1= prior mathematical knowledge (PMK)

x_2= statistical achievement (SA)

x_3= statistical misconception (MC)

x_4 = English Language (ENG)

x_5 = Gender (GEN)

The model is:

$$\mathrm{Y} = 43.61 + 0.05x_2 - 0.58x_3 + 3.45x_4$$

where Y= statistical reasoning (SR)

x_2= statistical achievement (SA) **4.15**

x_3= misconception (MC)

x_4 = English Language (ENG)

or

$$\mathrm{SR} = 43.61 + 0.05(\mathrm{SA}) - 0.58(\mathrm{MC}) + 3.45(\mathrm{ENG}) \quad \textbf{4.16}$$

In the standardized unit by employing the Standardized Coefficients (see Table 4.27), one can say that statistical reasoning has an inverse relation with misconception whereby an increase of approximately half a unit of misconception score will see a decrease of about one unit of SR score. Language mastery shows a positive effect on statistical reasoning as compared to misconception. This highlighted the case that language plays a major role in determining the students' reasoning skills. Statistical achievement plays only a minor positive role in this model.

V. The moderating effect of the variables language mastery (ENG) and gender (GEN) on the relationships of the response variables like SR, PMK and MC.

The findings from the moderation analysis show that neither language mastery nor gender has any indirect effect on the different relationships among PMK and MC on SR.

VI. The Final Models are:

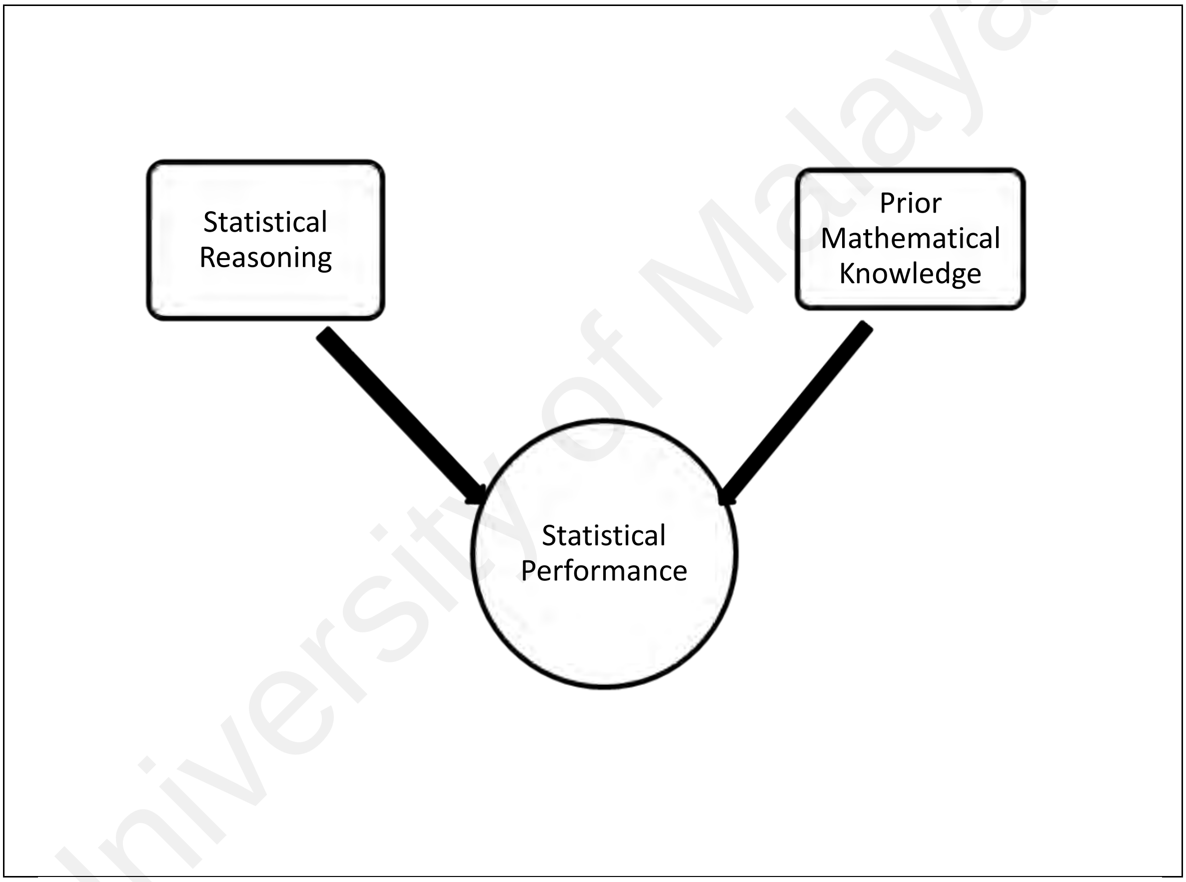

Figure 4.14: The best model showing the relationships prior mathematical knowledge, statistical reasoning and statistical achievement

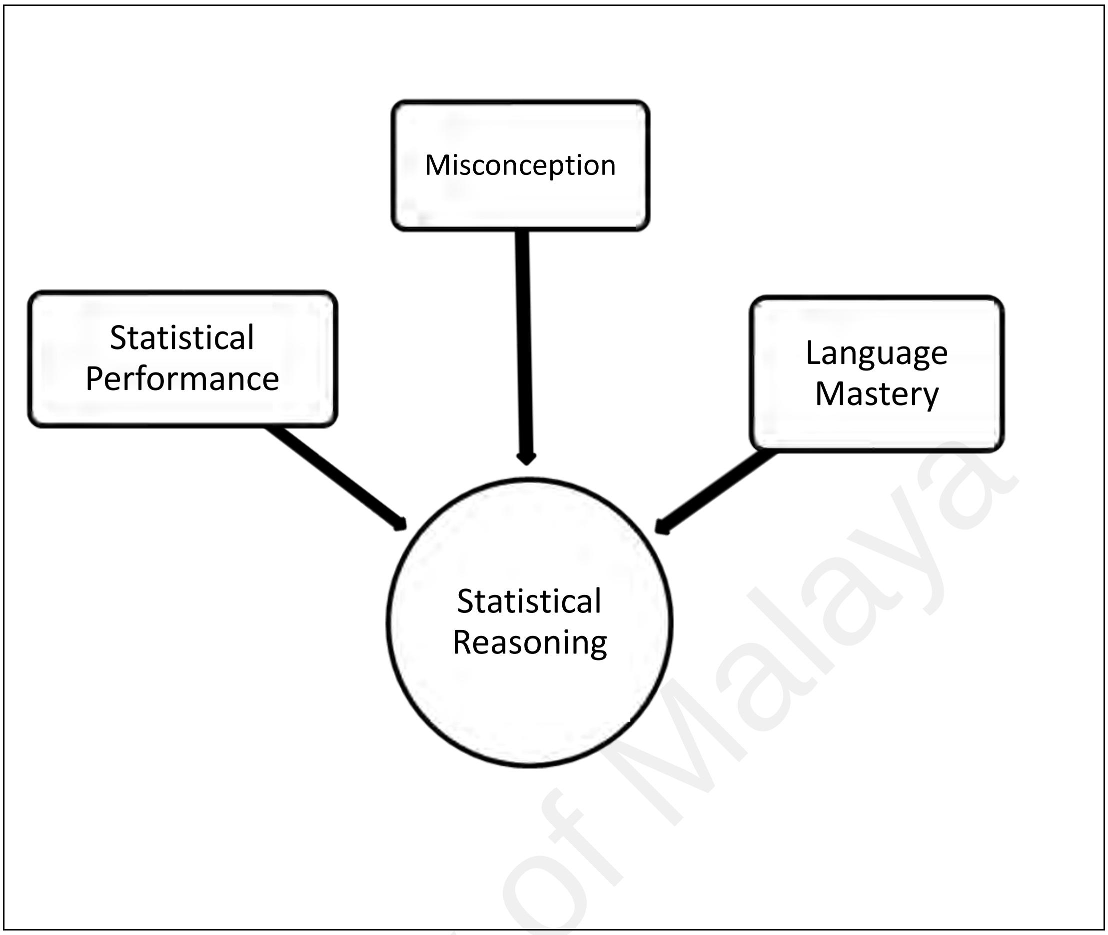

Figure 4.15: The best model showing the relationships between statistical achievement, misconception, language mastery and statistical reasoning

CHAPTER 5 : DISCUSSION AND CONCLUSION

5.1 Introduction

Chapter 5 revisits the purpose, problem statement, literature review and approaches to the data collection and analysis strategy in the light of the findings from the current study. Subsequently, a short presentation of the contributions and its implications to the current teaching and learning of statistics in a diploma classroom is discussed. The chapter closes with some recommendations for future studies

This study has explored, analyzed and characterized the findings by looking at the statistical achievement of Diploma Science students in a large Malaysian university and its relation to selected cognitive determinants like statistical reasoning, misconception and mathematical prior knowledge. In addition it studied the influence of gender and language mastery on the hypothesized relationships among the independent variables and dependent variables.

This study investigated the various hypothesized relationships of cognitive determinants like prior knowledge, statistical reasoning and statistical misconceptions, gender and language mastery that had been identified a priori to influence statistical achievement of Malaysian students. In addition, this study was carried out to determine the direct and indirect effect of gender and language mastery on the various relationships among the variables.

5.2 Discussion

The extensive amount of findings elicited from the chapter can be summarised according to research questions designed.

The academic profile of the respondents showed an above average proficiency level in term of mastery of prior mathematical knowledge, statistical achievement and language competency. However, they did not do too well in Statistical Reasoning (SR) and had a substantially high level of Misconception (MC) about statistics. Statistical achievement among Malaysian students was found to be mediocre. PMK (M = 78.54, SD = 11.72) and SA (M = 64.63, SD = 24.78) as compared to SR (M= 38.17, SD = 13.83) and MC (M = 34.44, SD = 11.56). Noraidah et al. (2011) noted that in a Malaysian public university the statistical achievement is only average. In another public university, the diploma students were found lacking too in this area. These findings concurred with those found in this study. Malaysian students need to pay more attention to the teaching and learning of statistics to counter the declining trend of statistics achievement. The level of reasoning skills among diploma students in Malaysia is low. This concurred with results from studies by Zamalia & Nor Hasmaniza (2010) and Chan, Zaleha and Bambang (2014). TIMSS reports on Malaysian students' achievement in the 'Data and Chance' category similarly indicated the same trend (Mullis et al., 2000, 2008, 2012).

The first research objective was answered using the results of the multiple regression analysis on statistical achievement with the assigned cognitive determinants. Results showed that there exists a significant relationship between statistical achievement and two predictors, i.e. prior mathematical knowledge (PMK) and statistical reasoning (SR).

The best model is given by the equation

$$\mathrm{SA} = 8.75 + .58\,(\mathrm{PMK}) + .27(\mathrm{SR})$$

PMK represented almost twice as much of the total variance as compared to that of SR. However both of them only contributed a lowly (10%) to the variance in Statistical Achievement. Achievement is a rather complex construct that has many dimensions to it. Studies have shown many cognitive and non-cognitive determinants like student previous course of study, their grade point average, language skills, self-efficacy, student's attitude towards statistics or student perception of statistics as a difficult subject is partially responsible for this state of affair (Lalonde & Gardner, 1993; Hardre et al, 2006; Dempster & McCorry, 2009; Chang and Cheo, 2012). In reality it is not surprising that PMK and SR only accounted for 10% of the variance found as many cognitive and non-cognitive factors have not been included in this current study.

IPT and in particular the Schema Theory can partly explain the findings earlier. Schema theory (Eysenck & Keane, 2015) has explained the importance of students' prior knowledge in influencing the understanding and construction of new statistical knowledge. Human mind utilizes schemata to organize, retrieve and encode large amount of information. If encoding, organizing and retrieving are not done well or correctly, the process will lead to distortion and mistakes. The newly 'revised' schemata will cause misconceptions to develop. Studies had shown that prior knowledge is an important determinant of undergraduates' academic achievement (Chang & Cheo, 2012). This study confirmed the importance of PMK in influencing achievement in statistics class just as those found in studies by Chiesi, Primi and Morsanyi (2009); Chiesi and Primi (2010) and Zuraida et al. (2012). However this study did not look at the type of mathematical content (e.g. Operations, Fractions, Set theory, first order Equations, Relations and Probability) that has an effect on achievement. It is recommended that future studies look into this aspect of prior mathematical knowledge.

Statistical Reasoning in this study showed a positive effect on Statistical Achievement. This finding provides more evidence about the differential effect of statistical reasoning on achievement where some studies showed low or negligible effect while others indicated moderate effect of SR on performance (Liu, 1998; Garfield, 2002, 2003; Tempelaar, 2004; Zuraida et al., 2012). One possible reason for the different results was due to the reliability and validity issues of the data collection instrument (SRA). The reliability of the instrument by Garfield (1998, 2003); Garfield and Chance (2000); Liu (1998); Sundre (2003) and Tempelaar et al. (2007) were average ranging from r=.70 to r=.75. Tempelaar et al. (2007) attempted with a similar approach using aggregated scores and found similar reliability indices as Garfield. Their studies showed that Cronbach alpha for both the scales were 0.24 and 0.06 respectively while the present study showed Cronbach alpha to be low too (.50). In addition, Gigerenzer and Goldstein (1996) noted that everyone displays bounded rationality with constraints due to factors like limited capacity of working memory and one's cognitive goals. The fact that each one has different cognitive goals each time one uses the reasoning power, was well supported from the research of Mercier (2013). There are times when a person is a good reasoner but at other times one may just reason badly. Hardman and Macchi (2003) explained the cognitive threesomes of reasoning, judgment and decision making as closely related and overlapping as talking about one will invoke the others. This is also true for statistical reasoning as invoking statistical reasoning one is invariably led to statistical thinking and statistical literacy. In other words, psychologists agreed that when individuals reason about something, invariably they will need to make a judgment call as well as make some kind of decision after considering all the options opened to them. This then can be extrapolated to the case of the threesome of statistical reasoning, statistical literacy and statistical thinking (delMas, 2004a). Martin (2013) commented on the multiple facets of statistical reasoning making assessment of the reasoning complicated. Many statisticians agreed on the importance of acquiring these abilities (Chance & Garfield, 2002; delMas, 2002; Garfield, 2002; Rumsey, 2002; Garfield & Ben-Zvi, 2008) but there is less consensus as to their actual use and operationalization of those constructs (Ben-Zvi & Garfield, 2004a, 2004b; delMas, 2004a; Garfield & Ben-Zvi, 2008). Unless a study controls for extraneous variables stringently, it is inevitable that results about the influence of SR on SA will vary due to the many factors described earlier. Herein lays a limitation of this study. An observational study design is inappropriate under such stringent circumstances. A better design would be an experimental approach that can control for the various extraneous factors.

The literature in chapter 2 has recounted the various factors and circumstances under which, a student operates to be a successful reasoner but ultimately from an educator's perspective what is important is how one is going to 'make' a good reasoner.

It is important to note that other factors like gender, language mastery and statistical misconception did not affect the performance. From literature, the impact of MC on achievement is significant too but using SRA instrument to measure both reasoning and misconception concurrently ran the problem of common variance shared as there is quite a strong correlation between these two variables. This is the most probable reason for seeing the insignificant effect of MC on SA. Furthermore gender and language mastery do not seem to affect SA as found in some of the studies mentioned in Chapter 1 and 2.

The third research objective was answered by further regression analysis of the relationships between statistical reasoning and the predictors (i.e. prior mathematical knowledge, statistical misconception, language, statistical achievement and gender). It was found that only three cognitive determinants had significant effect on reasoning.

The best model is:

$$\text{Y} = 43.61 + 0.05x_2 - 0.58x_3 + 3.45x_4$$

or

$$\text{SR} = 43.61 + 0.05(\text{SA}) - 0.58(\text{MC}) + 3.45(\text{ENG})$$

Statistical reasoning showed an inverse relation with misconception while language mastery shows a positive effect. In the standardized unit an increase of approximately half a unit of misconception score will see a decrease of about one unit of SR score. Statistical achievement plays a lesser positive role in this model. The inverse relationship between SR and MC is expected as students with lesser misconception would imply they have better understanding of statistics. Conversely students having high level of misconceptions would be bringing these to classes preconceptions and statistical misunderstanding that would hamper their construction of new and correct conceptions of statistics. It has been warned by many statistics educators including Newton (2000) that understanding failure is just due to factual error and could be rectified quite easily but if ideas or concepts are theoretically based they are much more difficult to overcome especially those of psychological nature (Huck, 2004; Shaughnessy, 1981; Kahneman & Tversky, 1972). Schema Theory provides some explanations about the consequences of developing misconception schemata. When errors are developed, there is a tendency to retrieve a similar or incorrect schema resembling the original schema. This is one reason for the occurrence of a variety of cognitive biasness that was discussed in the previous chapter (Huck, 2004; Shaughnessy, 1981a; Kahneman & Tversky, 1972). Once a schema is developed, it tends to be stable over a long period of time and to unlearn is much more difficult to relearn.

In addition, the Schema theory highlighted the effect of memory distortions and reconstructive memory. These two important concepts can in part explained the misconceptions among the Diploma students. The theory states that the accuracy of storage of any information presented to a student depends on the following: i) the level of attention paid to the original information, ii) the time that passes, iii) the matching of contexts, and iv) the presence of interference (Loftus, 2003). In essence, memory does not store the exact duplicate of information. It abstracts the gist and essential components only and fits them into schemas that make sense to the receiver of the information. Reconstructive memory suggests that in the absence of all information, one fills in the gaps to make more sense of what happened. This is why reconstructive memory contains distortions, deletions and omissions (Bartlett, 1932). The theory can then accounts for the failure of students in understanding basic concepts in statistics. Wrong understanding then leads to misconceptions due to the brain's attempt to make sense of that incorrect information by trying to fit in to a schema that does not match the original information. The new constructed schema in effect, contains distortions, deletions and omissions. By investigating a limited numbers of cognitive determinants one cannot paint a clear picture of the effect of these factors on achievement or reasoning. It is undeniable that the constructs of achievement, reasoning or other related terms like judgment or decision making are complex and cannot be studied comprehensively using a few variables. More advanced research design is needed and incorporating sophisticated modeling tool like Structural Equation Modeling may serve this purpose.

There was no moderating effect of the variables language mastery (ENG) and gender (GEN) on the various relationships of the IV variables on the DV variable. This effectively answered the second and fourth objective of the study.

Interestingly enough this study found language mastery to be a factor in the acquisition of statistical reasoning in answering the third research objective. The ability to understand the language structure and morphology of the information is important (Reed, 2011; Shaughnessy, 1992 and Gigerenzer & Hoffrage, 1995). The linguistic schema requires the learner to decode in order to understand how words are organized and fit together in a sentence. This implies that learner needs repetitions and recalls to develop good language mastery for understanding a question or a comprehension passage.

As seen in the previous chapter, Girotto (2004) asserted that much of the difficulty of reasoning lies with understanding the language of the problems. This finding is in line with the Schema Theory that linguistic schema and content schema need to be activated simultaneously at the LTM. Activation of these schemata is one thing but activating the correct schemata becomes a priority.

Literature has consistently shown mixed results when it comes to the effect of gender (Elmore & Vasu, 1986; Schram, 1996; Noor Azina & Azmah, 2008; Reed, 2011; Chang & Cheo, 2012; Reilly, 2012). Results of the various studies indicated that under different conditions the outcomes can differ. These extraneous variables can only be controlled effectively using an experimental design. This study showed gender did not affect any of the purported relationships.

5.3 Research Design, Sample and sampling technique

The correlational design used in the present study successfully answers the research questions though it could not confirm cause and direction affirmatively. Correlation does not allow us to go beyond the data that is given. For that multiple linear regression (MLR) models were created to test for assumed cause and effect from literature and past studies.

This study used 381 respondents out of a total of over 70,000 students. The constraint of getting a larger and random sample was due to the ability of the researcher to collect them from a population that was spread out all over Malaysia. A random sampling technique was out of the question by virtue that the selection of respondents must come from the classes taught by the researcher and colleagues.

Thus the results could not be generalized due to the problem of non-random sample selection. In addition the correlational design employed could not account for the large variance found in some of the relationships and the influence of a third variable. It could not handle too many variables well concurrently. As the constructs studied here were found to be complex variables, a more flexible and efficient analytical approach would be the answer to handling tens of these variables simultaneously

Future study of this nature where a large random sample is accessible, Structural Equation Modeling (SEM) obviously could counter some of the limitations of this study. SEM is a highly flexible multivariate data analysis method that can handle three types of relationships: 1) association (correlational analysis which is non-directional), 2) causation (multiple regression models which is directional) and 3) indirect effect (mediating or moderating effect) (Chou and Bentler, 1995).

5.4 Data collection instrument

Both primary and secondary data were utilized in the analysis. Secondary data like Prior Mathematical Knowledge and Statistical Achievement were collected using the survey form distributed to students at the start of the research. The data for Prior

Mathematical knowledge comprises of aggregated score which were self-reported data. As for the Statistical Achievement score, primary data were collected using scores from their semester test scores and final examination results. The instruments used to collect these scores were standard examination papers set by the Examination Council of Malaysia as well as carefully vetted examination and test papers set for all students in this university.

Demographic profile of participants and scores for Statistical Reasoning and Misconception variables were collected through the use of the Statistical Reasoning Assessment (SRA), an adapted version by Garfield (2003). The 15-item multiple-choice instrument was piloted and checked for validity and reliability. Each multiple-choice item has between 3-6 options depending on the complexity of the items constructed to gauge the reasoning and misconception. Each correct answer contributes to an aggregated score for statistical reasoning. The other incorrect options in each item are specially designed to identify the type of misconceptions. Item scoring depends on two scoring rubrics designed to measure the respondents' reasoning and misconception. This instrument suffered from the following weaknesses.

a) Low test-retest reliability as attested by Garfield (2002). This study ran two rounds of pilot testing on the instrument and the Cronbach alpha calculated from the two sets of data were still not impressive leading to the question of the SRA as the best instrument to measure statistical reasoning and misconceptions. Additional items were needed to overcome the big variance detected in the findings of this study.

b) Coverage of statistical reasoning skills was limited. A small subset of reasoning strategies/skills was covered leading to a rather skewed interpretation of what statistical reasoning is and consequently affecting the interpretations of the findings

c) There were some items with only 3 options. These items gave room to guessing and thus creating large unaccounted variances. In addition the item format and scoring omitted potentially important information. Items with 3-4 options are not really good to use in SRA.

d) In addition, the study depended heavily on self-reported scores from the various tests and examinations to compute their prior mathematical knowledge and statistical achievement. To access the examination records of students involved a lot of bureaucracy and time. However the researcher felt that collecting secondary data from students if well carried out could still reflect their real achievement.

e) Missing values or incomplete data are quite common occurrences in the data collection. Incomplete data set has implication on the analysis which one must be aware of. A sample of 381 Diploma students was drawn from Diploma students coming from two different states of the country out of which only 374 was usable. The number of unusable survey forms was low and missing value was treated according to standard procedure.

5.4.1 Data analysis technique using Multiple Linear Regression approach

The choice of statistical analysis technique was determined by the research questions. The Multiple Linear Regression models were developed to answer these questions. MLR was successfully used within the limits and constraints of this study. All assumptions were also taken care of. Goldberger and Duncan (1973) noted that the regression models were sufficient for circumstances where the relationships investigated were far less complex.

The MLR approaches have their inherent weaknesses. One major conceptual limitation of the regression technique is that one can only investigate the relationships but not the cause and effect. The sample size too can be an issue if the variables are too many. More importantly the assumptions of this regression technique have to be fulfilled. This study paid very close attention to the fulfilment of all the stated assumptions before any interpretations were made.

5.5 Implications

The implications of the present study are discussed at several levels. In addition to a treatise of the practical implications, the current study's implications to theory building are given an equal importance in this section.

Improving teaching and learning practices.

Findings arising from this research indicated that *Bumiputera* students showed moderate achievement in prior mathematical knowledge, statistical achievement and language competency. In addition, they achieved poorly in Statistical Reasoning (SR) and possessed a substantially high level of Misconception (MC) about statistics. Many of the conclusions mentioned earlier have been explicitly addressed using Information Processing Theory and in particular the Schema Theory. Armed with the findings and the reasons for the outcomes of this research, there are ways that IPT has found to be effective in improving the teaching and learning process in class.

The Information Processing Theory states that the memory storages in the brain are very limited i.e. sensory and working memory. To overcome this problem, cognitive psychologists recommend two strategies to cope with this problem, namely selectively focusing one's attention on important information and engaging in repetitions and reinforcements to help processing of information automatic where possible. From an educational perspective, it is essential for students to become masters of basic skills and simple procedural skills. This is related to prior knowledge of which will be discussed next. It has been found that the ability to put basic cognitive skills on an automatic mode can help free more processing resources to do complex mental tasks like thinking, reasoning or problem solving (Orey, 2001 ;Schraw et al., 2001; Sternberg, 2001; Zimmerman, 2000). In the context of reasoning, Stanovich (1999) and Evans and Over (1996) entertained the idea of dual processing. Implicit thinking or System 1 thinking provides automatic input to the brain to act pragmatically utilizing knowledge and beliefs residing in the long-term memory of which Stanovich called it fundamental computational bias. This is the basis for students to resort to heuristics to reason or solve problems. Heuristics work sometimes but most of the time causes biasness and errors in human cognition. This would help to explain why students still come to class with preconceived ideas or even misconceptions about basic foundational statistical concepts. To unlearn is more difficult than relearn – a fact well-known to educators.

The other type of thinking - explicit thinking or System 2 thinking is "linked to language and the reflective consciousness, and providing the basis for reasoning" (Evans, 2007). This concurred with the results of this study where statistical reasoning was found to be influenced by language mastery. According to Evans (2008), System 2 operation requires large space in the limited working memory where information is processed linearly. It has been established that effective functioning of this system is related to the IQ. However, due to the inherent 'inefficiency' of this site to process large amount of information, there is a tendency that most of us will fall back to System 1 regularly and that is where one makes errors and acquire misconceptions.

The second implication is that relevant prior knowledge helps in encoding and retrieval of information from the long-term memory. Thus for highly sophisticated learners or experts, they possess a great deal of organized knowledge within a particular domain such as reading, mathematics, or science. They are also found to have general problem-solving and critical-thinking scripts that enable them to apply their knowledge across different domains. This knowledge guides information processing in sensory and working memory by making retrieval from the memory networks situated either in working or long-term memory (Alexander, 2003; Ericsson, 2003). Thus, making sure students come to class with the correct prior mathematical knowledge is essential to promote effective statistical learning.

Another implication is that good learning strategies in statistics classrooms help learners to process information better and with deeper understanding. Some of the strategies or methods are automated as in System 1 but deep processing and metacognition requires System 2. Thus 'activating existing knowledge prior to instruction, or providing a visual diagram of how information is organized like flowchart, mind-maps or graphics, is one of the best ways to facilitate learning new information' (Schraw & McCrudden, 2013).

The current research provides the foundation for the development of future research that has been laid out in the chapters. The literature review in chapter 2 provided much arguments and rationale to consider what informal or intuitive beliefs held by researchers who are in the initial stages of their studies. There are many Dos and Don'ts to comply or avoid to ensure that the research can be run smoothly and timely in terms of selection of variables, conceptual framework, methodology, analysis techniques and writing of the findings.

More importantly this research had used a single data collection instrument incorporating the SRA tool to assess statistical reasoning. Findings indicated that there are obvious limitations to using this instrument in terms of reliability and validity as discussed in the previous sections. There are statistical reasoning tools being constructed recently that could complement the SRA i.e. the Quantitative Reasoning

Quotient (QRQ) and the Comprehensive Assessment of Outcomes in a first Statistics course (CAOS). It was naive to think that one instrument can measure such a complex construct like statistical reasoning. SRA is an important tool to assess statistical reasoning among diploma students doing an introductory course but its usefulness can be greatly enhanced by tackling the low reliability of the instrument through the following:

i) One SRA instrument is designed for only one topic – Probability, Hypothesis Testing, Multivariate Analysis, Basic Concepts, Variability, or Misconceptions.

ii) The number of items used to assess each concept in the SRA must be at least 3 as found in CAOS instrument.

iii) The number of options for each item in the SRA must be at least 5 as found in CAOS instrument.

iv) All concepts to be assessed must be well-defined.

v) Each multiple choice item must be followed by a short answer question to check for guessing as has been done in the QRQ instrument.

Information Processing Theory has largely been used to explain many of the outcomes of the current study with respect to reasoning, prior knowledge, memory capacity, memory retrieval, memory distortions, gender and language effects as well as achievement. However there are aspects of IPT that do not account for complex cognitive processes that are studied here. One of the major drawbacks of this theory is that it assumes a serial processing information proposed by Atkinson and Shiffrin (1968) may be too simplistic to explain complex mental processes like reasoning, decision making and higher order thinking. Alternative models like the parallel-distributed processing model and the connectionist model are found to be a better replication of these processes (Huitt, 2003). The connectionistic model expounded by

Rumelhart and McClelland (1986) is by far a better model as shown by the brain research carried out by Rumelhart and McClelland. This model can explain how a person attempts to make sense of the happening around him/her by employing a 'two-way flow of information' known as 'bottom-up processing' and 'top-down processing' depending on whether the information is from outside or information retrieves from the long-term memory (Huitt, 2003).

The reductionist approach of IPT to break up a complex system like the brain into smaller manageable units of study has a great impact on how one interprets the way that the brain works. The analogy between the human brain and a computer is far too simple. It may be good for surface understanding of how the brain works but one does not bring forth real understanding that is really needed in studying complex cognitive processes like reasoning or memory distortions. As has been proven by brain researchers (Anderson, 2015, Rumelhart & McClelland, 1986), human brain has the ability to make extensive parallel processing and make connections through its extensive networking web while the computer resort mostly to serial processing. In addition, cognition is also influenced by a host of emotional and motivational determinants. The findings of the IPT are based largely from experiments under controlled scientific conditions lacking what McLeod (2008) lack 'ecological validity'. Obviously the new models described earlier hold better potentials in furthering the understanding of the human cognition.

Schema theorists like Fischbein and Grossman (1997) and Eysenck and Keane, (2015) differentiate the schema into various categories of which linguistic and content schemata are especially helpful in explaining how students acquire prior knowledge, reasoning and memory distortions. Darley and Gross (1983) found that schema theory was effective in explaining processes like perception, reconstructive memory, misconceptions, stereotyping and reasoning. However the theory remains ineffective as the present conception of what a schema is, remain vague and does not explain how schemata are acquired (Cohen, 1993 as cited by McLeod, 2009). The ideas of reconstructive memory and memory distortions by Schema theorists (Loftus, 2003; Darley & Gross, 1983; Bartlett, 1932) to explain misconceptions, reasoning failure or memory lapses are largely theoretical rather than empirically based.

5.6 Future Research

Based on this study there are several recommendations for future research.

Firstly, since it is impossible to examine all variables simultaneously only three variables that were believed to have stronger effect on *Bumiputera* students' achievement were studied. The current study has clearly shown that statistical achievement and reasoning are complex constructs that require researchers to test out a whole range of cognitive and non-cognitive determinants to account for the remaining variances. Future studies should look in this direction to understand the contributing factors to high achievement in statistics. These studies should include other motivational variables such as goals, value, or interest and examine how the various variables operate in concert. Additionally, the study should be replicated with samples from a population that includes Diploma students in various institutions of higher learning in all parts of Malaysia. The pursuit to understand the influence of learner variables on achievement or achievement needs to continue.

Secondly, even though findings of this study can be partially explained by the Information Processing Theory, future research may want to study them using a different paradigm like qualitative research methodologies where in-depth examination of these few determinants across cultures and creed using the diversity in this country to the best of its advantage. This study is suggested to be repeated with the same type of sample to compare the results with different samples and classes at the postgraduate level and with a statistics class at the undergraduate level from different research paradigms.

In addition this study should also be repeated with a larger sample to compare results and explore if some of the trends toward significance for variables like gender, misconceptions, language would become significant with this increased sample size. In this research the correlations between language mastery with both statistical achievement and prior mathematical knowledge are not significant (see Table 4.3). Further investigations may validate these results with different sample sizes or even under different circumstances.

Another suggestion for future study is to use primary data for prior mathematical knowledge, and language mastery by creating new instruments to measure these criterion variables. Findings could have been different if primary data were used.
Finally, definition of terms used in research varies. A term used by psychologists can significantly differ from that of an educationist. The term 'achievement' is loosely defined as 'achievement' or 'ability'. Future studies must clearly choose or redefine the important constructs. A point in case is the term 'reasoning'.

From a psychologist perspective, reasoning, noted Galotti (2008) involves cognitive processes that turn bits and bytes of data into useful information so that the person can come to a conclusion. Mercier and Sperber (2011) see reasoning as a means to improve knowledge and make better decisions.

From an educationist point of view, reasoning being a higher order thinking skill is required for many of the thought processes in learning thus definition of the term varies greatly under different circumstances. This construct has been named differently

- informal reasoning versus formal reasoning, implicit vs. explicit reasoning, deductive vs. inductive reasoning, spatial reasoning, geometrical reasoning, proportional reasoning, argumentative reasoning, abductive reasoning, analogical reasoning and many more. Why are there so many different forms of reasoning? The problem is analogous to the different types of intelligences introduced by Howard Gardner. This could only imply that reasoning is a complex construct that has direct relation to a variety of cognitive processes.

As statistical reasoning is a complex construct and with the way it is defined, problem with using the SRA as the only instrument to measure this construct can be traced to the 'undefined' term that had given rise to different interpretations of the construct. Take for example the definition suggested by Garfield (2003). Statistical reasoning was defined as 'the way students reason with statistical ideas and make sense of statistical information'. The usage of the term 'reason' in its definition provokes thoughts of a circular definition as the meaning of the term 'reason' is not being addressed. In addition the term 'making sense' could be interpreted differently by different researchers. In this sense, it would be good for those involved in statistical reasoning research to redefine it. The researcher suggested a definition along the line of "the mental process of using statistical ideas and turn them into information to be able to judge and decide on best option to overcome an unsolved statistical situation".

Further evidence why the construct cannot be measured well comes from the PCA analysis of the SRA instrument – the number of dimensions keeps on changing with different population and different sample sizes. This is reflected in the different reliability indices for different studies and most of them are mostly low (Garfield (1998, 2003); Garfield, delMas and Chance (2002); Liu (1998); Sundre (2003) and Tempelaar et al., (2007). The results on the relationship between some well-known variables change for different studies indicating that probably the researchers were measuring different things. The language issue and its influence on student's interpretations of the SRA instrument must be taken into account too. Different students understand the items differently in relation to their language mastery. As a final analysis to this issue, it is highly recommended that a series of instruments must be used to cover the different aspects of this construct.

5.7 Summary

This study started out to determine the various relationships of cognitive determinants on statistical achievement of *Bumiputera* Diploma students. Furthermore, the study was intended to identify the direct and indirect effect of gender and language mastery on the various relationships. The research showed that on an average, learners achieved moderately well on prior mathematical knowledge (PMK) and statistical achievement (SA). Unfortunately, they did not do well in statistical reasoning (SR) and had a substantially high level of misconception (MC) about statistics. The best regression model on statistical achievement was:

$\mathrm{SA} = 8.75 + .58\,(\mathrm{PMK}) + .27(\mathrm{SR})$ with only prior mathematical knowledge (PMK) and statistical reasoning (SR) being significant contributors. The best model on statistical reasoning was: $\mathrm{SR} = 43.61 + 0.05(\mathrm{SA}) - 0.58(\mathrm{MC}) + 3.45(\mathrm{ENG})$ where SA, MC and ENG were significant contributors to SR. The findings found that gender and language mastery did not moderate the hypothesized relationships.

The study corroborated many of the predictions from Information Processing Theory as described in the previous sections. Important findings that emerged from this study can be explained through this theory and implications for learning and instructions were recommended as a direct result of these findings. Some promising new quantitative methods like SEM and newly verified data collection methods like QRQ and COAS are suggested to be used in future studies involving the construct of reasoning. Implications from this study can have far-reaching influence on future studies to confirm the roles played by the various cognitive and non-cognitive determinants on achievement or reasoning.

As a final thought, the end of any research is but the beginning of a series of new ones. A good research should be able to generate renewed interest and excitement to other researchers who want to take up the challenges of solving the unsolved. It is hope that this present study can generate enough interest and provide the necessary guideline for future research seeking to evaluate the relationships among cognitive determinants and statistical achievement.

REFERENCES

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House*, 78(5), 218-223.

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32*, 10–14.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *98* (4), 369-406.

Anderson, J.R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51* (4), 355-365.

Anderson, J. R. (2015). *Cognitive psychology and its implications* (8th ed.). New York, NY: Worth Publishers.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, N.H. (1970). Functional measurement and psychophysical judgment. *Psychological Review*, *77*, 153-170.

Anderson, R.C. (1977). The notion of schema and the educational enterprise: General discussion of the conference. In R.C. Anderson, J. Spiro, & W.E. Montague (Eds.), *Schooling and the Acquisition of knowledge* (pp. 415-431). Hillsdale: Erlbaum.

Arbuckle, J.L. (1996) Full information estimation in the presence of incomplete data. In G.A. Marcoulides and R.E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and Techniques* (pp. 243-277) Mahwah, NJ: Lawrence Erlbaum Associates.

American Statistical Association (ASA) (2005a). *Guidelines for assessment and instruction in statistics education* (GAISE) *college report*. Alexandria, VA: ASA. Retrieved from www.amstat.org/education/gaise/

American Statistical Association (ASA) (2005b). *Guidelines for assessment and instruction in statistics education* (GAISE): A curriculum framework for PreK-12 statistics education. Retrieved from http://www.amstat.org/education/gaise/GAISEPreK-12.htm.

American Statistical Association (ASA) (2005c). *Guidelines for assessment and instruction in statistics education (GAISE).*. Retrieved from http://www.amstat.org/education/gaise/GAISECollege.htm.

American Statistical Association (ASA) (2007). *ASA vision, mission and history*. Retrieved from www.amstat.org/about/

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *The Psychology of Learning and Motivation*, *2*, 89-195.

Axelrod, R. (1973). Schema theory: An information processing model of perception and cognition. *American Political Science Review, 67*(4), 1248-1266.

Babbie, E. (1990). The essential wisdom of sociology. *Teaching Sociology*, *18*(4), 526-530.

Bakker, A. & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben- Zvi and J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 147-168). Dordrecht, The Netherlands: Kluwer.

Baron, J. (2004). Normative models of judgment and decision making. In D. J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making,* (pp. 19–36). London: Blackwell.

Baron, R. M. & Kenny D. A. (1986). The Moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Penality and Social Psychology*, *51*(6), 1173-1182.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, UK: Cambridge University Press.

Ben-Zvi, D., & Garfield, J. (1999). *Statistical reasoning, thinking, and literacy: Selected readings*. Rehovot, Israel: Weizmann Institute of Science.

Ben-Zvi, D., & Garfield, J. B. (2004a). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 3-15). Dordrecht, The Netherlands: Kluwer Academic Publishing.

Ben-Zvi, D., & Garfield, J. B. (Eds.). (2004b). *The challenge of developing statistical literacy, reasoning, and thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishing.

Best, J. B. (1982). Misconceptions about psychology among students who perform highly. *Psychological Reports, 51*, 239-244.

Bloom B. S. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain.* New York, NY: David McKay Co Inc.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in Lisrel maximum likelihood estimation. Psychometrika, *50*(2), 229-242.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (1999). *How people learn: brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brewer, W. F., & Samarapungavan, A. (1991). Children's theories versus scientific theories: Differences in reasoning or differences in knowledge? In R. R Hoffman & D. S. Palermo (Eds.), *Cognition and the symbolic processes: Applied and ecological perspectives* (Vol. 3, pp. 209–232). Hillsdale, NJ: Erlbaum.

Broers, N. J. (2009). Using Propositions for the Assessment of Structural Knowledge of Statistics. *Journal of Statistics Education* [Online], *17*(2). Retrieved from www.amstat.org/publications/jse/v17n2/Broers.html.

Brooks, C. (1987). Superiority of women in statistics achievement. *Teaching of Psychology*, *14*, 45.

Broadbent, D. (1958). *Perception and communication.* London: Pergamon Press.

Brown, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, Bruce, B. C., & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education* (pp. 453-481). Hillsdale, NJ: Erlbaum.

Brown, A.L. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science, 14*, 107-133.

Byrne, B. M. (2001). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.

Buck, J. (1985). A failure to find gender differences in statistics achievement, *Teaching of Psychology, 12*, 100.

Carmona, J. (2004). Revising the reliability and validity evidence on attitudes and anxiety towards statistics questionnaires. *Statistics Education Research Journal*, *3*(1), 5-28. Retrieved from http://www.stat.auckland.ac.nz/~iase/serj/.

Chance, B. L., & Garfield, J. B. (2002). New approaches to gathering data on student learning for research in statistics education. *Journal of Educational Statistics, 1*, 38-41.

Chang D. W. & Cheo, R. K. (2012). Determinants of Malaysian and Singaporean economics undergraduates' academic performance. International Review of Economics Education, *11*(2), 7-27.

Chan, S. W., Zaleha Ismail, & Bambang Sumintono. (2014). A Rasch model analysis on secondary students' statistical reasoning ability in descriptive statistics. Procedia-Social and Behavioral Sciences, *129*, 133-139.

Cheng, P. & Holyoak, K. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391-416.

Chiesi, F., Primi, C., & Morsanyi, K. (2009). *The effects of education, instructions and cognitive abilities on probabilistic reasoning: A test of a theory*. Society for Research in Child Development. Paper presented in SRCD, Biennial Meeting, Denver.

Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' achievement. *Statistics Education Research Journal*, *9*(1), 6-26.

Chou, C., and Bentler, P.M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.). *Structural equation modeling: Concepts, issues, and application* (pp. 37-55) Thousand Oaks, CA: Sage Publications.

Cobb, G. (1998). *The Objective-Format Question in Statistics: Dead Horse, Old Bath Water, or Overlooked Baby?* Paper presented in the Annual Meeting of American Educational Research Association, San Diego, CA.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112* (1), 155-159.

Cohen, S., Smith, G., Chechile, R. A., Burns, G., & Tsai, F. (1996). Identifying impediments to learning probability and statistics from an assessment of instructional software. *Journal of Educational and Behavioural Statistics, 21*(1), 35–54.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1-73.

Coştu, S., Serhat A., & Mehmet, F. (2009). Students' conceptions about browser-game-based learning in mathematics education: TTNetvitamin case. *Procedia-Social and Behavioral Sciences*, *1*(1), 1848-1852.

Crane, J. & Hannibal, J. (2009). *Psychology: Course companion*. Oxford: Oxford Press.

Creswell, J. W. (2009). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, California, CA: SAGE Publications, Inc.

Danili, E. & Reid, N. (2006). Cognitive factors that can potentially affect pupils' test performance. *Chemistry Education Research and Practice*, *7*, 64-83.

Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, *44*, 20-33.

Darling-Hammond, L. & Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning.* Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Dawson, J.F. (2014). Moderation in management research: what, why, when, and how. *Journal of Business Psychology*, *29*, 1–19.

delMas, R. C. (2002). Statistical literacy, reasoning, and learning. *Journal of Statistics Education*, *10*(3). Retrieved from http://www.amstat.org/publications/jse/v10n3/delmas_intro.html.

delMas, R. C. (2004a). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79-95). Dordrecht, The Netherlands: Kluwer Academic Publishing.

delMas, R. C. (2004b). Overview of ARTIST website and Assessment Builder. *Proceedings of the ARTIST Roundtable Conference*, Lawrence University. Retrieved from http://www.rossmanchance.com/artist/Proctoc.html

delMas, R. C. & Garfield, J. (1991). Using multiple items to assess misconceptions. In *Research Papers from ICOTS III*, International Study Group for Research in Learning Probability and Statistics.

delMas, R. C., Garfield, J., & Ooms, A. (2005). Using assessment to study students' difficulty reading and interpreting graphical representations of distributions. In K. Makar (Ed.), *Proceedings of the Fourth International Research Forum on Statistical Reasoning, Literacy, and Reasoning* (on CD). Auckland, New Zealand: University of Auckland.

delMas, R. C., Ooms, A., Garfield, J., & Chance, B. (2006). Assessing students' statistical reasoning. *Proceedings of the Seventh International Conference on Teaching Statistics*. Salvador de Bahia, Brazil: International Association of Statistics Education and International Statistical Institute. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/17/6D3_DELM.pdf

Dempster, M., & McCorry, N.K. (2009). The role of previous experience and attitudes toward statistics in statistics assessment outcomes among undergraduate psychology students. *Journal of Statistics Education*, *17*(2). Retrieved from www.amstat.org/publications/jse/v17n2/dempster.html.

Dennett, D. C. (1998). *Brainchildren: Essays on designing minds*. Cambridge, Massachusetts: MIT Press.

Ding, C. S., Song, K., & Richardson, L. I. (2006). Do mathematical gender differences continue? A longitudinal study of gender difference and excellence in mathematics performance in the U.S. *Educational Studies*, *40*(3), 279-295

Dwyer, C.A. (1973). Sex differences in reading: An evaluation and a critique of current theories. *Review of Educational Research, 43*, 455–467.

Elmore, P.B., & Vasu, E.S. (1986). A model of statistics achievement using spatial ability, feminist attitudes and mathematics-related variables as predictors. *Educational and Psychological Measurement*, *46*, 215-222.

Ericsson, K. A. (2003). The acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice. In J. E. Davidson and R. J. Sternberg (Eds.). *The psychology of problem solving* (pp. 31–83). Cambridge, England: Cambridge University Press.

Evans, J. St. B. T. (2007). *Hypothetical thinking: dual processes in reasoning and judgement*. Hove: Psychology Press.

Evans, J. St. B. T. (2008). Dual processing accounts of reasoning, judgement and social cognition. *Annual Review of Psychology*, *59*, 255-278.

Evans, J. St. B. T. & Over, D. E. (1996). *Rationality and reasoning.* Hove: Psychology Press.

Eysenck, M.W. & Keane, M.T. (2015). *Cognitive psychology: a student's handbook*. (7th ed.). New York, NY: Psychology Press.

Feinberg, L. B., & Halperin, S. (1978). Affective and cognitive correlates of course performance in introductory statistics. *The Journal of Experimental Education*, *46*(4), 11-18.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: Sage Publications Ltd.

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling*, *2*, 87–105.

Fischbein, E. (1999). Intuitions and schemata in mathematical reasoning, *Educational Studies in Mathematics*, *38*, 11-50.

Fischbein, E. and Grossman, A.(1997). Schemata and intuitions in combinatorial reasoning. *Educational Studies in Mathematics*, *34*, 27-47.

Foo, K.K. (2011). *Null hypothesis significance testing: An Asian perspective*. Shah Alam, Malaysia: Pusat Penerbitan Universiti, Universiti Teknologi MARA.

Foo, K. K. & Noraini Idris. (2010). A comparative study on statistics competency level using TIMSS data: Are we doing enough? *Journal of Mathematics Education*, *3*(2), 126-138.

Franklin, C., & Garfield, J. (2006). The Guidelines for assessment and instruction in statistics education (GAISE) project: Developing statistics education guidelines for pre K-12 and college courses. In G. F. Burrill (Ed.), *Thinking and reasoning about data and chance* (Vol. 68, pp. 345-375). Reston, VA: National Council of Teachers of Mathematics.

Gal, I. (2004). Statistical literacy, meanings, components, responsibilities. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 47-78). Dordrecht, The Netherlands: Kluwer Academic Publishing

Gal, I., & Garfield, J. (Eds.) (1997) *The assessment challenge in statistics educatio*n. Amsterdam: IOS Press.

Gal, I., Ginsburg, L. & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education,* (pp. 37-51). Amsterdam: IOS Press and the International Statistical Institute.

Galagedera, D. (1998). Is remedial mathematics a real remedy? Evidence from learning statistics at tertiary level. *International Journal of Mathematical Education in Science & Technology, 29* (4), 475 - 480.

Galotti, K. M. (2008). *Cognitive psychology: In and out of the laboratory* (4th ed.). Singapore: Thomson Wadsworth.

Gardner, P. L., and Hudson, I. (1999). University students' ability to apply statistical procedures. Journal of Statistics Education, *7*(1). Retrieved from http://www.amstat.org/publications/jse/secure/v7n1/gardner.cfm

Garfield, J. (1994). Beyond testing and grading: Using assessment to improve student learning. *Journal of Statistics Education*, *2*(1). Retrieved from http://www.amstat.org/publications/jse/v2n1/garfield.html

Garfield, J. (1998). The statistical reasoning assessment: Development and validation of a research tool. In Proceedings of the Fifth International Conference on Teaching Statistics, L. Pereira-Mendoza (Ed.), Voorburg, The Netherlands: International Statistical Institute, (pp. 781-786).

Garfield, J. (2002) The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education*, *10*(3). Retrieved from www.amstat.org/publications/jse/v10n3/garfield.html

Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, *2*(1), 22-38. Retrieved from www.stat.auckland.ac.nz/%7Eiase/serj/SERJ2(1).pdf.

Garfield, J. B., & Ahlgren, A. (1988). Difficulties in learning: Implications for research. *Journal for Research in Mathematics Education*, *19*, 44-63.

Garfield, J. B., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 397-409). Dordrecht, The Netherlands: Kluwer Academic Publishing.

Garfield, J. B., & Ben-Zvi, D. (Eds.) (2005). Reasoning about variation [Special section]. *Statistics Education Research Journal*, *4*(1). Retrieved from http://www.stat.auckland.ac.nz/serj

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer Science & Business Media.

Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematics Thinking and Learning*, *2*(1-2), 99-125.

Garfield, J., delMas, R., & Chance, B. (2002). *The Assessment Resource Tools for Improving Statistical Thinking (ARTIST) Project*. NSF CCLI grant ASA-0206571. Retrieved from https://app.gen.umn.edu/artist/

Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. K. C. Lewis (Ed.), *A handbook for data analysis in the behavioural sciences: Methodological issues* (pp. 311-339). Hillsdale, NJ: Erlbaum.

Gigerenzer, G. (1998). We need statistical thinking, not statistical rituals. *Behavioural and Brain Sciences, 21*, 199-200.

Gigerenzer, G. & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103*, 650–669.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684-704.

Giraud, G. (1997), Cooperative Learning and Statistics Instruction. *Journal of Statistics Education*, *5*(3). Retrieved from www.amstat.org/publications/jse/v5n3/giraud.html

Girotto, V. (2004). Task understanding. In J. P. Leighton & R. J.Sternberg (Eds.), *The nature of reasoning* (pp. 103–125). New York: Cambridge University Press.

Glöckner, A., & Witteman, C. (2010). Beyond dual-process models: A categorisation of processes underlying intuitive judgement and decision making. *Thinking & Reasoning*, *16*(1), 1-25.

Goldberger, A.S. & Duncan, O.D. (1973). *Structural equation models in the social sciences.* New York: Seminar Press.

Gonzales, P., Guzmán, J.C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., and Williams, T. (2004). Highlights from the trends in international mathematics and science study (TIMSS) 2003 (NCES 2005-005). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Gutman, A. (1979). Misconceptions of psychology and performance in the introductory course. *Teaching of Psychology, 6*, 159-161.

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly, 28*(2), 116–159

Hailikari, T. (2009). *Assessing university students' prior knowledge: Implications for theory and practice*. Department of Education, Research Report 227. University of Helsinki.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C. (1999) *Multivariate data analysis* (5th ed.). Upper Saddle River, New Jersey: Prentice Hall.

Haller, H., & Krauss, S. (2002). Misinterpretations of Significance: A problem students share with their teachers? Methods of Psychological Research – Online [Online serial], *7* (1), 1-20.

Hardman, D & Macchi, L. (2003). *Thinking: Psychological perspectives on reasoning, judgment, and decision making*. Wiley & Sons.

Hardre, P.L., Chen, C.H., Huang, S.H., Chiang, C.T., Jen, F.L., & Warden, L. (2006). Factors affecting high school students' academic motivation in Taiwan. *Asia Pacific Journal of Education*, *26* (2), 189-207.

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioural Decision Making, 12*(4), 275 - 305.

Hirsch, L., & O'Donnell, A. M. (2001). Representativeness in statistical reasoning: identifying and assessing misconceptions. *Journal of Statistics Education*, 9(2). Retrieved from http://www.amstat.org/publications/jse/v9n2

Hubbard, R. (1997). *Assessment and the Process of Learning Statistics. Journal of Statistics Education.* 5(1). Retrieved from http://www.amstat.org/publications/jse/v5n1/hubbard.html

Huck, D. W. (2004). *Reading Statistics and Research*. In NCTM (Ed.), Teaching Statistics and Probability *(4th ed.). NCTM 1981 Yearbook.* Boston: Pearson Education Inc.

Huitt, W. (2003). The information processing approach to cognition. *Educational Psychology Interactive*. Valdosta, GA: Valdosta State University. Retrieved from, http://www.edpsycinteractive.org/topics/cognition/infoproc.html

Hulsizer, M. R., & Woolf, L. M. (2008). *Guide to teaching statistics: Innovations and best practices*. Wiley-Blackwell.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*(5888), 494-495.

International Association for the Evaluation of Educational Achievement (IEA). (2009). *Trends in International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: IEA TIMSS & PIRLS International Study Center.

International Association for the Evaluation of Educational Achievement (IEA). (2013). *Trends in International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: IEA TIMSS & PIRLS International Study Center. Retrieved from http://timssandpirls.bc.edu/timss2011/international-results-mathematics.html

Johnson-Laird, P.N. (2006). *Promoting academic achievement and motivation: A discussion & contemporary issues based approach*. Oxford: Oxford University Press.

Kahneman, D. (1991). Judgment and decision making: A personal view. *American Psychological Society*. *2*(3): 142-146

Kahneman, D. Slovic, P. & Tversky, A. (1982). *Judgement under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review,* 80, 237– 57.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.

Kalat, J.W. (2011) *Introduction to Psychology* (9^{th} ed.). Wadsworth Publishing.

Kersten, D., Mamassian, P. & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.

Kline, R. B. (1998). *Principles and practice of structural equation modelling* (2nd ed.). New York: Guilford Press.

Knofczynski, A. & Mundfrom, D. (2008). Sample sizes when using multiple linear regression for prediction. In Gregory T. (Ed) *Educational and Psychological Measurement, 68* (3):431-442: Sage Publications

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6*(1), 59-98.

Konold, C. (1991). Understanding students' beliefs about probability. In E. V. Glaserfeld (Ed.), *Radical constructivism in mathematics education*. (pp. 139-156) Dordrecht, The Netherlands: Kluwer Academic Publishers.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, *3*(1). Retrieved from http://www.amstat.org/publications/jse/v3n1/konold.html

Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin & D. Schifter (Eds.), *A Research Companion to Principles and Standards for School Mathematics* (pp. 193-215). Reston, VA: National Council of Teachers of Mathematics.

Konold, C., Pollstek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education*, *24*(5), 392–414.

Kooi, L. T., & Ping, T. A. (2006). Factors Influencing Students Performance in Wawasan Open University: Does Previous Education Level, Age Group and Course Load Matter?. Retrieved from http://www1.open.edu.cn/elt/23/2.html.

Lalonde, R. N., & Gardner, R. C. (1993). Statistics as a second language? A model for predicting performance in psychology students. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *25*(1), 108.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics, 23*, 557-568.

Lèonard and Sackur-Grisvard (1987). Necessity of a triple approach of erroneous conceptions of students, example of the teaching of relative numbers: Theoretical analysis. In J.C. Bergeron (Ed.). *Proceedings of the Eleventh International Conference for the Psychology of Mathematics Education* (Vol. 2, pp. 444-448). Montreal.

Levitin, D.J. (2002). *Foundations of cognitive psychology: Core readings.* Bradford Book.

Li, J. (2007). *Regression diagnostics for complex survey data*. (Unpublished doctoral dissertation), University of Maryland

Lipson, A. (1990). Learning: A momentary stay against confusion. *Teaching and Learning*. *The Journal of Natural Inquiry*, *4*, 2-11.

Lipson, K. (2002). *The role of computer based technology in developing understanding of the concept of sampling distribution.* In the Proceedings of the sixth international conference on teaching statistics, Voorburg, The Netherlands.

Liu, H. J. (1998). *A cross-cultural study of sex differences in statistical reasoning for college students in Taiwan and the United States*. (Doctoral dissertation). University of Minnesota, Minneapolis. Brandsford

Liu, H. J. & Garfield, J. B. (2002). Sex differences in statistical reasoning. *Bulletin of Educational Psychology*, *32*, 123-138.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.

Loftus, E.F. (2003). Make-believe memories. *American Psychology*, *58*, 864–873

Ministry of Education (MOE). (2012). *Malaysia Education Blueprint (2013-2025): Preliminary Report and Executive Summary.* Malaysia Ministry of Education: Kuala Lumpur. Retrieved from http://www.moe.gov.my/userfiles/file/PPP/Preliminary-Blueprint-Eng.pdf.

Manitoba Education, Citizenship and Youth (2006). *Rethinking classroom assessment with purpose in mind.* Ministry of Education, Citizenship and Youth: Manitoba.

Martin, N. (2013) *Exploring the mechanisms underlying gender differences in statistical reasoning: A multipronged approach*. (Unpublished PhD thesis.) University of Waterloo, Ontario, Canada.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*, 391-410.

Maxwell, S. E. (2000). Sample size and multiple regression analysis. *Psychological Methods*, *5*, 434-458

MacCallum, R. C., & Austin, J. T. (2000). Applications of Structural Equation Modeling in psychological research. *Annual Review of Psychology*, *51*(1), 201-226.

McLeod, S. A. (2008). Qualitative vs. Quantitative. Retrieved from www.simplypsychology.org/qualitative-quantitative.html

McLeod, S. A. (2009). Eyewitness Testimony. Retrieved from www.simplypsychology.org/eyewitness-testimony.html

McCutcheon, L. E. (1991). A new test of misconceptions about psychology. *Psychological Reports*, *68*, 647-653.

Mercier, H. (2013). The function of reasoning: Argumentative and pragmatic alternatives. *Thinking and Reasoning, 19*(3-4), 488-494.

Mercier, H. & Sperber, D. (2009) Intuitive and reflective inferences. In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 149 – 70). Oxford University Press.

Mercier, H, & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioural And Brain Sciences, 34*, 57–111.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97. Retrieved from http://www.musanim.com/miller1956

Miller, K. (1999). Which assessment type should be encouraged in professional degree courses - continuous, project-based or final examination-based? In K. Martin, N. Stanley & N. Davison (Eds.), *Teaching in the Disciplines/ Learning in Context*, (pp. 278-281). Proceedings of the 8th Annual Teaching Learning Forum, The University of Western Australia, Perth: UWA. Retrieved from http://cleo.murdoch.edu.au/asu/pubs/tlf/tlf99/km/miller.html

Miller, N. (1997). *Assessment: Alternative forms of formative & summative assessment*. *The Handbook for Economics Lecturers.* Glasgow: Caledonian University.

Moore, D. S. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy*. (pp. 95-137) Washington, D.C: National Academy Press.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J. and Smith, T.A. (2000) *TIMSS 1999 International Mathematics Report. Findings from IEA's repeat of the Third International Mathematics and Science Study at the eighth grade*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, Foy, P., Olson, J.F., Preuschoff, C., Erberber, E., Arora, A., and Galia, J. (2008). TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College

Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International results in mathematics.* Chestnut Hill, MA: Timss & Pirls International Study Center, Boston College. Retrieved from http://timss.bc.edu/timss2011/downloads/T11_IR_Mathematics_FullBook.pdf

Nasser, F. M. (1999). Prediction of statistics achievement. *In Proceedings of the International Statistical Institute 52nd Conference*, (Vol 3, pp. 7-8 Helsinki, Finland.

Nasser, F. M. (2004). Structural model of the effects of cognitive and affective factors on the achievement of Arabic-speaking pre-service teachers in introductory statistics. *Journal of Statistics Education*, *12*(1), 1-28.

National Council of Teachers of Mathematics (NCTM). (1995). *The assessment standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Newton, D. P. (2000). *Teaching for understanding: What it is and how to do it.* London: Routledge Falmer.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgments.* Englewood Cliffs, NJ: Prentice Hall.

Noor Azina Ismail & Azmah Othman (2008). Comparing university academic performances of HSC students at the three Art-based faculties. *International Education Journal*, *7*(5), 668-675.

Noraidah Sahari Ashaari, Hairulliza Mohamad Judi, Hazura Mohamed & Tengku Siti Meriam Tengku Wook. (2011). Student's attitude towards statistics course. *Procedia Social Behavioral Sciences*, *18*, 287-294.

Oakes, M. (1986). Statistical inference: A commentary for the social and behavioral sciences. Chichester: Wiley.

Organisation for Economic Cooperation and Development. (OECD) (2001), *Knowledge and skills for life* – First Results from PISA 2000, OECD, Paris.

Organisation for Economic Cooperation and Development. (OECD) (2004). *Learning for tomorrow's world. First results from PISA 2003*. Paris. Retrieved from http://www.pisa.oecd.org/dataoecd/1/60/34002216.pdf

Organization for Economic Cooperation and Development. (OECD). (2010). *PISA 2009 results: What makes a successful school?* Resources, Policies, and Practices (Volume 4). Paris.

Organisation for Economic Cooperation and Development. (OECD). (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: Author. Retrieved from http://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20e-book_final.pdf

Olivier, A. (1989). Handling pupils' misconception. *Pythagoras*. *21*, 10-19.

Onwuegbuzie, A. J., & Seaman, M. A. (1995). The effect of time constraints and statistics test anxiety on test performance in a statistics course. *The Journal of Experimental Education*, *63*(2), 115-124.

Ooms, A. (2005). *The iterative evaluation model for improving online educational resources*. (Doctoral Thesis). University of Minnesota.

Orey, M. (2001). Information processing. In M. Orey (Ed.), *Emerging perspectives on learning, teaching, and technology*. Retrieved from http://epltt.coe.uga.edu/

Overton, T. (2008) *Assessing learners with special needs* (6th ed.). Prentice Hall.

Pappas, C. (2014)**.** Frederic Bartlett 's Schema Theory**.** Retrieved from http://elearningindustry.com/schema-theory.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: the science and design of educational assessment*. Washington, DC: National Academy Press.

Pfannkuch, M. (2005). Probability and statistical inference: how can teachers enable learners to make the connection? In G.A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 267-294). New York: Springer.

Pfannkuch, M., & Wild, C. J. (2003). *Statistical Thinking: How can we develop it?* In the 54th International Statistical Institute Conference. Reston: NCTM Inc

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 17-46). Dordrecht, The Netherlands: Kluwer Academic Publishing.

Piattelli-Palmarini, M. (1994). Ever since language and learning: afterthoughts on the Piaget-Chomsky debate. *Cognition. 50*, 315-346.

Pinilla, B., & Munoz, S. (2005). *Educational opportunities and academic performance: A case study of university student mothers in Venezuela*. Higher Education. Volume 50(2).

Pinker, J. (1999). *Words and rules.* New York: Harper-Collins.

Plotnik, R. & Kouyoumdjian, H. (2011) *Introduction to Psychology* (9th ed.). Wadsworth Publishing

Radke-Sharpe, N. (1998). *Assessment issues in introductory and advanced statistics courses.* Paper presented in the Joint Statistical Meetings, Dallas, TX.

Randolph, K. A., & Myers, L. L. (2013). *Basic statistics in multivariate analysis*. New York: Oxford University Press

Reading, C. (2002). The International Research Forum on Statistical Reasoning, Thinking and Literacy: Summaries of presentations at STRL-2. *Statistics Education Research Journal,1*(1), 30-45. Retrieved from http://fehps.une.edu.au/f/s/curric/creading/serj/past_issues/SERJ1(1).pdf)

Reed, D. K. (2011). A review of the psychometric properties of retell instruments. *Educational Assessment*, *16*, 123–144

Reed, S.K. (2013). *Cognition: Theories and application* (9th ed.). Wadsworth: Cengage Learning.

Reilly D (2012) *Gender, culture, and sex-typed cognitive abilities*. PLoS ONE 7(7): e39904. doi:10.1371/journal.pone.0039904

Riegler, B., & Riegler, G.L. (2004). *Cognitive psychology: Applying the science of the mind*. Allyn & Bacon.

Rienties, B., Tempelaar, D., Bossche, P. V., Gijselaers, W., & Segers, M. (2009). The role of academic motivation in computer-supported collaborative learning. *Computers in Human Behavior*, *25*(6), 1195-1206.

Roseth, C. J. Garfield, J. B., & Ben-Zvi, D. (2008). Collaboration in learning and teaching statistics. *Journal of Statistics Education* [Online], 16(1). Retrieved from http://www.amstat.org/publications/jse/v16n1/roseth.html.

Rubin, A., Bruce, B., & Tenney, Y. (1991). *Learning about sampling: Trouble at the core of statistics.* Paper presented in the Third International Conference on Teaching Statistics. Voorburg, The Netherlands.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. T. Spiro, B. C. Bruce and W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Erlbaum, Hillsdale, NJ.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press.

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, *10*(3). Retrieved from http://www.amstat.org/publications/jse/v10n3/rumsey2.html

Saldanha, L. A. (2004). *'Is this sample unusual?' An investigation of students exploring connections between sampling distribution and statistical inference*. (Unpublished doctoral thesis.) Vanderbilt University.

Schau, C., & Mattern, N. (1997). Assessing students' connected understanding of statistical relationships. In I. Gal, & Garfield, J. B. (Ed.), *The Assessment Challenge in Statistics Education* (pp. 91-104). Amsterdam: IOS Press.

Schoenfeld, A. H. (1985). *Mathematical problem solving.* Orlando, FL: Academic Press.

Schram, C. (1996). A meta-analysis of gender differences in applied statistics achievement. *Journal of Educational and Behavioral Statistics*, *21*, 55 - 70.

Schraw, G., Flowerday, T., & Lehman, S. (2001). Increasing situational interest in the classroom. *Educational Psychology Review*, *13*(3), 211-224.

Schraw, L., & McCrudden, M. (2013). *Information processing theory*. Retrieved from Education.Com: http://www.education.com/reference/article/information-processing-theory/

Schwartz, D. L., Goldman, S.R., Vye, N.J. & Barron, B.J. (1998). Aligning everyday and mathematical reasoning: The case of sampling assumptions. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 233-273). NJ: Lawrence Erlbaum Associates, Mahwah.

Schwartz, S.J. (2001). The evolution of Eriksonian and neo-Eriksonian identity theory and research: A review and integration. *Identity*, *1*, 7–58.

Seddon, G. (1978). The properties of Bloom's taxonomy of educational objectives for the cognitive domain. *Review of Educational Research, 48*(2), 303-323.

Sedlmeier, P. (1999). *Improving statistical reasoning. Theoretical models and practical implication*. Mahwah, NJ: Erlbaum.

Shaughnessy, J. M. (1981). Misconceptions of probability: From systematic errors to systematic experiments and designs in teaching probability and statistics. In NCTM 1981 Yearbook. (pp. 90-99) Reston, VA: National Council of Teachers of Mathematics.

Shaughnessy, J. M. (1981b). Teaching and Learning specific topics. In NCTM (Ed.), *Teaching statistics and probability*. NCTM 1981 Yearbook. Reston. VA: NCTM Inc.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In A. Grouws (Ed.), *Handbook of research on mathematics teaching and* learning (pp. 465-494). New York: MacMillan Publishing Company.

Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review*, *63*,129–138.

Smith, J. P., III, diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, *3*(2), 115–163.

Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W. & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review 2*, 98-113.

Sousa, D. A. (2008). How the brain learns mathematics. Thousand Oaks, CA: Corwin Press

Stanovich, K. E. (1999) *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

StatPac (2010). StatPac user's guide. StatPac Inc. 1200 First Street, Pepin, WI 54759. Retrieved from https://statpac.com/manual/index.htm?turl=collinearitydiagnostics.htm.

Sternberg, R. J. (2001). Metacognition, abilities, and developing expertise: What makes an expert student? In H. J. Hartman (Ed.), *Metacognition in learning and instruction: Theory, research, and practice* (pp. 247–260). Dordrecht, The Netherlands: Kluwer.

Sundre, D. L. (2003), *Assessment of Quantitative reasoning to enhance educational quality*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago. Retrieved from http://www.gen.umn.edu/artist/articles/AERA_2003_QRQ.pdf

Suthers, D. (1996). Attention and automaticity. Pittsburgh: University of Pittsburg, Learning Research and Development Center. Retrieved from http://www.pitt.edu/~suthers/infsci1042/attention.html

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.

Tempelaar, D. (2004). Statistical reasoning assessment: An Analysis of the SRA instrument. In Proceedings of the ARTIST Roundtable Conference, Lawrence University. Retrieved from http://www.rossmanchance.com/artist/Proctoc.html

Tempelaar, D. (2006). A structural equation model analyzing the relationship students' statistical reasoning abilities, their attitudes toward statistics, and learning approaches. In Proceedings of the 7th International Conference on the Teaching of Statistics. Retrieved from http://iase-web.org/documents/papers/icots7/2G3_TEMP.pdf

Tempelaar, D., Gijselaers, W. H., & Van der Loeff, S. (2006). Puzzles in statistical reasoning. *Journal of Statistics Education*, *14*(1) 1-26. Retrieved from http://www.amstat.org/publications/jse/v14n1/tempelaar.html

Tempelaar, D., Van der Loeff, S., & Gijselaers, W.H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities and course performance. *Statistics Education Research Journal, 6*(2), 78-102. Retrieved from http://www.stat.auckland.ac.nz/serj

Tremblay, P. F., Gardner, R. C., & Heipel, G. (2000). A model of the relationships among measures of affect, aptitude, and performance in introductory statistics. *Canadian Journal of Behavioral Science*, *32*, 40-48.

Trochim, W. (2006). *The research methods knowledge base*. (2nd ed.). Retrieved from the Internet at http://www.socialresearchmethods.net/kb

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

Van Merrienboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147-177.

Ware, M. E., & Chastain, J. D. (1991). Developing selection skills in introductory statistics. *Teaching of Psychology, 18*(4), 219–222.

Watson, J. M. (1997). Assessing statistical thinking using the media. In Gal & Garfield (Eds.), *The assessment challenge in statistics education.* Amsterdam: IOS Press.

Watson, J.M. (2009). The influence of variation and expectation on the developing awareness of distribution. *Statistics Education Research Journal*, 8(1), 32-61. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications.php?show=serjarchive.

Wild, C., Triggs, C., & Pfannkuch, M. (1997). Assessment on a budget: using traditional methods imaginatively. In Gal & Garfield (Eds.), *The assessment challenge in statistics education.* Amsterdam: IOS Press.

Wild, C.J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*, 223–265.

Wilkins, J. L., & Ma, X. (2002). Predicting student growth in mathematical content knowledge. *The Journal of Educational Research*, *95*, 288-298.

Wolpert, D. M. & Kawato, M. (1998) Multiple paired forward and inverse models for motor control. *Neural Networks, 11*(7–8), 17–29.

Wu, A. D. & Zumbo, B. D. (2008). Understanding and using mediators and moderator. *Social Indicator Research*, *87*, 367–392

York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. Practical Assessment, Research, and Evaluation, 20 (5). Available online: http://pareonline.net/getvn. asp?v=20&n=5

Zuraida Jaafar, Foo, K. K., Rosemawati Ali. & Haslinda Abdul Malek. (2012) Cognitive factors influencing statistical performance of diploma science students: A structural equation model approach. *In Proceedings of Langkawi Conference-ICSSBE*. (pp 562-566)

Zamalia Mahmud & Nor Hasmaniza Osman (2010). Statistical competency and attitude towards learning elementary statistics: A case of SMK Bandar Baru Sg Buloh. *In Proceedings of the Regional Conference on Statistical Sciences* (RCSS'10) (pp 335-348).

Zimmerman, B. J. (2000). Attaining self-regulation: a social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation*. (pp. 13-39) San Diego: CA: Academic Press.