Application of Principal Component Analysis on English Proficiency among Technical University Students

Siti Zahariah Abdul Wahab 1*, Yasmin Yahya1, Norhatta Mohd 1, Nurul Sharaz Azmanuddin 1, Nurhadiana Nurulmatin 2

1 Applied Statistics and Data Science Cluster (ASADS), Universiti Kuala Lumpur Malaysian Institute of Information Technology (UniKL MIIT), 1016 Jalan Sultan Ismail, Kuala Lumpur, Malaysia
2 Universiti Kuala Lumpur Malaysian Institute of Information Technology (UniKL MIIT), 1016 Jalan Sultan Ismail, Kuala Lumpur, Malaysia

Received: 29/06/2020  Accepted: 05/10/2020  Published: 10/10/2020

Abstract

The purpose of this study was to explore the level of English proficiency among students from Universiti Kuala Lumpur, Malaysian Institute of Information Technology (UniKL MIIT). This study also observed the relationship between the level of four main skills: Reading, Writing, Listening and Speaking by distributing an online questionnaire to 116 students from bachelor programme. The data was then analysed using the Principal Component Analysis (PCA) to reduce the dimensionality of the dataset. The research findings revealed that the students' level of Reading, Writing, Listening and Speaking proficiency are intermediate. The findings also showed that there is a significant relationship between each Reading, Writing, Listening and Speaking proficiency of the students. On the other hand, the finding also reflects that students acquired the lowest in writing skill. Finally, the research draws conclusion and recommendations based on the level of students' English Proficiency and the ways to improve students' achievement in English.

Keywords: Data analysis, Principal Component, Eigenvector, English proficiency

1 Introduction

English is recognized as the world's intermediate language. Million people all over the world speak English as their first language. Undeniably, English has become a global language and it is very significant in the world trade, international relations, technological advances, science, medicine and education. English is not only a language, but it is a skill that one must have to grow and excel in this challenging world. Today, Malaysia uses English as a second language and most of the universities in Malaysia use English as the dominant language in the classroom. Moreover, all the lecture notes, assignment questions and exam questions are also in English.

Looking at the vast importance of English to the worldwide, every student is required to learn this language so that at least they can read and write in English. However, acquiring and mastering only reading and writing skills are not enough in order to become a competitive student. As a great student, he must be able to read, write, listen and speak the language well. These four skills, reading, listening, writing and speaking should be integrated in order to have good English proficiency and develop students' communication skill (Nan, 2018). In fact, listening and speaking are crucial skills that are correlated in real circumstances. To achieve an effective communication, one must listen and understand the conversation in order to respond in an understandable language (Sadiku, 2015). Apart from that, a student needs to master reading and writing skills since these two skills are very associated. This is to ensure an effective written communication can be achieved.

Furthermore, English proficiency is the main requirement for students to go for a job interview. Most of the companies in Malaysia looking for candidates who can converse and voice out ideas in English fluently. The goal of this study was to investigate the level of English proficiency among University Kuala Lumpur (UniKL MIIT) students based on the four main English language proficiency skills namely reading, writing, listening and speaking. Principal Component Analysis (PCA) was employed to evaluate the data throughout this study. The following research questions were addressed in this study:

a) Is there any relationship between the four main skills: Reading, Writing, Listening and Speaking?
b) What is the students’ level of English proficiency for the four main skills?

c) What is the performance of PCA on the students’ English proficiency based on the four main skills?

The objectives of the study are as below:

a) To determine the significant correlation between the four main skills: Reading, Writing, Listening and Speaking.

b) To identify the students’ level of English proficiency for the four main skills.

c) To determine the performance of PCA on the students’ English proficiency based on the four main skills.

2 Literature Review

2.1 The Importance of Mastering Reading, Writing, Speaking and Listening

Reading, writing, speaking and listening are the four foundational skills of language learning that complement each other. In order to be a well-rounded person who can be fluent in English language learning, a very strong foundation is needed to ensure someone can fully utilize the skills acquired. The importance of mastering the four skills are necessarily crucial especially to students.

Reading is a method for communicating via the written form and it is the most powerful way of gaining knowledge on many aspects of life including Science and Technology. Reading can improve students’ vocabulary. In reading, students will be exposed to a variety of new words and it can be used in writing and speaking session. This skill is important to students in high school or university since most of the academic textbooks are in English. If students cannot understand English, it will be hard for them to understand the subject and they will be left out in class.

Besides reading, good English writing is also an important factor for academic growth. It is necessary for career development (Mathew Nalliveettil & Mahasneh, 2017). This is because writing helps to strengthen the grammatical structure and enhance the students’ vocabulary. Students need this skill to do assignments, reports, take notes in class and do exams (Indrilla & Ciptaningrum, 2018).

Another important skill is listening as it is one of the main channels for language learning. Listening can be considered as the first skill in learning a language. Before we learn how to speak, we learn how to hear. Language learners with good listening skills can successfully engage in a conversation and learn to speak, read and write by listening to others (Hong, 2016). A successful conversation must have a good listener and a good speaker (Leong & Ahmadi, 2017). Listening ability is essential for students to understand the lectures in class. Eighty percent of gaining knowledge in university is delivered by lectures. A student learns better when he can listen better (Sadiku, 2015). Thus, students with good listening skills are easy to associate with Western students because Westerners have different English accents.

Finally, speaking is also an important skill to deliver a message to people (Sadiku, 2015). Language learners normally find conversing in English as a daunting process despite the fact that with good speaking skills, someone will be able to socialise better and communicate effectively with other people. Having good proficiency in English speaking will help students to voice out their opinion and ideas in the most appropriate way either during presentation or even discussion.

2.2 Factors that Hinder the Mastery of the Students’ Reading, Writing, Speaking and Listening Skills

There are many factors that can hinder the mastery of the students’ reading, writing, speaking and listening skills. These obstacles are sometimes realised clearly by the students and sometimes undetected. Reading habit for example needs to be nurtured from childhood. Nowadays, students do not like to read book, especially English book (Rachman, 2018). Many students are demotivated to read English reading materials because they do not understand the meaning. Lack of the vocabulary makes the students refused to read English book. There are some English words which are spelt the same but bring different meanings and connotations. This makes students confused and misinterpret the whole text. Complex and long sentences with many clauses will also hinder students’ understanding.

Apart from reading, writing is the way people express their thoughts and emotions with visible signs and it is understandable not only for themselves but also for other people (Indrilla & Ciptaningrum, 2018). Writing is one of the core English competences. In a study conducted by David et. al (2015), the data analysis found that writing was the hardest and most complex skill students had to master. This is supported by Javed et. al (2013) where they stated that even English speakers have problems in demonstrating good writing command. Nevertheless, research studies have shown that students are not inspired and willing to develop their language skill, particularly their writing skills (David et. al, 2015).

Several students are also very afraid to speak English because they are nervous to use the language. Malaysian students especially are reluctant to communicate in English because they felt embarrassed and they want to look good when speaking in English. This attitude restricts them to speak English because they are afraid to make even one mistake while speaking (Rusli et. al, 2018). Furthermore, some schools are still implementing the classical way of teaching where the teachers apply one-way communication and have less interaction with the students. This will make students feel comfortable in their own zone which hinder the mastery of the speaking skill.

2.3 Methods to Improve the Four English Skills

To develop and improve the four English skills, students and teachers need to play their role in ensuring the mastery of English can be done. Reading skill can be developed by students if they work hard to learn English and teachers must give full support to them (Nur & Ahmad, 2017). As stated by Teevno and Raisani (2017), reading is a basic skill to learn English because from reading students can enrich their vocabulary and will have better understanding when reading English materials (Sadiku, 2015). Interest in reading English materials should be inculcated from not only home but also school. In Muhammadiyah Yogyakarta, the teachers used the translation method to encourage the students to read English (Ismail et. al, 2017). This will help the students to understand the English materials better and eventually increase the students’ passion in reading who find English as a foreign language and hard to be learnt. Even though students have notes in hardcopy form, they also need an explanation from lecturers or teachers to enhance their understanding in the subject (Picard & Velautham, 2016).
Therefore, the role of facilitators in learning cannot be ignored. Apart from that, some universities in Malaysia have visiting lecturers from western countries and students should take this opportunity to hear any academic talks and sharing sessions which are conducted in English so that they become familiarised with the language.

Another way is to practice spontaneous speaking. It is the best way to train students to become more efficient and confident to talk in English because students are trained to talk in the most natural and spontaneous way (Lumettu & Runtuwene, 2018). Hence, teachers play an important role to encourage students to speak English in class and develop their confidence to communicate in front of the crowd (Leong & Ahmadi, 2017).

3 Method

3.1 Study Design

This research utilised the quantitative study by collecting information among UniKL MIIT students for two semesters by online survey. A total of 116 students were participated in the survey. The survey link was broadcasted to third semester degree students through random sampling technique. Random sampling was used as it is the suitable technique to collect data since it is a quick and easy method to give equal chance of anyone being picked within one large target group. The programs that participated in this study were Bachelor in Computer and Entrepreneur Management (BCEM), Bachelor in Networking System (BNS) and Bachelor in Software Engineering (BSE). The average time taken to answer the questionnaire was 30 minutes.

3.2 Study Instrument

3.2.1 Questionnaire

The instrument used was adapted from various studies which incorporated the values of attitude, understanding and performance of students towards the use of English. English proficiency consists of four elements: Speaking, Writing, Reading and Listening. The questionnaires done by (Atetwe, 2013; Latu, 1994; Marian et al., 2007; Yushau and Hafidz, 2015) were adapted to suit the scope of this study. There were two sections in the questionnaire: the first section is seeking for students’ demographic background where closed-ended questions were asked. The second section consists of questions regarding students’ attitude, performance and understanding towards the use of English as well as their score on the four main skills: Speaking, Writing, Reading and Listening. For the purpose of the participants to understand the items better, some amendment, enhancement and improvement have been made for several phrases and sentence construction of the questionnaire. Likert Scale items were used for the rating from 1 to 5 with 1 for totally disagree, 2 for disagree, 3 for neutral, 4 for agree and 5 for totally agree.

3.2.2 Data analysis: Principal Component Analysis

Principal Component Analysis is a powerful mathematical procedure in multivariate analysis. The main purpose of the PCA technique is to reduce the dimensionality of a correlated and high dimensional dataset (Shlens, 2005). In this procedure, it reduces and transforms a larger number of explanatory variables dataset to a smaller number of factors or latent variables of PCA that can be visualized graphically, with minimal loss of information (Li, Liu, Peng, & Sun, 2016). These orthogonal latent variables are known as principal components (PCs) in PCA and can be defined as a linear combination of variables.

Let a data matrix,

\[ X = n \times p \]  

(1)

Where \( n \) is a number of observations and \( p \) is a number of independent variables.

In PCA, the linear combination of variables for PCs is formulated as,

\[ l_i = x_{i1} p_1 + x_{i2} p_2 + ... + x_{ip} p_m \]

\[ i = 1, 2, ..., n \]  

(2)

\( p_m \) is a loading vector of unknown coefficients and \( l_i \) is a score vector of PCs. It can be written as

\[ T = X P \]  

(3)

and \( P \) is an orthogonal matrix. The number of \( P \) is chosen based on the first \( r \) of eigenvector columns, where \( r < m \). The PCs are uncorrelated and ordered according to the magnitude of the eigenvalues. The first principal component, PC1 holds the largest variance or contains as much information of dataset. The variance decreases as the number of PCs increase. To choose the ample number of PCs, they must have a cumulative variance explained at least 80% of the total variance.

The PCA technique can be summarized as follows:

a) First, we normalize the data set by using mean and standard deviation to convert the dataset into the same scale.

\[ u_i = \frac{x_i - \bar{x}}{D_x} \]  

(4)

Where \( \bar{x} \) is a mean of \( x \) and \( D_x \) is a standard deviation of \( x \).

b) Then, calculate the covariance matrix of standardized data set in step (a). This covariance matrix measures the correlation among the independent variables. Given the formula as below,

\[ \sum = \frac{1}{n - 1} [ (u - \bar{u})^T (u - \bar{u}) ] \]  

(5)

It is crucial to identify correlation variables because they comprehend noise and redundant information.

c) Calculate the eigenvalues based on the covariance matrix in (b). Then, determine the eigenvector corresponding to the eigenvalues.

\[ (U - \lambda I)E = 0 \]  

(6)
E is an eigenvector corresponding to the eigenvalue, \( \lambda \) of \( U \). The magnitude of eigenvalues is equivalence to the score variance and the eigenvectors signify the directions of PCs.

d) Then, the eigenvectors are sort according to their eigenvalues in decreasing order.

\[
\lambda_1 \quad 0 \quad \ldots \quad 0 \\
0 \quad \lambda_2 \quad \vdots \\
\vdots \\
0 \quad 0 \quad \ldots \quad \lambda_p \tag{7}
\]

e) The first \( r \) of eigenvectors will be chosen and that will form a new \( k \) dimensions of the matrix. The first PC will capture the most information of the dataset because it has the highest magnitude of the eigenvalue as it holds a lot of information about the data and any changes along that vector will give large variations. PCs with small eigenvalues have low variance and thus our data does not give significantly different when there are changes along the vector.

f) Transform the original \( n \)-dimensional data points into \( k \) dimensions onto the PCA subspace.

\[
T = X P_k 
\]

4 Result and Discussion

Descriptive statistics were performed to explore the characteristics of the sample. As can be seen from Table 1, the distribution of the sample by gender found that 58.6 percent of the students were males and 41.4 percent were females. Most of the respondents come from the city (83.6%) compared to the village which is 16.4 percent. The SPM (Malaysia Education Certificate) result for English subject demonstrates that approximately half of the respondents scored grade B, 34.5 percent scored grade A and 14.7 percent scored grade C and below. Table 2 shows the mean, standard deviation, minimum and maximum values for the four main skills of English proficiency. There is not much variation of the mean values among the four main skills with Speaking shows the highest score of 2.90 with standard deviation of 1.14. The lowest mean score is Writing with 2.60 and standard deviation of 1.00. The range score for all the main skills are from 1 to 5.

Table 1: Descriptive statistics of the respondents

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number (n)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>68</td>
<td>58.6</td>
</tr>
<tr>
<td>Female</td>
<td>48</td>
<td>41.4</td>
</tr>
<tr>
<td>City</td>
<td>97</td>
<td>83.6</td>
</tr>
<tr>
<td>Village</td>
<td>19</td>
<td>16.4</td>
</tr>
<tr>
<td>SPM English (A)</td>
<td>40</td>
<td>34.5</td>
</tr>
<tr>
<td>SPM English (B)</td>
<td>59</td>
<td>50.9</td>
</tr>
<tr>
<td>SPM English (C and below)</td>
<td>17</td>
<td>14.7</td>
</tr>
</tbody>
</table>

In our study, 4 elements were measured to determine the level of English proficiency among the UniKL students. There are Speaking, Writing, Reading, and Listening skills. The PCA method has been used to reduce the dimensionality of a dataset and making easier to visualize the issue in English proficiency among the students.

Table 2: Basic info of the four main skills

<table>
<thead>
<tr>
<th>Main Skill</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>2.78</td>
<td>1.38</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Writing</td>
<td>2.60</td>
<td>1.00</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Speaking</td>
<td>2.90</td>
<td>1.14</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Listening</td>
<td>2.75</td>
<td>1.22</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3 shows that there are correlations among all the elements, and it proves that those elements are the important criteria in measuring the students’ English proficiency. Students who have good Writing skill will be good in Speaking skill and vice versa. Writing skill is essential for students to write good research thesis, prepare lab reports and score in English subject. Speaking skill is very much needed for UniKL students to present their final year presentation (FYP), interacting in class with lecturers and international students. The ability of English-speaking skill is commonly being a benchmark for students to be more confident for their future career preparation. Listening skill is the second-highest correlation to Speaking skill. Students must be able to listen and understand English well for them to catch-up the lesson in class since all the lectures in UniKL MIIT will be held in English. If they cannot understand the language well, they will miss what they are learning in class. Reading skill also shows a positive correlation with all the elements. Reading skill is crucial to be mastered by UniKL MIIT students so that they can read and understand the academic textbooks while learning English terms and eventually enhancing their English vocabulary. Thus, reading is a basic skill that students must possess to become proficient in writing, listening and speaking skills.

Table 3: The correlation matrix among the selected criteria

<table>
<thead>
<tr>
<th></th>
<th>Speaking</th>
<th>Writing</th>
<th>Reading</th>
<th>Listening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speaking</td>
<td>1</td>
<td>0.667</td>
<td>0.634</td>
<td>0.646</td>
</tr>
<tr>
<td>Writing</td>
<td>0.667</td>
<td>1</td>
<td>0.590</td>
<td>0.595</td>
</tr>
<tr>
<td>Reading</td>
<td>0.634</td>
<td>0.590</td>
<td>1</td>
<td>0.697</td>
</tr>
<tr>
<td>Listening</td>
<td>0.646</td>
<td>0.595</td>
<td>0.697</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4: The Importance of Components

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalue</td>
<td>1.7074</td>
<td>0.6770</td>
<td>0.5698</td>
<td>0.5492</td>
</tr>
<tr>
<td>Variance</td>
<td>0.7288</td>
<td>0.1146</td>
<td>0.0812</td>
<td>0.0754</td>
</tr>
<tr>
<td>Percent</td>
<td>0.7288</td>
<td>0.8434</td>
<td>0.9246</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

PCA method was conducted to avoid multicollinearity in the dataset because there was a redundancy among the variables. The eigenvalues in Table 4 represent the calculation of the amount of variation held by each principal component. These eigenvalues are decreasing subsequently. The sum of all the eigenvalues gives a total variance of 3.5034. The second row for Table 4 represents
the proportion of variation explained by each PC. We can describe that for PC1, there is 73 percent holds the information of the original data and PC2 contains 11.46 percent of the variation. In the analysis, the first two PCs were only used with the cumulative variance percent of 84.34 and this percentage is large enough to explain the whole dataset.

Generally, a larger variance percent means a better variable. Eighty percent is considered as the number of the percentage which describes the data well (Wold et. al, 2002). It is shown in the Scree plot in Figure 1 that variance for PC3 and PC4 are relatively small. PC3 and PC4 can be just ignored because their percentage of variance is small and almost the same. Thus, they will give the same information as PC1 and PC2 do. In this manner, the number of variables was decreased from 4 to only 2 variables.

![Figure 1: Scree plot (Y-axis = variance; X-axis = number of PCs)](image)

As can be seen in Table 5, Principal Components are the linear combinations for each of the variables. For the first loading vector places approximately equal weight on Speaking, Reading and Listening skills and less weight on Writing skill. PC1 has much information for Speaking, Reading and Listening skills. The second loading vector places the highest weight on Writing skill and less weight on the other 3 features. Hence, PC2 has more information on Writing skill. The real situation for English language proficiency among UniKL students can be observed based on Figure 2 and Figure 3. Most of the individual data are far away from the skill projections and some are scattered around the origin. It can be explained that most of UniKL students do not have good skills in reading, listening, speaking and writing. Only a few of them can read and listen to English well, and only a few of them can speak English well. Many students have incorrect use of English grammar and inappropriate use of sentence structures during a conversation. This is because most of the students learn English for the exam purpose (Nishanthi, 2018) and they study English just to satisfy the university requirement as English is a compulsory subject (Jasima et al., 2017). Fewer students engage in English conversations between themselves, and this leads them to have a low level of speaking skills. In fact, 90% of the UniKL MIIT students are Malay. They prefer to use their native language rather than speaking English among themselves (Wahid & Pilus, 2017). Some students were faced with difficulties in understanding the questions from the lecture and this eventually make them become confused and misinterpret the lesson. Most of them have endured frustration, humiliation, shyness, and lack of confidence in the classroom. These factors hinder the mastery of the students’ English language skills.

![Figure 2: Individuals-PCA Biplot](image)

![Figure 3: PCA Biplot](image)

PC2 describes the skill of Writing among UniKL students. Very few students could write English well. This skill was evaluated based on how the students constructed good sentences, used correct usage of grammar, expressed ideas in writing and based on the previous English semester result. Most of the students did not focus their attention on the ability to write at the school level and this weakness is brought by them to the university level. As a result, it can be seen from the result of their English in SPM (Malaysia Education Certificate). Only few of
them scored A on their English paper for SPM Exam. This is relevant to the correlation score as presented in Table 3. Students with good Writing, Listening and Reading skills will be able to speak well.

5 Conclusions

The results of this study indicate that the students who have been introduced to English language in UniKL, MIIT demonstrate different level of understanding among the four main skills which are Reading, Writing, Listening and Speaking. The correlation test suggests that the four main skills are highly correlated between each other. The high correlation between the main skills may lead to multicollinearity; hence, Principal Component Analysis (PCA) is used to further investigate with the aim of focusing on the most significant skills. The results revealed that the students have a moderate level in English proficiency based on the four main skills with the lowest capability possesses among students is the Writing skill. Despite the contribution to the field of English Language proficiency, some limitations of this study need to be recognized in interpreting the results. The respondents come from similar field of study and the number of respondents is also quite low compared to the population of the whole UniKL MIIT; consequently, the conclusion drawn from this study is limited by the nature of the particular sample used, which comprises merely on technical students in UniKL, MIIT. Future research should consider these and other related issues concerning English proficiency as to deliver empirical evidence of its effectiveness over the four main skills.

Acknowledgement

The authors would like to thank the members of the Applied Statistics and Data Sciences Cluster (ASADS) for their comments and suggestions throughout the study. Special thanks to Universiti Kuala Lumpur for funding the conference fees.

Competing interests

The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

References
