Modified Gompertz Model to Predict Language Proficiency in Proportion to Language Learning Strategy

¹J.J. Kiram, ²J. Sulaiman, ³S. Suwanto and ¹W.A. Din
¹Preparatory Center for Science and Technology,
²Faculty of Science and Natural Resources,
³Faculty of Psychology and Education, University Malaysia Sabah,
88400 Kota Kinabalu, Sabah, Malaysia

Abstract: A perpetuation from previous studies, it aims to fit a nonlinear mathematical model, specifically the modified Gompertz Model into a data involving language learning strategies. These language learning strategies are regarded as the six independent variables. Whereas, the language proficiency is regarded as the dependent variable. The language proficiency is measured according to the student's Malaysian University English test result where two hundred and thirty pre-university students of Universiti Malaysia Sabah participated. These language learning strategies are based upon a self-report questionnaire called the strategy inventory for language learning. The model's goodness-of-fit was tested using Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Residual Standard Error (RSE).

Key words: Mathematical modelling, language learning strategies, Gompertz Model, variable, strategies, data

INTRODUCTION

Constructing a mathematical model is possible for almost any situations be it in health science (El-Marouf and Al-Ahmady, 2012), economy or even the weather (Fithriasari *et al.*, 2013), each with its own standard of procedures and different mathematical methods. Mathematical modelling also goes through a process of trial and error until the best model is obtained. In this study of language proficiency and Language Learning Strategy (LLS), it has often been assumed to be a linear model (Park, 1997; Bremner, 1999; Magogwe and Oliver, 2007). However, with six language learning strategies introduced as independent variables (i.e., memory, cognitive, compensation, metacognitive, affective and social) the linearity of the model becomes questionable as to what is the best model that FITD these variables with minimal amount of error. Selecting the model is subjected to the main objective of modelling the relationship in the study. Previously, these data has been tested as a linear model of Eq. 1:

$$P(\tilde{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \mu \qquad (1)$$

Where:
k   = The number of independent variables β_i
j = 0, 1, 2, ..., k = Unknown parameters
μ   = The error term

$\tilde{x} = [x_1, x_2, x_3, x_4, x_5, x_6]^T$ (Kiram *et al.*, 2014, 2015a, b). Other than that the study also introduced fitting these social data into a model introduced by Gompertz (1825). The Gompertz Model with a few assumptions made which has resulted in an even lower error count compared to the linear models (Kiram *et al.*, 2014).

In spite of the importance of searching for the best model to describe the relationship between the LLS and language proficiency, this study attempts to fit a modified Gompertz Model introduced by Bhaduri *et al.* (1991). This study uses data from the pre-university students of Universiti Malaysia Sabah. Thus, this study fits these data into a nonlinear modified Gompertz Model and tests it for its model adequacy in order to explain the relationship between LLS and second language proficiency.

The English language in Malaysia is widely used language and was made compulsory to be taken as a subject throughout primary and secondary school for decades as instructed by the Ministry of Education, Malaysia (Thirusanku and Yunus, 2012). It is also compulsory to sit for the Malaysian University English Test (MUET) at a pre-university level prior to the enrolment for the student's first degrees (Allen and Lee, 2017). The results are a pre-requisite for entering most undergraduate programs at any Public Higher Educational

---

Corresponding Author: J.J. Kiram, Preparatory Center for Science and Technology, University Malaysia Sabah,
88400 Kota Kinabalu, Sabah, Malaysia

Institutions (IPTA). Despite the Ministry of Education initiatives towards promoting the use of English more often in the Malaysian education over the years, studies show that a typical student still finds it hard to adept the language adequately especially in terms of verbal fluency, writing compositions in English and also applying proper grammar. As the country's second language, it greatly influences a student's future. The inability of a student to adept the language is causing unsatisfactory examination result which brings problems for academically able students to get into universities. Moreover, most universities in Malaysia set English as the main language for most courses as an upper hand for students to gain information and all sorts of different knowledge easily.

In 1999, MUET was introduced and apparently one of the most frequently used examination to test English proficiency in Malaysia. There are four components tested in MUET; Listening (45 marks), Speaking (45 marks), reading (120 marks) and writing (90 marks) where each of these component represents a certain strength, such as the ability to communicate well to listen and understand comprehensions and to write argumentatively. The proficiency is graded using 6 bands with band 6 being the highest and band 1 the lowest. The exam is conducted three times a year which is in March, July or November ("STPM and MUET Calendar 2012," 2013).

A student's language proficiency is often influenced by their LLS. 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. Numerous studies around the world have been done to study how a learner's LLS affects their language proficiency, most of it representing a specific country (Park, 1997; Hong-Nam and Leavel, 2006; Magogwe and Oliver, 2007; Ghavamnia *et al.*, 2011). Many models proposed by these studies differ in terms of which strategy is most favoured by learners. However, none of which was found to discuss the relationship in nonlinear terms (Park, 1997). Thus, the need to study the possibilities of the relationship as a nonlinear system on a molecular level is what this study aspires.

The current study is a perpetuation where two hundred and thirty pre-university students at Universiti Malaysia Sabah participated the study back in October 2013. These students received a background questionnaire where it includes basic background questions such as gender, age, nationality, state of origin, language used at home, Malaysian Certificate of Education (SPM) results for English language, previous secondary school type, household income and parent's education, together with a self-report questionnaire called

the Strategy Inventory for Language Learning (SILL) (Oxford, 1990) to spot their language learning strategies. Despite the many strategy questionnaires that have been constructed (Bialystok, 1981; Politzer, 1983; Politzer and McGroarty, 1985), the SILL has been reported to have a higher degree of reliability and validity (Oxford, 1990, 1996). The students who participated must also sit for MUET in November of 2013. Confidentiality of their background and their results were kept and the data was only recorded based on their matriculation number. In this context, it is worth mentioning that multiple tests have also been made using the pilot data of this study where the data was assumed linear and tested for goodness-of-fit (Kiram *et al.*, 2015a, b).

These constructed mathematical models are merely approximations of the actual projections. To obtain the absolute right mathematical model is impossible, nor is it possible to say that a mathematical model is completely wrong. However, to Fathom a model that closely estimates the outcomes is possible and indeed crucial. Especially, in this study where finding a good model helps in aiding teachers and academicians to understand on how they can improve their way of delivering knowledge to their pupils. In spite of the previous findings regarding a nonlinear relationship between the LLS and language proficiency (Kiram *et al.*, 2013a, b), this study is attempting to find an even better nonlinear model to compliment the data further. We begin by fitting the data into the modified Gompertz Model (Bhaduri *et al.*, 1991). We then calculate its goodness of fit using Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Residual Standard Error (RSE).

MATERIALS AND METHODS

A study conducted within Universiti Malaysia Sabah, 230 pre-university science students participated in this study. These students were under Preparatory Centre for Science and Technology 1 year pre-university program. They had been informed verbally that they were part of a study to identify their language learning strategies and that there were no right or wrong answers to the questionnaires given. These students were given 20 min to answer both the SILL and background questionnaire simultaneously during a class in October 2013.

The MUET scores were taken to measure their language proficiency. Typically, MUET scores range from 0-300. Based on these data, the minimum score is 96 (band 1) while the maximum score is 227 (band 5). The calculated mean score is 153.413. To minimize the variance of the data in spite of its wide range, we normalized the MUET score as (Green and Oxford, 1995) Eq. 2:

$$y_{\eta_i} = \frac{y_i - (y_{min} - \in)}{(y_{max} + \in) - (y_{min} - \in)}; i = 1, 2, 3, \ldots, n \qquad (2)$$

Where:
$\in$ = Relative to small real number
y_η = The normalized data of y

Theoretical model: In this context, it is worth mentioning that these data have been tested for linearity with all its assumptions (Kiram *et al.*, 2015a, b) and a nonlinear Gompertz Model also comes with assumptions (Kiram *et al.*, 2015a, b). Thus, taking one more step forward into fitting the data into a modified Gompertz Model (Bhaduri *et al.*, 1991) where the general form is Eq. 3:

$$\log y(t) = a - ce^{-e^{-B(t)}} + \mu \qquad (3)$$

Where:
y (t)   = A function of t
a and c = Unknown parameters
B(t)   = A function of time

However, the data of this study is not a time-series data. Thus, slight adjustments were made to suite the data of one dependent variable and six independent variables. From Eq. 3, y is changed to a function of x, where Eq. 4:

$$\log y(x) = a - ce^{-e^{-B(\tilde{x}, \beta_j)}} + \mu \qquad (4)$$

where, $B(\tilde{x}, \beta_j)$ is a function of x and $B(\tilde{x}, \beta_j) = \beta_0 + \sum_i^n \beta_i x_i$ for every i = 1, 2, 3, ..., n. With the normalization in Eq 2, the study needs to let log y (x) = y_{η_i} instead in order for Eq. 4 to be valid. Since, linearizing Eq. 4 gives 5:

$$B(\tilde{x}, \beta_j) = P(\tilde{x}, \beta_j) \qquad (5)$$

where, Eq. 6:

$$B(\tilde{x}, \beta_j) = -\ln\left(\frac{-\ln(a - y_{\eta_i})}{\ln c}\right) \qquad (6)$$

and Eq. 7:

$$P(\tilde{x}, \beta_j) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} +, \ldots + \hat{\beta}_k x_{ik} + \mu_i \qquad (7)$$

Including all 6 independent variables, the full form of the model becomes Eq. 8:

$$y_{\eta_i} = 1 - e^{-e^{(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6})}} + \mu \qquad (8)$$

for every i = 1, 2, 3, ..., n.. Equation model (Eq. 8) assumes a and c from Eq. 3, equals 1 to reduce the number of unknown parameters that needs to be estimated. As a preliminary test, the limited source of statistical tools

withholds the ability to estimate further amount of unknown parameters. However, further researches are already on going to search for good estimators. Since, Eq. 8 is a nonlinear model, the value of all unknown parameters, β will be determined by linearization using R statistical software.

Moving on, we then calculate the corrected Akaike Information Criterion (AIC$_c$) (Hurvich and Tsai, 1989) of the model using the formula Eq. 9:

$$AIC_C = -2\ln(L) + 2K + \frac{2K(K+1)}{n - K - 1} \qquad (9)$$

Where:
L = The maximum likelihood that estimates for the model
K = Must be the total number of parameters in the model including the intercept and L

There are many methods to test the goodness-of-fit of the model, be it graphical or numerical. However, due to the fact that this model considers 6 independent variables and fitted into a nonlinear data to assess it graphically would be difficult as most statistical software do not offer this method, especially with a modified Gompertz Model as Eq. 8. Therefore, in this study, we show the goodness-of-fit using the methods of Root Mean Square Error (RMSE), Mean Square Error (MAE) and the Residual Standard Error.

Strategy Inventory for Language Learning (SILL): The SILL Version 7.0 (Oxford, 1990) was used to measure learning strategies preferences. It is a self-report questionnaire that is divided into 6 sections where each section represents a particular strategy. The direct strategies were memory, cognitive and compensation and the indirect strategies were metacognitive, affective and social. The questionnaire had 50 items and students responded to each item using a 5-point Likert scale with 1 being "Never or almost never true of me", 2 "Usually not true of me", 3 "Somewhat true of me", 4 "Usually true of me" and 5 "Always or almost always true of me". The questionnaire was prepared in both English and Malay.

Background questionnaire: The background study of each participant is crucial to set the study's sample criteria. This questionnaire requires the research subjects to provide the following information: matriculation number, gender, age, nationality, state, language used at home, SPM result for English language, previous secondary school type, household income, parent's highest education and the student's use of English whether as a first language, second language or foreign language.

RESULTS AND DISCUSSION

The proposed assumption of Eq. 8 was fit into the data with $P(x) = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{\beta}_4 x_{4i} + \hat{\beta}_5 x_{5i} + \hat{\beta}_6 x_{6i}$. The independent variables $x_1, x_2, ..., x_6$ each representing the language learning strategy which are memory, cognitive, comprehensive, metacognitive, social and affective, respectively. We first begin by fitting the model and calculating the convergence. Next, we calculate the AIC_c of the model. Finally, we test the goodness of fit of the model using the earlier mentioned methods.

Fitting the model: Fitting the model requires starting values. All seven unknown parameters are denoted as β_0, $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ and β_6. By ad hoc guessing, considering the y_η to be a normalized value ranging (0, 1), thus, we assume all of our unknown parameter's starting values as zero.

Next is to check the algorithm convergence. Using R statistical software, the function gives the number of iterations to convergence as 5 and the achieved convergence tolerance is 4.813e-07. Table 1 shows the estimated value for the unknown parameters. According to the estimates, cognitive strategy and metacognitive strategy are the only ones showing non-negative values, all ranging from (0, 1). The standard errors calculated were of small values. However, this may be due to the fact that y_η and the predicted $\hat{y}_\eta$ values must all be within (0, 1), thus making the sum of error to be small.

The corrected akaike information criterion: The AIC_c of the model was calculated using Eq. 9. The AIC_c calculated is -158.1829. The AIC_c is designed to compare the performance of models. In this study, model in Eq. 8 was built to compare with the linear model in Eq. 1 as follows Eq. 9:

$$\hat{y}_\eta = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{\beta}_4 x_{4i} + \hat{\beta}_5 x_{5i} + \hat{\beta}_6 x_{6i} \qquad (9)$$

Where:

$\hat{y}_\eta$ = The predicted normalized dependent variable
ε = The term for error

The AIC_c calculated for this model is -156.6422. Thus, making the modified Gompertz Model to be the better model with a more negative AIC_c.

Goodness of fit: The goodness-of-fit of this model was evaluated using three methods that are the RMSE, MAE and the residual standard error. Table 2 shows the calculated values obtain for the nonlinear Modified Gompertz Model using these methods. All three methods have a similar rule which is the closer it is to zero, the better the model. Table 2 shows the RMSE calculated showed a small value-approaching zero, same as the MAE and residual standard error. The comparison shows that

Table 1: Estimates and standard error of the unknown parameters

Parameters	Estimate	SE
$\hat{\beta}_0$	-0.77417	0.25201
$\hat{\beta}_1$	-0.09789	0.06813
$\hat{\beta}_2$	0.30078	0.09083
$\hat{\beta}_3$	-0.07870	0.06254
$\hat{\beta}_4$	0.24926	0.07797
$\hat{\beta}_5$	-0.28758	0.06211
$\hat{\beta}_6$	-0.06284	0.06738

Table 2: Goodness-of-fit for nonlinear model

Variables	RMSE	MAE	Residual (SE)
Linear model	0.1660	0.1297	0.1686
Nonlinear model	0.1659	0.1295	0.1685

the nonlinear model modified Gompertz is obviously better than the linear model in terms of the amount of error.

CONCLUSION

The nonlinear modified Gompertz Model was formed and its AIC_c and goodness of fit was tested. The estimated parameters were concluded after 5 iterations to convergence and showed acceptable values of standard errors. The AIC_c was calculated compared to the linear model's AIC_c and it showed improvements as the AIC_c for the nonlinear model was even more negative. The goodness of fit test was done and the RMSE, MAE and residual standard error all showed smaller values approaching zero as compared to the linear model.

Despite the data fitting the modified Gompertz Model, the model shows uncertainties as to whether there could be a better nonlinear model that will fit these data better or other ways to approach the assumptions to make the model even more accurate.

RECOMMENDATION

Hence, further research must be done to model out these LLS in order to help educators produce students with better English proficiency.

REFERENCES

Allen, S.H.L. and E.W. Lee, 2017. Involvement of higher order thinking skills within a preparatory course for the Malaysian university English test. Eng. Teach., 40: 91-112.

Bhaduri, S., P.W. Smith, S.A. Palumbo, C.O. Turner-Jones and J.L. Smith *et al.*, 1991. Thermal destruction of *Listeria monocytogenes* in liver sausage slurry. Food Microbiol., 8: 75-78.

Bialystok, E., 1981. The role of conscious strategies in second language proficiency. Mod. Lang. J., 65: 24-35.

Bremner, S., 1999. Language learning strategies and language proficiency: Investigating the relationship in Hong Kong. Canadian Modern Lang. Rev., 55: 490-514.

El-Marouf, S.A.A. and A.A.O. Al-Ahmady, 2012. Global analysis of tumors and drugs model. Intl. J. Appl. Math. Stat., 27: 80-101.

Fithriasari, K., N. Iriawan and B.S.S. Ulama, 2013. On the multivariate time series rainfall modeling using time delay neural network. Intl. J. Appl. Math. Stat., 44: 193-201.

Ghavamnia, M., Z. Kassaian and A. Dabaghi, 2011. The relationship between language learning strategies, language learning beliefs, motivation and proficiency: A study of EFL learners in Iran. J. Lang. Teach. Res., 2: 1156-1161.

Gompertz, B., 1825. On the nature of the function expressive of the law of human mortality and on a new mode of determining the value of life contingencies. Philos. Trans. R. Soc. London, 115: 513-585.

Green, J.M. and R. Oxford, 1995. A closer look at learning strategies, L2 proficiency and gender. TESOL Quart., 29: 261-297.

Hong-Nam, K. and A.G. Leavell, 2006. Language learning strategy use of ESL students in an intensive English learning context. Syst., 34: 399-415.

Hurvich, C.M. and C.L. Tsai, 1989. Regression and time series model selection in small samples. Biometrika, 76: 297-307.

Kiram, J.J., J. Sulaiman, S. Swanto and W.A. Din, 2013a. The relationship between English language learning strategies and gender among pre-university students: An overview of UMS. Proceedings of the 3rd International Conference on Mathematical Sciences, June 17-19, 2013, AIP Publishing LLC., Kuala Lumpur, Malaysia, ISBN:978-0-7354-1236-1, pp: 502-507.

Kiram, J.J., J. Sulaiman, S. Swanto and W.A. Din, 2013b. The relationship between English language learning strategies and proficiency of pre-university students: A study case of UMS. Proceedings of the 21st International National Symposium on Mathematical Sciences (SKSM21), November 6-8, 2013, AIP Publishing LLC., Penang, Malaysia, ISBN:978-0-7354-1241-5, pp: 775-780.

Kiram, J.J., J. Sulaiman, S. Swanto and W.A. Din, 2014. A Gompertz model of the strategy inventory for language learning questionnaire. Intl. J. Appl. Math. Stat., 53: 212-219.

Kiram, J.J., J. Sulaiman, S. Swanto and W.A. Din, 2015a. Modeling the language learning strategies and English language proficiency of pre-university students in UMS: A case study. Proceedings of the 22nd National Symposium on Mathematical Sciences (SKSM22): Strengthening Research and Collaboration of Mathematical Sciences in Malaysia Vol. 1682, November 24-26, 2014, AIP Publishing LLC., Selangor, Malaysia, pp: 1-1.

Kiram, J.J., J. Sulaiman, S. Swanto and W.A. Din, 2015b. Performance of variable selection methods in predicting language proficiency using language learning proficiency. Appl. Math. Sci., 9: 557-568.

Magogwe, J.M. and R. Oliver, 2007. The relationship between language learning strategies, proficiency, age and self-efficacy beliefs: A study of language learners in Botswana. System, 35: 338-352.

Oxford, R.L., 1990. Language Learning Strategies: What Every Teacher should Know. Newbury House Publishers, New York, USA., ISBN-13:9780066326078, Pages: 342.

Oxford, R.L., 1996. Employing a questionnaire to assess the use of language learning strategies. Appl. Lang. Learn., 7: 25-45.

Park, G.P., 1997. Language learning strategies and English proficiency in Korean university students. Foreign Lang. Ann., 30: 211-221.

Politzer, R.L. and M. McGroarty, 1985. An exploratory study of learning behaviors and their relationship to gains in linguistic and communicative competence. Tesol Quart., 19: 103-123.

Politzer, R.L., 1983. An exploratory study of self reported language learning behaviors and their relation to achievement. Stud. Second Lang. Acquisition, 6: 54-68.

Thirusanku, J. and M.M. Yunus, 2012. The many faces of Malaysian English. ISRN. Educ., 1012: 1-14.