Analyzing the Factors Associated with Perceived English Oral Communication Skills among Undergraduates: A Case Study

D.S. RUWANKUMARA, E.R.A.D. BANDARA

shashika14854@gmail.com, anjana@stat.cmb.ac.lk

University of Colombo, Faculty of Science
Department of Statistics

This case study explore the factors influence the English oral communication skills of undergraduates of the Faculty of Science at the University of Colombo. 332 undergraduate participants were surveyed using stratified sampling based on their academic year and subject stream. The assessment of their oral communication skills encompassed four critical aspects: presenting ideas verbally, listening skills, providing feedback, and presentation abilities. Employing exploratory data analysis (EDA) and advanced statistical models including multivariate ridge, lasso, elastic net regression, and partial least square regression, the study uncovers the most significant factors influencing English oral communication skills were the results of English language-related exams taken during secondary education and undergraduate years. Students who performed well in these exams demonstrated superior communication skills, alongside the positive impact of extracurricular participation as revealed through EDA. The results of the study guide oral English communication practice for students' future careers and educational.

Keywords: English oral communication, multivariate regularized regression, partial least square regression.

1. Introduction

This study underscores the critical importance of effective communication, particularly oral communication skills, among undergraduate students, especially in contexts where English is not the primary language. In countries like Sri Lanka, where English is often a second language, students encounter challenges in developing strong oral communication skills despite exposure to the language in their education.

Oral communication involves conveying information verbally, requiring clarity, active listening, and effective responses, alongside nonverbal elements such as tone, body language, and gestures. These skills are vital for academic, social, and professional success, as undergraduates must engage in presentations, and discussions, and seek guidance from mentors. Unfortunately, many students in state universities in Sri Lanka face confidence issues and subpar oral communication skills, affecting their academic performance and interactions with mentors [1], [2], [3]. Strong oral communication is also essential for globalized workplaces where organizations interact with stakeholders worldwide. Most undergraduates do not possess innate verbal skills and must actively work on developing them [4]. Therefore, understanding the components of effective oral communication and consistent practice is crucial for improvement. The primary aim of this study is to identify key factors influencing English oral communication skills among undergraduates at the University of Colombo's Faculty of Science. By identifying these factors, the study aims to contribute to the development of comprehensive graduates with strong oral communication skills, essential for success in academic, professional, and globalized contexts.

2. Methodology

The study was done using primary data that was collected using a survey. The objectives of the study were achieved through thorough exploratory data analysis and fitting multivariate statistical models. Therefore, the study was done in 3 main steps: data collection and sampling, exploratory data analysis, and an advanced analysis.

Sampling is the process of drawing a representative sample from the population to analyze the characteristics of the whole

population using a specialized sampling procedure [5].

The sample size was calculated using Slovin's formula. After adding 20% non-response rate to the result obtained from the Slovin's formula, the survey was sent to 410 current active undergraduates from a population of 2251 active undergraduates.

$$n = \frac{N}{\left(1 + Ne^2\right)} \qquad (1)$$

where N is the population size and e is the margin of error.

It is necessary to have a sample that is a representation of all parts of the population. Thus, the multistage stratified sampling technique is employed as the sampling technique because it better covers the population and gives the researchers more control over the subgroups and the ability to ensure their inclusion. Grouping variables that were considered for stratification were the academic year and the subject stream. Proportional allocation was used to allocate observations for each stratum. The questionnaire was created using Google Forms. 32 questions were there in the questionnaire. The responder's oral communication skill level was categorized under four different aspects (presenting ideas verbally, listening skills, skills in giving feedback, and presentation skills) and the four aspects were selected from a previous study done by Iksan et al. (2012) [6] which was used to evaluate verbal communication skills in their study. Iksan et al. (2012) focused on communication skills among undergraduates while this study focuses on the English oral communication skills of undergraduates. Each aspect was measured by number of sub-items. Those sub-items were evaluated using a point likert scale and take the average. The pilot test was carried out for 24 undergraduates. A reliability analysis was done for the response variables in the pilot study to check the internal consistency of responses using the Alpha Cronbach. All Alpha Cronbach values were higher than 0.8, indicating good internal consistency.

Exploratory Data Analysis (EDA) is a procedure for reviewing and analyzing data sets in order to highlight their key features using statistical and visual techniques. EDA was done by considering univariate and bivariate and the bivariate analysis was done using graphical illustrations and numerical methods. In the univariate analysis it was found that out of the four aspects of oral communication skills, the listening skills of undergraduates are higher compared to others. In bivariate analysis, graphical illustrations highlighted that type of the school attended for Advanced level, medium of the language in which A/L examination was attended, result for English subject in Ordinary level examination and Advanced level examination, result for EC 1001 (English) course that was done in the undergraduate level at the faculty, current GPA of the student and participating in clubs/societies at school have a comparably high potential impact on the four components of the oral communication skills. A nonparametric multivariate analysis of variance method called permutational MANOVA test [7] was conducted to identify if there were any difference in oral communication skill levels of undergraduates among the levels of the academic year as well as the subject stream. For both variables, the test was insignificant. The factors that was highlighted in the EDA was also founded to be significant in the results of permutational MANOVA. The Spearman's rank correlation coefficient test implemented to find the association between quantitative variables and the response variables and the Z score of the A/L examination was significant for each of the four components of oral communication abilities. Cramer's V was used to identify associations between categorical predictors and there were no important associations to consider despite some obvious relationships.

The researcher attempts to accomplish the main goal, which is to identify the critical factors that influence undergraduates' oral communication abilities, by employing statistical learning approaches. Since the oral communication abilities were evaluated using four aspects, a multivariate approach is used in the advanced analysis. The advanced analysis began with identifying and removing two outlier observations that were identified in the exploratory data analysis. One-hot encoding was done to convert categorical variables into numerical format. The data was then split into a train and test set. Min-max normalization was applied to bring qualitative variables to the same scale. The response variables were shown to be correlated with one another in the findings of the exploratory data analysis, and multicollinearity may be evident given a large number of predictor factors. So, methods that address those problems must be put into practice. Consequently, regularized regression methods (multivariate ridge regression, multivariate lasso regression, and multivariate elastic net regression), and partial least square regression

were applied as the statistical learning techniques. Due to the scenario of multiple dependent variables and the correlation between predictors (multicollinearity) as well as the response variables, modelling such data is not a straightforward case. For handling such data in the context of univariate regression, there are several tools available. One of the most widely used techniques to counteract effects like multicollinearity is regularization regression. However, in the context of multivariate regression modelling, these techniques have not been widely studied. Finch & Hernández-Finch (2017) [8] have carried out a simulation study to check the performance of regularization regression for the multivariate case along with Bayesian regression with non-informative normal priors method and Bayseian regression with spike and slab priors method. The results revealed that the Ridge Regression was found to be particularly effective in terms of parameter estimation accuracy and control over the Type I error rate. These regularized regressions are more useful when the response variables are correlated with each other [9]. When more than one response variable depends on the same set of predictor factors, multivariate ridge regression, an extension of the conventional ridge regression model, is utilized.

The multivariate linear regression model without an intercept can be written as;

$$Y = X\beta + \epsilon \qquad (2)$$

where:
- Y is a $n \times k$ matrix, where n is the number of observations and k is the number of dependent variables;
- X is the data matrix, where the dimension of the matrix is $n \times p$ (p is the number of independent variables);
- β is coefficient matrix, where the dimension of the matrix is $p \times k$.

In ridge regression, a L2 norm penalty term is introduced to the loss function to counterpart issues like overfitting. The loss function for multivariate ridge regression is as follows.

$$\arg\min_{\beta \in \Re^{p \times k}} \left\| Y - X\beta \right\|^2 + \lambda \left\| \beta \right\|^2 \qquad (3)$$

where $\left\| . \right\|$ is the L2 norm and λ is the tuning parameter that controls the amount of shrinkage applied to the regression coefficients. 10-fold cross-validation is utilized to find the optimal λ.

The only difference in the lasso regression from ridge regression is the loss function. The loss function of the multivariate lasso regression is as follows.

$$\arg\min_{\beta \in \Re^{p \times k}} \left\| Y - X\beta \right\|^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right| \qquad (4)$$

where β_j is the j-th row of the $p \times k$ coefficient matrix β and λ is a group-lasso penalty on each coefficient k-vector β_j for a single predictor (i.e. column of the data matrix). The group lasso penalty behaves similarly to the regular lasso penalty, except that it applies to the entire group of coefficients for each response. Either all of the coefficients are zero, or none of them are zero but are shrunk by a certain amount depending on λ. Here also the optimal value for λ is determined by the cross-validation technique in practice.

Elastic net regression is a generalization of both ridge and lasso regression, combining the two penalties. The loss function of the multivariate elastic net regression is as follows.

$$\arg\min_{\beta \in \Re^{p \times k}} \left\| Y - X\beta \right\|^2 + \lambda \left[(1-\alpha) \left\| \beta \right\|^2 + \alpha \sum_{j=1}^{p} \left| \beta_j \right| \right] \qquad (5)$$

Optimal $\alpha \left(0 < \alpha < 1 \right)$ for the multivariate elastic net regression is determined using a validation set.

Partial least square (PLS) regression is a technique that generalizes information from multiple regression and principal component analysis [10]. PLS regression is used to predict or analyze a set of response variables (especially when they are correlated) from a set of predictor variables. This technique also counterpart the multicollinearity issue. PLS regression decomposes both X matrix and Y matrix as a product of a set of orthogonal factors and a set of loadings. The X matrix is decomposed as

$$X = TP' \qquad (6)$$
$$T'T = I \qquad (7)$$

where I is the identity matrix. Here T is called the "score matrix" and P is the "loading matrix". Similarly, the response matrix is decomposed as

$$\widehat{Y} = TBC' \qquad (8)$$

where B is a diagonal matrix with "regression weights" as diagonal elements and C is the "weight matrix" of the dependent variables. The columns of the T matrix are the "latent vectors" (Abdi, 2010). PLS regression has two main algorithms, named "NIPALS" and

"SIMPLS", to accomplish its basic objective, which is to identify components that best capture the covariance between X and Y. In this study, the "plsr()" function in the "pls" library in R, which uses the "SIMPLS" algorithm, was used to perform PLS regression. A 10-fold cross-validation approach is used to determine the optimal number of latent variables.

3. Results and discussion

From the fitted models Partial Least Square regression model has the lowest test RMSE (0.3466) and it has the highest R2 value for all four response variables (0.2105, 0.2267, 0.0663, 0.2654). Therefore, the Partial Least Square regression model with 3 latent variables is selected as the best-fitted model.
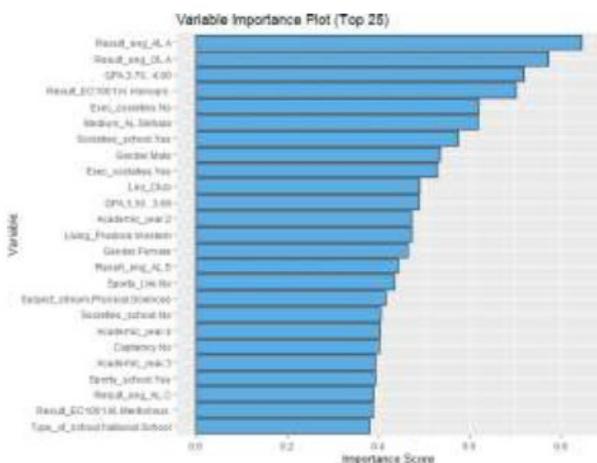

Fig. 1. Variable Importance Plot

The importance score for each variable was calculated by taking the summation of the absolute value of loadings of the 3 latent variables. Then the variable importance plot was drawn for the top 25 variables. According to the variable importance plot, result for English subject in Ordinary level examination and Advanced level examination, current GPA of the student, and whether the student held any executive positions in the clubs or societies has the highest association with the response. The main goal of the study is to analyze the factors that influence the English oral communication among undergraduates. An exploratory data analysis and an advanced analysis were carried out to achieve the mentioned objectives. The study reveals that out of the most 5 important factors that associated with undergraduates' English oral communication skills, three of them are the results of examinations that are related to English language in the advance analysis of

the study, which shows that those students who performed well on those exams have a higher level of oral communication abilities than other students. Also, the study has emphasized the importance of involvement in extracurricular societies/clubs on the development of oral communication abilities. Another interesting fact revealed from the EDA is that the students who went to private schools have higher English oral communication abilities that the ones who went to national and provincial schools.

4. Bibliography

[1] Crosling G., Ward I., "Oral communication: The workplace needs and uses of business graduate employees", *English for Specific Purposes*, vol. 21(1), 41–57 (2002).

[2] Mohamed A.-A., Radzuan N.R., Kassim H., Azmat Ali M.M., "Conceptualizing english workplace communication needs of professional engineers", *Selangor Business Review*, vol. 1(1), 1–11 (2017).

[3] Darling A.L., Dannels D.P., "Practicing engineers talk about the importance of talk: A report on the role of oral communication in the workplace", *Communication Education*, vol. 52(1), 1–16 (2003).

[4] Collins J., "Education techniques for lifelong learning: principles of adult learning", *Radiographics*, vol. 24(5), 1483–1489 (2004).

[5] Rahman Md M., et al., "Sampling techniques (probability) for quantitative social science researchers: a conceptual guidelines with examples", *Seeu Review*, vol. 17(1), 42–51 (2022).

[6] Iksan Z.H., et al., "Communication skills among university students", *Procedia – Social and Behavioral Sciences*, vol. 59, 71–76 (2012).

[7] Anderson M.J., "A new method for nonparametric multivariate analysis of variance", *Austral Ecology*, vol. 26(1), 32–46 (2001).

[8] Finch W.H., Hernández-Finch M.E., "Multivariate regression with small samples: A comparison of estimation methods", *General Linear Model Journal*, vol. 43(1), 16–30 (2017).

[9] Friedman J., et al., "Package *glmnet*", *CRAN R Repositary*, 2021.

[10] Abdi H., "Partial least squares regression and projection on latent structure regression (pls regression)", *WIREs Computational Statistics*, vol. 2(1), 97−106 (2010).

Analiza czynników związanych z postrzeganą znajomością języka angielskiego. Umiejętności komunikacyjne wśród studentów: studium przypadku

D.S. RUWANKUMARA, E.R.A.D. BANDARA

W tym studium przypadku zbadano czynniki wpływające na umiejętności komunikacji ustnej w języku angielskim wśród studentów Wydziału Nauk Uniwersytetu w Kolombo. Przebadano 332 studentów studiów licencjackich, stosując dobór warstwowy na podstawie roku akademickiego i kierunku studiów. Ocena ich umiejętności komunikacji ustnej obejmowała cztery krytyczne aspekty: werbalne prezentowanie pomysłów, umiejętności słuchania, przekazywanie informacji zwrotnej i umiejętności prezentacji. Wykorzystując eksploracyjną analizę danych (EDA) i zaawansowane modele statystyczne, w tym wielowymiarowy grzbiet, lasso, regresję siatki elastycznej i regresję częściową metodą najmniejszych kwadratów, badanie odkrywa, że najważniejszymi czynnikami wpływającymi na umiejętności komunikacji ustnej w języku angielskim były wyniki egzaminów z języka angielskiego zdawanych podczas wykształcenia średniego i studiów licencjackich. Uczniowie, którzy dobrze wypadli na tych egzaminach, wykazali się doskonałymi umiejętnościami komunikacyjnymi, a także pozytywnym wpływem zajęć pozalekcyjnych, ujawnionym przez EDA. Wyniki badania stanowią wskazówki dotyczące praktyki komunikacji ustnej w języku angielskim dla przyszłych karier zawodowych i edukacyjnych uczniów.

Słowa kluczowe: komunikacja ustna w języku angielskim, wielowymiarowa regresja uregulowana, regresja częściowa metodą najmniejszych kwadratów.