CONSTRUCTION OF MATHEMATICAL MODELS IN LANGUAGE LEARNING STRATEGY AND ENGLISH PROFICIENCY OF PRE-UNIVERSITY STUDENTS

JOHANNAH BINTI JAMALUL KIRAM

PERPUSTAKAAN
UNIVERSITI MALAYSIA SABAH

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

FACULTY OF SCIENCE AND NATURAL RESOURCES
UNIVERSITI MALAYSIA SABAH
2016
JUDUL: CONSTRUCTION OF MATHEMATICAL MODELS IN LANGUAGE LEARNING STRATEGY AND ENGLISH PROFICIENCY OF PRE-UNIVERSITY STUDENTS

IJAZAH: SARJANA SAINS

Saya, JOHANNAH BINTI JAMALUL KIRAM Sesi Pengajian 2013-2015 mengaku membenarkan tesis Sarjana ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis ini adalah hak milik Universiti Malaysia Sabah.
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagain bahan pertukaran antara pengajian tinggi.
4. Sila tandakan (/)

☐ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

☐ TERHAD (Mengandungi Maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

☐ TIDAK TERHAD

Alamat Tetap: No. 5, Lorong Aman 2, Taman Aman, 88450 Kota Kinabalu, Sabah.

Tarikh: 25 Februari 2016
DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, equations, summaries and references, which have been duly acknowledged.

Date: 4 MARCH 2016

Johannah Binti Jamalul Kiram
MS1311005T
CERTIFICATION

NAME : JOHANNAH BINTI JAMALUL KIRAM

MATRIC NO. : MS1311005T

TITLE : CONSTRUCTION OF MATHEMATICAL MODELS IN LANGUAGE LEARNING STRATEGY AND ENGLISH PROFICIENCY OF PRE-UNIVERSITY STUDENTS

DEGREE : MASTER OF SCIENCE (MATHEMATICS)

VIVA DATE : 27 JANUARY 2016

DECLARED BY

1. MAIN SUPERVISOR
   ASSOCIATE PROFESSOR DR. JUMAT SULAIMAN

[Signature]

2. CO SUPERVISOR
   ASSOCIATE PROFESSOR DR. SUYANSAH SWANTO

[Signature]
ACKNOWLEDGEMENTS

First and foremost, Alhamdulillah. I would like to express my sincere thanks to those who have helped me throughout my study in Universiti Malaysia Sabah.

My deepest gratitude goes to my supervisors, Assoc. Prof. Dr. Jumat Sulaiman and Assoc. Prof. Dr. Suyansah Swanto for their never ending support, guidance and continuous encouragement throughout the course of my study.

Special thanks goes to Puan Wardatul Akmam Din for helping me with gathering the data. May Allah bless her.

Lastly, for my family. To my parents, for the understanding and support. To my husband, Awang Mohamad Yazrif, thank you for always being by my side through good times and hard times.

Johannah Binti Jamalul Kiram
MS1311005T
ABSTRACT

This study aims to construct the best mathematical model that describes the relationship between six Language Learning Strategies and English proficiency of the pre-university students of Universiti Malaysia Sabah. Two hundred and thirty pre-university students of Universiti Malaysia Sabah took part in this study by answering a background questionnaire and the Strategy Inventory for Language Learning self-report questionnaire. Also, these students sat for the Malaysian University English Test (MUET) and their results are taken as the student’s proficiency in English as a second language. Three linear regression models was used as initial models, two of which went through stepwise variable selection method. However, in this study, we proposed two families of nonlinear models, which are the nonlinear Gompertz model, and a modified Gompertz model. Based on the five proposed models, the dependent variable, $y$, is the normalized English proficiency of the students and the independent variables, $X_j$, are the six different language learning strategies where $j = 1, 2, 3, ..., 6$. These six strategies are memory, cognitive, comprehensive, metacognitive, social and affective respectively. Goodness-of-fit tests and information criterion were used to assess these models, which are the root mean square error (RMSE), mean absolute error (MAE), residual standard error (RSE), the corrected Akaike information criterion ($AIC_c$) and Bayesian information criterion (BIC). The results were then used to compare these models. The linear model showed good readings with its RMSE, MAE and RSE approaching zero indicating that the model is strong. However, the Gompertz model and the modified Gompertz model showed even better values of RMSE, MAE, RSE and $AIC_c$. Especially for the modified Gompertz model, the values for the errors were even closer to zero. The goodness-of-fit tests and $AIC_c$ show that the modified Gompertz model’s results were better than the nonlinear Gompertz model, and the linear regression model. Despite that, only the BIC calculated for the linear model M35 has the weightiest amount of information. In conclusion, four out of five metrics calculated suggested that the modified Gompertz model is the best amongst investigated models.
## LIST OF CONTENTS

<table>
<thead>
<tr>
<th>TITLE</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DECLARATION</td>
<td>ii</td>
</tr>
<tr>
<td>CERTIFICATION</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF CONTENTS</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF APPENDIXES</td>
<td>xiv</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xv</td>
</tr>
<tr>
<td>LIST OF SYMBOLS</td>
<td>xvi</td>
</tr>
</tbody>
</table>

## CHAPTER 1: INTRODUCTION

1.1 Malaysian Education System and Language Learning Strategy  1
1.2 English Language Proficiency in Malaysia  3
1.3 Language Learning Strategies  4
1.4 Development of Mathematical Models  6
1.5 Types of Models  8
    1.5.1 Time-Series Models  10
    1.5.2 Judgmental Models  11
    1.5.3 Cause-and-Effect Models  11
1.6 Regression Analysis  12
3.3 Models and Its Assumptions

3.3.1 Linear Regression Model

3.3.2 Nonlinear Regression Model
   a. Linearization of the Gompertz model
   b. Linearization of the modified Gompertz Model

3.4 Preparation of Normalized Data

3.5 Goodness-Of-Fit tests and Information Criterion

3.5.1 Root Mean Square Error

3.5.2 Mean Absolute Error

3.5.3 Residual Standard Error

3.5.4 The Corrected Akaike’s Information Criterion

3.5.5 Bayesian’s Information Criterion

CHAPTER 4: LINEAR MODELS

4.1 Introduction

4.2 Construction of Linear Model

4.3 Linear Model Assumptions Tests
   4.3.1 M63 Assumptions Test
   4.3.2 M45 Assumptions Test
   4.3.3 M35 Assumptions Test

4.4 Goodness-Of-Fit Tests and Information Criterion
   4.4.1 M63 Adequacy Test
   4.4.2 M45 Adequacy Test
   4.4.3 M35 Adequacy Test

4.5 Summary
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Examples of nonlinear models</td>
<td>16</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>ANOVA for Global Test</td>
<td>44</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>ANOVA of Global test for M63</td>
<td>56</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>ANOVA of Global test for M45</td>
<td>57</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>ANOVA of Global test for M35</td>
<td>57</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Correlation coefficients of the variables</td>
<td>58</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>Estimates for the unknowns, $\beta_j$ for model M63</td>
<td>60</td>
</tr>
<tr>
<td>Table 4.6</td>
<td>Estimates for the unknowns, $\beta_j$ for model M45</td>
<td>60</td>
</tr>
<tr>
<td>Table 4.7</td>
<td>Estimates for the unknowns, $\beta_j$ for model M35</td>
<td>61</td>
</tr>
<tr>
<td>Table 4.8</td>
<td>The Variance Inflation Factor of each strategy for M63</td>
<td>65</td>
</tr>
<tr>
<td>Table 4.9</td>
<td>The Variance Inflation Factor of each strategy for M45</td>
<td>68</td>
</tr>
<tr>
<td>Table 4.10</td>
<td>The Variance Inflation Factor of each strategy for M35</td>
<td>72</td>
</tr>
<tr>
<td>Table 4.11</td>
<td>Comparison of goodness-of-fit between the linear models</td>
<td>78</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Estimates for the unknowns, $\beta_j$ for the Gompertz model</td>
<td>81</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Estimates for the unknowns, $\beta_j$ for the Modified Gompertz model</td>
<td>83</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Comparison of goodness-of-fit between all five models</td>
<td>93</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Work flow in a regression analysis</td>
<td>8</td>
</tr>
<tr>
<td>1.2</td>
<td>Types of Mathematical models</td>
<td>9</td>
</tr>
<tr>
<td>3.1</td>
<td>Steps in getting the best Multiple Regression model</td>
<td>40</td>
</tr>
<tr>
<td>4.1</td>
<td>Residual versus fitted values for M63</td>
<td>64</td>
</tr>
<tr>
<td>4.2</td>
<td>Absolute studentized residuals versus fitted values for M63</td>
<td>64</td>
</tr>
<tr>
<td>4.3</td>
<td>Normal Q-Q plot for M63</td>
<td>65</td>
</tr>
<tr>
<td>4.4</td>
<td>Histogram of frequency versus residual for M63</td>
<td>66</td>
</tr>
<tr>
<td>4.5</td>
<td>Residual versus fitted values for M45</td>
<td>67</td>
</tr>
<tr>
<td>4.6</td>
<td>Absolute studentized residuals versus fitted values for M45</td>
<td>68</td>
</tr>
<tr>
<td>4.7</td>
<td>Normal Q-Q plot for M45</td>
<td>69</td>
</tr>
<tr>
<td>4.8</td>
<td>Histogram of frequency versus residual for M45</td>
<td>69</td>
</tr>
<tr>
<td>4.9</td>
<td>Residual versus fitted values for M35</td>
<td>71</td>
</tr>
<tr>
<td>4.10</td>
<td>Absolute studentized residuals versus fitted values for M35</td>
<td>71</td>
</tr>
<tr>
<td>4.11</td>
<td>Normal Q-Q plot for M35</td>
<td>72</td>
</tr>
<tr>
<td>4.12</td>
<td>Histogram of frequency versus residual for M35</td>
<td>73</td>
</tr>
<tr>
<td>5.1</td>
<td>Residuals versus fitted values for the Gompertz model</td>
<td>86</td>
</tr>
<tr>
<td>5.2</td>
<td>Residuals versus fitted values for the modified Gompertz model</td>
<td>86</td>
</tr>
<tr>
<td>5.3</td>
<td>Absolute residuals versus fitted values for the Gompertz model</td>
<td>87</td>
</tr>
</tbody>
</table>
Figure 5.4: Absolute residuals versus fitted values for the modified Gompertz model

Figure 5.5: Q-Q plot for the Gompertz model

Figure 5.6: Q-Q plot for the modified Gompertz model

Figure 5.7: Lag plot for the Gompertz model

Figure 5.8: Lag plots for the modified Gompertz model
**LIST OF APPENDIXES**

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPENDIX A</td>
<td>Background Questionnaire</td>
<td>111</td>
</tr>
<tr>
<td>APPENDIX B</td>
<td>List of all 63 possible models</td>
<td>112</td>
</tr>
<tr>
<td>APPENDIX C</td>
<td>Model adequacy result for all 63 possible models</td>
<td>116</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>AIC&lt;sub&gt;c&lt;/sub&gt;</td>
<td>The Corrected Akaike’s Information Criterion</td>
<td></td>
</tr>
<tr>
<td>ANNs</td>
<td>Artificial Neural Networks models</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian’s Information Criterion</td>
<td></td>
</tr>
<tr>
<td>FL</td>
<td>Foreign languages</td>
<td></td>
</tr>
<tr>
<td>L2</td>
<td>Second language</td>
<td></td>
</tr>
<tr>
<td>LLS</td>
<td>Language Learning Strategy</td>
<td></td>
</tr>
<tr>
<td>MAE</td>
<td>Mean Absolute Error</td>
<td></td>
</tr>
<tr>
<td>MLR</td>
<td>Multiple Linear Regression</td>
<td></td>
</tr>
<tr>
<td>MUET</td>
<td>Malaysian University English Test</td>
<td></td>
</tr>
<tr>
<td>OLS</td>
<td>Ordinary Least Square</td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
<td></td>
</tr>
<tr>
<td>RSE</td>
<td>Residual Standard Error</td>
<td></td>
</tr>
<tr>
<td>SILL</td>
<td>Strategy Inventory for Language Learning strategies</td>
<td></td>
</tr>
<tr>
<td>VIF</td>
<td>Variance Inflation Factor</td>
<td></td>
</tr>
</tbody>
</table>
LIST OF SYMBOLS

\( \beta_j \)  Unknown parameters

\( \hat{\beta}_j \)  Estimated parameters

\( \sigma^2 \)  Variance

\( \mu_t \)  Random error

\( Y_t \)  Independent variable

\( X_t \)  Dependent variable

\( e \)  Exponential

\( n \)  Sample size

\( k \)  The number of independent variables

\( L \)  Maximum likelihood

\( Y_\eta \)  The normalized data of \( Y \)

\( K \)  Number of parameters in the model, including the intercept and \( L \)

\( X_{i1} \)  Memory strategy

\( X_{i2} \)  Cognitive strategy

\( X_{i3} \)  Comprehensive strategy

\( X_{i4} \)  Metacognitive strategy

\( X_{i5} \)  Affective strategy

\( X_{i6} \)  Social strategy

\( \hat{X} \)  \([X_{i1} \; X_{i2} \; ... \; X_{i6}]^T\)
$N$ Number of possible models

$k$ Number of independent variables

$H_o$ Null hypothesis

$H_1$ Alternative hypothesis

$\bar{Y}$ Mean of dependent variable

$\hat{Y}_\eta$ Estimated dependent variable
CHAPTER 1

INTRODUCTION

1.1 Malaysian Education System and Language Learning Strategy

The Malaysian education system has always been ever changing to bring improvement across the country, regardless of those living in rural areas or in the urban areas. According to Zakaria (2000), the Malaysian national education system was established when Malaysia (then Malaya) gained its independence from the British and was modeled based on the English grammar school system of the British colonial era. However, throughout the years, Education Acts have been introduced from time to time to create an education system based on our national philosophies which is a centralized system with a common curriculum using Bahasa Malaysia as its main medium of instructions (Zakaria, 2000). As a step towards taking care of aspirations and sensitivities of Malaysia’s multiethnic population, Chinese and Tamil schools are allowed to be established up to primary level education (Zakaria, 2000). However, the implementation of language learning strategies in English language into the education system has been vague as there are hardly any documented records of this (Wong, 2005).

English is considered the country’s second language (L2) where it is a compulsory subject taught throughout primary and secondary schools in its formal education system. At a pre-university level, it is compulsory for those who plan to continue their studies at any Public Higher Educational Institutions (IPTA), as it is a pre-requisite for these students to take the Malaysian University English Test (MUET) prior to enrolment for their first degrees (“Bajet 2015.” 2015). However, despite having English language taught in primary and secondary schools for at least eleven years, an average Malaysian student still finds it hard to master the language adequately especially in terms of verbal fluency, writing compositions in English and also applying proper grammar. This is causing problems, academically,
for students to get further education, moreover when most public and private universities in Malaysia now set English as their main language for most courses as an advantage for students to gain information and all sorts of different knowledge easily because they are mostly written in English. This greatly influences the student’s future and the Malaysian’s Vision 2020 where the nation achieves the status of a fully developed and industrialised country.

Over the years in the language-learning world, a system called the Language Learning Strategy (LLS) was introduced and highly researched on up until today. This system is focused on strategies that are used when learning a second language (L2) and foreign languages (FL). A number of types of strategies were introduced, as it will be further discussed in Section 2.2. However, as Wong (2005) reported in her study that in literature, there are hardly any documented records of local studies on English as a second language and the language learning strategies used in the Malaysian education system. A recent literature by Teh et al. (2009) was done to examine the relationship between the use of language learning strategies and language learning motivation, but it was for the Arabic language among students from Malaysia religious secondary schools students. This study chose to introduce another strategy called the “metaphysic strategy”, as an addition to Oxford’s (1990) Strategy Inventory for Language Learning strategies (SILL), which are memory strategy, cognitive strategy, compensation strategy, metacognitive strategy, social strategy and affective strategy. The study found that motivation level affects language learners is such a way that the higher it is, the richer the repertoire of strategies and the more frequently they employ strategies than less motivated language learners (Teh et al. 1990).

The Malaysian education system recently introduced English as a medium of instruction in schools starting in the year 2003. However, as of 2012, the medium of instruction has since been reverted back to the Malay language. This has brought many debates amongst linguists and activists, not forgetting politicians. In fact, the Sultan of Johor, Sultan Ibrahim Ibni Almarhum Sultan Iskandar expressed his concerns as to why English should have remained the main medium of instruction in schools through the news reported by The Straits Times, Singapore (“The sad state of English in Malaysia,” 2015). Hence, this leaves limited options as
to who should be held to take the responsibility to improve the English proficiency of our students, as parents who started their education after the year 1970 (Putheh and Sintok, 2012), has faced little exposure to the language when the government decided to make the Malay language as the main medium of instruction. As a step forward, researches as to how to improve the practice towards the language should be taken, so despite the limited exposure towards the language, the impact of the knowledge gives obvious results in these pupils’ language learning process. As cited in Park (1997), in order for L2 learners to become autonomous learners who take charge of their learning, they must use appropriate language learning strategies (Wenden, 1991). Thus, putting as much effort to aid those in the pedagogical practice to improve the English language proficiency amongst pupils is highly recommended. Introducing the existence of language learning strategies into the system should be the step forward that needs to be taken, not only among students, but also towards present and future educators of any languages.

1.2 English Language Proficiency in Malaysia

According to The Malay Mail (“Malaysians most English proficient in Asia, ahead of Singaporeans, survey claims,” 2015), a survey was done and Malaysia tops the survey as the most English proficient country in Asia. However, worldwide, Malaysia is only ranked at number twelve (“EF English Proficiency Index,” 2015). These claims were made out of a survey that also said the survey index was self-selected and not guaranteed to represent the whole country, which was administered online. Thus, limiting their sample size to only those who are frequent Internet users and excluding those in rural areas, those that are not interested in taking tests via the Internet, and those with limited source for connecting to the Internet.

However, since these claims were made out of bias data, it is unfair to conclude entirely the state of English proficiency in Malaysia. In fact, Pandian (2002) stressed the concern among teachers, scholars and the members of public about how the proficiency of English amongst students has declined since the implementation of the Malay medium of instruction. Not only will this effect a student’s admission to institutions of higher learning, but also occupational needs in the industry (Abdullah, 2012). In addition to this, the government has made a
number of initiatives to help solve the problems. Amongst which was the 10-year period from 2003 to 2012 of when English was made the main medium of instruction in primary and secondary school for the Science and Mathematics subjects. However, this initiative was reverted back as many linguists and activist disagrees with the system, afraid that these students will lose their interest towards the National language of Malaysia. Another initiative of the Malaysian government was announced by the Former Deputy Prime Minister of Malaysia, Tan Sri Muhyiddin Yassin on the 8th of March, 2013, where it has been decided that English language will be a compulsory pass subject in the Malaysian Certificate of Education (SPM) starting 2016 ("Muhyiddin: English a compulsory pass subject as early as 2016," 2015). This is in hopes that it will make students realize the importance of English language before entering into their first degree due to the fact that most medium of knowledge these days are written in English.

As one of the stepping stone on how to make studying L2 easier, this study is important to ensure that those in the pedagogical practice will be more certain about their approaches on how to instill more knowledge about the English language among their students. Recognizing a student’s main strengths and weaknesses when it comes to language learning should be the initial move. This study aims to recognize how, as a whole, each of the language learning strategies aid these students of Universiti Malaysia Sabah in achieving better proficiency. After all, each of these strategies introduced by Oxford (1990) are according to the norm of language learning world. Thus, analyzing their ways in using these strategies will give an advantage to the educators as to how to give these students a more influential learning system.

1.3 Language Learning Strategies
This Section discusses the strategies that are used throughout this study. The context of this study portray, language learning strategy (LLS) is a type of learning process which involves actions, practice and interactions that are consciously utilized with the intentions of improving a learner’s ability to grasp a certain language. Using Oxford’s (1990) concept introduced in the book “Language Learning Strategy: What Every Teacher Should Know”, this study focuses on the
relationship between these strategies and the outcome of their language proficiency.

Oxford (1990) introduced six strategies, each with its specific definition. These strategies were divided into two; direct and indirect strategies, for which the book defined and showed their applications to develop four skills which are reading, writing, listening and speaking. Direct strategies are LLS that directly involve the target language such as memory, cognitive and compensation strategies, whereas indirect strategies are strategies that support and manage the language learning without directly involving the target language, such as metacognitive, social, and affective strategies (Oxford 1990).

Oxford (1990) defined these strategies short and precise using diagrams. It was said that;

i. Memory strategy is creating mental linkages, applying images and sounds, reviewing well and employing action.

ii. Cognitive strategy is practicing the language, receiving and sending messages, analyzing and reasoning, and creating structure for input and output.

iii. Compensation strategy is guessing intelligently and overcoming limitations in speaking and writing.

iv. Metacognitive strategy was described as centering your learning, arranging and planning your learning, and evaluating your learning.

v. Affective strategy is lowering your anxiety, self-encouraging, and taking your emotional temperature.

vi. Social strategy is asking questions, cooperating with others, and empathizing with others.

According to previous studies of LLS at a university level (Park, 1997; Wharton, 2000; Nisbet et al., 2005; Hong Nam and Leavel, 2006; McMullen, 2009; Ghavamnia et al., 2011), these strategies have been examined to be closely related to the language proficiency of the targeted students. In addition to that, this study focusses on on how a pre-university student in Universiti Malaysia Sabah utilizes all
six strategies and how it affects their language proficiency where the results can be used to analyze which strategy actually strengthens a student's language proficiency. As a result, both the English language education world and the research on LLS can be understood. Studies of the different movements and strategy used from students around the world is seen through researches from Korea (Park, 1997), China (Nisbet et al., 2005), Saudi Arabia (McMullen, 2009) and Botswana (Magogwe and Oliver, 2007). By referring to the previous statement, this could be a stepping-stone for researchers in Malaysia to go more in depth as to what makes a better language learner. Inspired by previous results, the same concept of study will be conducted to examine the LLS of the pre-university students, specifically from Universiti Malaysia Sabah.

1.4 Development of Mathematical Models
Mathematical models are a perception of reality. A single line of mathematical symbols and numbers may represent a number of things. When a mathematical model is done right, it will help to understand the system in a better way, and predict outcomes subject to its accuracy. To build a mathematical model, there has always been a system to follow. Beginning from the raw data up until its results, there are always tests and observations that need to be done.

Figure 1.1 is an adaptation from Archontoulis and Miguez (2015), where it was simplified to keep it within the scope of this discussion. It shows a diagram of steps to follow when building a mathematical model. It begins with raw data. To check for outliers, errors, and assumptions of a certain type of distribution, the discussion begins with graphs and plots such as quantile-quantile plots, histograms, scatter plots and residual plots. From the observations of the plot, the decision as to remove the outliers or to consider generalized linear models, which are models with errors that do not follow normal distributions, is to be considered. Then, determining the type of model the data might fit into is next. To build a nonlinear mathematical model, it will always have starting values because models are not guaranteed to converge to the global optimum. Thus, starting values are needed to begin estimation of the parameters. The starting values need to be obtained or dealt with by stating assumptions. However, in this study, the nonlinear model will


Chai, T., and Draxler, R. R. 2014. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. 7(3):1247-1250.


Muñoz-Cuevas, M., Metris, A., and Baranyi, J. 2012. Predictive modelling of Salmonella:
From cell cycle measurements to e-models. *Food Research International*. 45(2):852-862.


Climate research. 30(1):79-82.


