

**DEVELOPMENT OF ESL SPEAKING  
PROFICIENCY ASSESSMENT FOR  
UNDERGRADUATES**

**KARWAN MUSTAFA SAEED**

**UNIVERSITI SAINS MALAYSIA**

**2018**

**DEVELOPMENT OF ESL SPEAKING  
PROFICIENCY ASSESSMENT FOR  
UNDERGRADUATES**

**by**

**KARWAN MUSTAFA SAEED**

**Thesis submitted in fulfilment of the requirements  
for the degree of  
Doctor of Philosophy**

**July 2018**

## ACKNOWLEDGEMENT

First of all, I would like to thank God, the one above all of us and omnipresent for answering my prayers for giving me the strength, health, and ability to accomplish this thesis despite all the circumstances.

I would like to extend my deepest gratitude to my academic supervisor, associate professor Dr. Shaik Abdul Malik Mohamed Ismail. Under his guidance, I have been able to complete my Ph.D. thesis successfully. He was always ready to listen as well as provide valuable feedback throughout the research. His feedback has always been insightful and greatly enhanced this thesis. Indeed, he has always been a tremendous support in my tedious journey.

I am also indebted to my co-supervisor, Dr. Lin Siew Eng. Indeed, she has always shown her support. I thank her for her continuous encouragement in the course of completing my thesis.

Last but not least, I truly would like to thank and extend my sincere gratitude to my beloved family members for always being there for me, in particular, my father who has always supported and encouraged me. Words fail me in expressing my heartfelt thanks to my mother for her unquestioning love and patience, understanding and believe in my effort and for her constant support and prayers.

Karwan Mustafa Saeed

July 2018

## TABLE OF CONTENTS

|                                      |           |
|--------------------------------------|-----------|
| Acknowledgement.....                 | ii        |
| Table of Contents.....               | iii       |
| List of Tables.....                  | ix        |
| List of Figures.....                 | xi        |
| List of Abbreviations.....           | xiii      |
| Abstrak.....                         | xiv       |
| Abstract.....                        | xvi       |
| <b>CHAPTER 1 – INTRODUCTION</b>      | <b>1</b>  |
| 1.1 Introduction.....                | 1         |
| 1.2 Background of the Study.....     | 4         |
| 1.3 Statement of the Problem.....    | 8         |
| 1.4 Rationale for the Study.....     | 12        |
| 1.5 Objectives of the Study.....     | 13        |
| 1.6 Research Questions.....          | 15        |
| 1.7 Significance of the Study.....   | 16        |
| 1.8 Definition of the Terms .....    | 18        |
| 1.9 Summary.....                     | 20        |
| <b>CHAPTER 2 – LITERATURE REVIEW</b> | <b>21</b> |
| 2.1 Introduction.....                | 21        |
| 2.2 Speaking.....                    | 21        |
| 2.3 Importance of Speaking.....      | 25        |
| 2.4 Assessment.....                  | 27        |
| 2.4.1 Speaking Assessment.....       | 29        |

|                                |   |           |
|--------------------------------|---|-----------|
| 2.4.2                          | Summative Assessment.....   | 31        |
| 2.4.3                          | Formative Assessment.....   | 33        |
| 2.4.4                          | Validation of Test Instrument.....  | 35        |
| 2.5                            | Speaking Proficiency Descriptors.....   | 39        |
| 2.6                            | Identifying Respondents' Speaking Proficiency.....  | 42        |
| 2.7                            | Profiling Undergraduate Students' Speaking Proficiency.....                                       | 44        |
| 2.8                            | Theories Related to Speaking Proficiency .....  | 48        |
| 2.8.1                          | Littlewood's Methodological Framework.....  | 48        |
| 2.8.2                          | Interaction Hypothesis.....   | 50        |
| 2.9                            | English Language Syllabus offered by the School of Languages, Literacies,<br>and Translation..... | 52        |
| 2.10                           | Kathleen Bardovi-Harlig's Operationalizing Conversation Speech Acts.....                          | 53        |
| 2.11                           | Public Exams: IELTS and MUET Past Year Examinations.....  | 54        |
| 2.12                           | Conceptual Framework.....   | 55        |
| 2.13                           | Conclusion.....   | 58        |
| <b>CHAPTER 3 – METHODOLOGY</b> |   | <b>59</b> |
| 3.1                            | Introduction.....   | 59        |
| 3.2                            | Research Design.....  | 59        |
| 3.3                            | Research Procedure.....   | 62        |
| 3.4                            | Development of Prototype Speaking Proficiency Tests (PSPT).....                                   | 62        |
| 3.5                            | Validity.....   | 64        |
| 3.6                            | Reliability.....  | 66        |
| 3.7                            | Scoring Rubrics.....  | 67        |
| 3.8                            | Interview Protocols.....  | 72        |
| 3.9                            | Administering Questionnaire and ESL Speaking Proficiency Test.....                                | 74        |
| 3.9.1                          | Population and Sampling Procedure.....  | 74        |

|   |   |           |
|---|---|-----------|
| 3.9.2   | Gathering Data for Analysis.....  | 76        |
| 3.10  | Data Analysis.....  | 77        |
| 3.10.1  | Quantitative Data.....  | 77        |
| 3.10.2  | Qualitative Data.....   | 78        |
| 3.11  | Identifying the Undergraduates’ Speaking Proficiency.....   | 80        |
| 3.12  | Profiling the ESL Speakers’ Speaking Proficiency.....   | 80        |
| 3.13  | Summary.....  | 80        |
| <b>CHAPTER 4 – DEVELOPMENT OF ESL SPEAKING PROFICIENCY TEST</b> |   | <b>83</b> |
| 4.1   | Introduction.....   | 83        |
| 4.2   | Development of ESL Prototype Speaking Proficiency Test.....   | 83        |
| 4.2.1   | Selection of Relevant Sources for the Construction of the ESL Prototype Speaking Proficiency Test.....  | 85        |
| 4.2.1(a)  | The Syllabus of the English Language Courses Offered at the School of Languages, Literacies, and Translation at USM.....                      | 85        |
| 4.2.1(b)  | Littlewood’s Methodological Framework (1981).....   | 87        |
| 4.2.1(c)  | Kathleen Bardovi-Harlig’s Operationalizing Conversation Speech Acts (2015).....   | 89        |
| 4.2.1(d)  | Public Examinations: IELTS Part B Speaking test and MUET Past Year Examinations.....  | 89        |
| 4.2.1(e)  | Long’s Interaction Hypothesis (1981).....   | 90        |
| 4.2.2   | Determining the Best Possible Combination of ESL Speaking Proficiency Questions Selected for the ESL Prototype Speaking Proficiency Test..... | 92        |
| 4.2.3   | Pilot Study to Test Validity and Reliability of the ESL Prototype Speaking Proficiency Test.....  | 96        |
| 4.2.3(a)  | Validity of ESL Prototype Speaking Proficiency Test...  | 100       |
| 4.2.3(a)(i)   | Content Validity.....   | 100       |
| 4.2.3(a)(ii)  | Construct Validity.....   | 103       |

|   |  |            |
|---|--|------------|
| 4.2.3 (b)                                     | Reliability of ESL Prototype Speaking Proficiency Test...  | 107        |
| 4.2.3(b)(i)                                   | Parallel-Forms Reliability.....  | 108        |
| 4.2.3(b)(ii)                                  | Internal Consistency Reliability.....  | 109        |
| 4.2.3(b)(iii)                                 | Interrater Reliability of Tests Scores<br>(Pilot Study: 96 Respondents).....   | 110        |
| 4.2.4   | Time Allocated for the ESL Speaking Proficiency Test.....  | 113        |
| 4.3   | Development of ESL Speaking Proficiency Test.....  | 114        |
| 4.4   | Summary.....   | 115        |
| <b>CHAPTER 5 – DATA ANALYSIS AND FINDINGS</b> |  | <b>117</b> |
| 5.1   | Introduction.....  | 117        |
| 5.2   | Developing a Set of Descriptors to Identify the Undergraduates’ ESL<br>Speaking Proficiency.....                           | 118        |
| 5.2.1   | Determining the Cut Scores for the Bands.....  | 118        |
| 5.2.2   | Identifying Speaking Skills Undergraduates have Acquired.....  | 125        |
| 5.3   | Identifying Undergraduates’ ESL Speaking Proficiency.....  | 144        |
| 5.3.1   | Respondents’ General ESL Speaking Proficiency .....  | 147        |
| 5.3.2   | Respondents’ Overall ESL Speaking Proficiency Performance<br>According to Gender, Ethnicity, University and<br>School..... | 148        |
| 5.3.2(a)                                      | Respondents’ Overall ESL Speaking Proficiency<br>According to Gender.....  | 149        |
| 5.3.2(b)                                      | Respondents’ Overall ESL Speaking Proficiency<br>According to Ethnicity.....   | 152        |
| 5.3.2(c)                                      | Respondents’ Overall ESL Speaking Proficiency<br>According to University.....  | 155        |
| 5.3.2(d)                                      | Respondents’ Overall ESL Speaking Proficiency<br>According to School.....  | 158        |
| 5.3.2(d)(i)                                   | Schools in University A.....   | 158        |
| 5.3.2(d)(ii)                                  | Schools in University B.....   | 159        |

|  |   |            |
|--|---|------------|
| 5.4  | Profiling Undergraduates' ESL Speaking Proficiency.....   | 160        |
| 5.4.1  | Profiles of University A Undergraduates' Performance.....   | 162        |
| 5.4.1(a)   | Profiles of Superior Performers.....  | 163        |
| 5.4.1(b)   | Profiles of Advanced Performers.....  | 165        |
| 5.4.1(c)   | Profiles of Intermediate Performers.....  | 167        |
| 5.4.1(d)   | Profiles of Novice Performers.....  | 169        |
| 5.4.2  | Profiles of University B Undergraduates' Performance.....   | 169        |
| 5.4.2(a)   | Profiles of Superior Performers.....  | 169        |
| 5.4.2(b)   | Profiles of Advanced Performers.....  | 171        |
| 5.4.2(c)   | Profiles of Intermediate Performers.....  | 173        |
| 5.4.2(d)   | Profiles of Novice Performers.....  | 175        |
| 5.4.3  | The Speakers' Speaking Proficiency Profile.....   | 175        |
| 5.4.3(a)   | University A Speakers' Speaking Proficiency Profile...  | 175        |
| 5.4.3(b)   | University B Speakers' Speaking Proficiency Profile...  | 176        |
| 5.5  | Conclusion.....   | 177        |
| <b>CHAPTER 6 – DISCUSSION OF FINDINGS AND CONCLUSION</b> |   | <b>179</b> |
| 6.1  | Introduction.....   | 179        |
| 6.2  | Discussion of Findings.....   | 179        |
| 6.2.1  | Development of an ESL Speaking Proficiency Test for Undergraduates.....                           | 180        |
| 6.2.2  | Development of a Set of Descriptors to Identify the Undergraduates' ESL Speaking Proficiency..... | 185        |
| 6.2.3  | Identifying Undergraduates' ESL Speaking Proficiency.....   | 187        |
| 6.2.4  | Profiling the Undergraduates' ESL Speaking Proficiency.....                                       | 192        |
| 6.3  | Overview of the Study.....  | 194        |
| 6.4  | Restatement of the Objectives.....  | 195        |



|     |   |            |
|-----|---|------------|
| 6.5 | Pedagogical Implications of the Findings..... | 196        |
| 6.6 | Limitations of the Study.....                 | 198        |
| 6.7 | Recommendations for Stakeholders.....         | 199        |
| 6.8 | Recommendations for Further Research.....     | 201        |
| 6.9 | Summary.....                                  | 202        |
|     | <b>REFERENCES.....</b>                        | <b>204</b> |
|     | <b>APPENDICES</b>                             |            |

## LIST OF TABLES

|            |   | <b>Page</b> |
|------------|---|-------------|
| Table 2.1  | The ACTFL Proficiency Guidelines-Speaking   | 45          |
| Table 2.2  | Common Reference Levels: Speaking   | 47          |
| Table 3.1  | Assessment Rubrics for Speaking Proficiency   | 71          |
| Table 3.2  | Respondents of Study  | 76          |
| Table 3.3  | Labels of Codes of Participants   | 79          |
| Table 3.4  | Research Matrix   | 81          |
| Table 4.1  | Speaking Tasks Found in the Syllabus of the English Courses at the School of Languages, Literacies, and Translation (SOLLAT)  | 86          |
| Table 4.2  | Littlewood’s Methodological Framework   | 88          |
| Table 4.3  | Summary of Sources of Data for PSPT   | 91          |
| Table 4.4  | Sections of ESL Prototype Speaking Proficiency Test   | 96          |
| Table 4.5  | Respondents of the Pilot Study  | 98          |
| Table 4.6  | Ratings of Content Experts  | 102         |
| Table 4.7  | Comparison of the Mean Scores of Respondents’ Speaking Proficiency Based on Their English Proficiency for Set One and Set Two | 105         |
| Table 4.8  | Reliability of the Prototype Standardized Speaking Proficiency Test   | 110         |
| Table 4.9  | Interrater Reliability of the Prototype Speaking Proficiency Test   | 112         |
| Table 4.10 | Pilot Test to Find Out Time Taken to Complete PSPT  | 114         |
| Table 5.1  | Descriptive Statistics for Pilot Study  | 121         |
| Table 5.2  | Establishing Scores for Bands   | 123         |
| Table 5.3  | Cut scores and Categories of Performers   | 124         |
| Table 5.4  | Mean Score of Respondents’ Speaking Proficiency Performance at Different Sections of PSPT                                     | 125         |
| Table 5.5  | Frequency and Percentages of Respondents in Performance Bands   | 126         |

|            |  |     |
|------------|--|-----|
| Table 5.6  | Terms Used in Speaking Proficiency Descriptors   | 128 |
| Table 5.7  | Clarity of Sections A, B, and C  | 137 |
| Table 5.8  | Difficulties Encountered During the Test   | 139 |
| Table 5.9  | Preference of Types of Speaking Questions  | 142 |
| Table 5.10 | Percentages of University A Respondents According to Performance Bands for Each School | 145 |
| Table 5.11 | University A Respondents' Speaking Proficiency   | 145 |
| Table 5.12 | Percentages of University B Respondents According to Performance Bands for Each School | 146 |
| Table 5.13 | University B Respondents' Speaking Proficiency   | 147 |
| Table 5.14 | University A and University B Respondents' Speaking Proficiency                        | 148 |
| Table 5.15 | School Performance   | 159 |
| Table 5.16 | School Performance   | 160 |
| Table 5.17 | University A Undergraduate Speakers' Profile (N=80 Respondents)                        | 176 |
| Table 5.18 | University B Undergraduate Speakers' Profile (N=60 Respondents)                        | 177 |

## LIST OF FIGURES

|             |  | <b>Page</b> |
|-------------|--|-------------|
| Figure 2.1  | Representation of conceptual framework of study  | 57          |
| Figure 3.1  | Research design  | 61          |
| Figure 4.1  | Relationship between respondents of English language proficiency levels and their speaking | 107         |
| Figure 5.1  | The normal curve relationship between z-score and location in a standard distribution      | 120         |
| Figure 5.2  | Cut scores based on z-scores   | 123         |
| Figure 5.3  | Respondents' acquired speaking skills  | 131         |
| Figure 5.4  | Percentages of male and female respondents   | 149         |
| Figure 5.5  | Percentages of male and female respondents at the various bands                            | 150         |
| Figure 5.6  | Gender performance   | 152         |
| Figure 5.7  | Percentages of respondents of different ethnic groups                                      | 153         |
| Figure 5.8  | Percentages of different ethnic groups at the various bands                                | 153         |
| Figure 5.9  | Ethnicity performance  | 155         |
| Figure 5.10 | Percentages of respondents of university A and university B                                | 156         |
| Figure 5.11 | Percentages of respondents in university A and university B at the various bands           | 156         |
| Figure 5.12 | University performance   | 158         |
| Figure 5.13 | Speaking proficiency performance bands/speaking proficiency descriptors                    | 161         |
| Figure 5.14 | University A superior performers   | 163         |
| Figure 5.15 | Percentages of respondents as superior performers  | 164         |
| Figure 5.16 | University A advanced performers   | 165         |
| Figure 5.17 | Percentages of respondents as advanced performers  | 166         |
| Figure 5.18 | University A intermediate performers   | 167         |

|             |   |     |
|-------------|---|-----|
| Figure 5.19 | Percentages of respondents as intermediate performers | 168 |
| Figure 5.20 | University B superior performers                      | 170 |
| Figure 5.21 | Percentages of respondents as superior performers     | 170 |
| Figure 5.22 | University B advanced performers                      | 171 |
| Figure 5.23 | Percentages of respondents as advanced performers     | 172 |
| Figure 5.24 | University B intermediate performers                  | 173 |
| Figure 5.25 | Percentages of respondents as intermediate performers | 174 |

## **LIST OF ABBREVIATIONS**

|        |   |
|--------|---|
| ESL    | English as a Second Language                              |
| TESOL  | Teaching English to Speakers of other Languages           |
| IELTS  | International English Language Testing System             |
| MUET   | Malaysian University English Test                         |
| TOEFL  | Test of English as a Foreign Language                     |
| PSPT   | Prototype Speaking Proficiency Test                       |
| ACTFL  | The American Council on the Teaching of Foreign Languages |
| CEF    | Common European Framework                                 |
| SOLLAT | School of Languages, Literacies, and Translation          |
| TOEIC  | Test of English for International Communication           |
| SPT    | Speaking Proficiency Test                                 |
| DCEST  | Diagnostic College English Speaking Test                  |

**PEMBANGUNAN PENILAIAN KEMAHIRAN BERTUTUR PELAJAR  
IJAZAH SARJANA MUDA DALAM BAHASA INGGERIS SEBAGAI  
BAHASA KEDUA**

**ABSTRAK**

Bahasa Inggeris dianggap sebagai aset yang boleh menghasilkan kejayaan dalam pasaran kerja abad ke-21. Oleh itu, menguasai seni bertutur dalam kalangan pelajar bahasa dilihat sebagai aspek bahasa yang paling penting ketika mempelajari bahasa Inggeris. Penilaian bertutur telah menjadi satu perkara yang penting dalam bidang pengajaran dan pedagogi bahasa dalam usaha melaksanakan penambahbaikan. Perkara ini penting kepada para pengajar dalam menilai kemahiran bertutur dalam kalangan pelajar mereka secara berterusan, dan oleh sebab itu, keputusan ujian mesti dapat membantu para pengajar menentukan kemahiran bertutur seseorang pelajar. Sungguhpun begitu, penilaian semasa yang berdasarkan sesuatu gred tidak memberi maklumat khusus mengenai kemahiran khusus bertutur pelajar. Gred yang disediakan, tidak menunjukkan maklumat berhubung cara membantu pelajar meningkatkan kemahiran bertutur mereka. Oleh itu, instrumen ujian yang tidak hanya menyediakan skor ujian tetapi juga membantu para pengajar bahasa untuk mengenal pasti kekuatan dan kelemahan pelajar mereka amat penting. Hal ini menjelaskan bahawa objektif pertama kajian semasa adalah untuk membangunkan satu ujian kemahiran bertutur bahasa Inggeris sebagai bahasa kedua bagi pelajar ijazah sarjana muda. Kajian ini memberi makna kepada penilaian kemahiran bertutur melalui satu instrumen ujian dan komponen deskriptor kemahiran bertutur untuk menerangkan kemahiran bertutur seseorang pelajar. Dengan berbuat demikian, ujian prototaip kemahiran bertutur

pertama kali telah dibuat, kemudian disaring berdasarkan cadangan dua ahli TESOL. Ujian itu kemudiannya diuji ke atas 96 orang pelajar ijazah sarjana muda di sebuah universiti awam di Malaysia. Penyelidik seterusnya membangunkan tahap skor untuk menetapkan band mengikut prestasi; Band 1, Band 2, Band 3, dan Band 4. Berdasarkan kemahiran bertutur responden, deskriptor kemahiran bercakap dibina untuk mendiagnosis kemahiran bertutur pelajar ijazah sarjana muda. Instrumen yang diuji ini dijalankan ke atas 140 pelajar tahun pertama di dua buah universiti awam di utara Semenanjung Malaysia dan data yang diperolehi dianalisis untuk menanda aras pelajar. Akhirnya, profil penutur dalam kategori prestasi berbeza telah ditakrif dan dirangkaikan. Dapatan kajian menunjukkan bahawa daripada 140 responden tersebut, 56.43% dikategorikan sebagai "penutur pertengahan", 36.43% dikategorikan sebagai "penutur maju", dan hanya 7.14% dikategorikan sebagai "penutur unggul". Seperti yang dijangkakan, tiada pelajar dikategorikan sebagai "penutur baharu". Pensyarah bahasa Inggeris sebagai bahasa kedua boleh menggunakan ujian yang dibina untuk mengenal pasti masalah bertutur dalam kalangan pelajar kemudiannya memaklumkan keputusan dalam usaha meningkatkan bahan pengajaran yang disesuaikan dengan keperluan pelajar. Kajian semasa ini hanya melibatkan pelajar ijazah sarjana muda dari dua universiti. Walau bagaimanapun, penglibatan pelajar ijazah sarjana muda universiti lain dari seluruh negara perlu dipertimbangkan bagi penyelidikan masa depan agar keputusan umum kajian akan dapat dijamin.



# **DEVELOPMENT OF ESL SPEAKING PROFICIENCY ASSESSMENT FOR UNDERGRADUATES**

## **ABSTRACT**

English is regarded as an asset that can result in success in the twenty-first-century job market. Hence, mastering the art of speaking to many language learners is seen as the most important language aspect when learning the English language. Speaking assessment has made significant inroads into the field of language teaching and pedagogy in pursuit of improvement. It is vital that instructors assess their students' speaking proficiency continuously and therefore, test results must assist instructors to determine students' speaking proficiency. However, the current assessment based on grades does not provide specific information regarding students' specific speaking proficiency. The grade provided does not indicate information on how to help students enhance their speaking proficiency. Therefore, a test instrument that not merely provides test scores but also aids language instructors to recognize their students' strengths and weaknesses is crucially vital. The first objective of the current study is to develop an ESL speaking proficiency test for undergraduates. The present study gives meaning to speaking proficiency assessment through a test instrument and speaking proficiency descriptor components to describe students' speaking proficiency. In doing so, a prototype speaking proficiency test was first devised and then refined based on recommendations of two TESOL experts. The test was then piloted over 96 undergraduates at a public university in Malaysia. The researcher then developed the cut scores to establish the performance bands namely; Band One, Band Two, Band Three and Band Four. Based on the respondents' speaking proficiency performance,

the speaking proficiency descriptors were established to identify the undergraduates' speaking proficiency. The tested instrument was administered over 140 first-year undergraduates at two public universities in northern region Peninsular of Malaysia and the data gathered were analyzed to identify the students' performance. Finally, the profiles of the speakers in the different performance categories were defined and described. The findings showed that out of the 140 students who served as participants of the study, 56.43% of the respondents were categorized as 'intermediate performers', 36.43% of the respondents were categorized as "advanced performers" and only 7.14% of the respondents were categorized as "superior performers". As expected, no students were categorized as "novice performers". The ESL language lecturers can use the developed test to identify undergraduates' speaking problems and inform decision making in pursuit of improving the teaching materials tailored towards students' needs. The current study has only included undergraduates from two universities. However, including undergraduates of other universities or nationwide involvement should be considered for future research so the generalization of the results would be guaranteed.

# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

Nowadays, learning a second or foreign language is considered as an essential aspect in the curricula at different levels in education across the world, in particular, teaching and learning English. Since the beginning of the new era, English has grown internationally and achieved the status of “great international language” (House, 2002). Hereafter, it has been recognized as the worldwide language for exchanging information and knowledge and communication purposes (Cenoz & Jessner, 2000). As a result, over the past few decades, the English language has obtained its magnificence as a lingua franca (Risager, 2007).

For the past three decades, a number of models of communicative competence have been developed in the field of language teaching and testing which included (Canale & Swain, 1980; Canale, 1983; Bachman, 1990; Bachman & Palmer, 1996; Bachman & Palmer, 2010). Generally speaking, these communicative models of language teaching and testing argue that language should be taught based on communicative activities in the classroom. In addition, these models of communicative language approach provide useful frameworks for designing language tests.

Communicating in the English language is of paramount importance across the world. The English language is widely spoken in Malaysia. As such, Malaysian students should be well prepared to remain competitive, in particular, communicating in English efficiently (Abdullah & Rahman, 2010). They added that the English language is extensively used in different fields of life such as legal and business. Therefore, it is crucially significant for Malaysian students to be well prepared to speak the language fluently. As a result, speaking proficiency has been a significant portion of the curriculum in language teaching and learning and this makes speaking proficiency an important object of assessment as well.

Moreover, language assessment plays a vital role in language teaching and learning; it is a matter of concern to those who are engaged in the education sector whether they are teachers or researchers (Dahan, 2012). Furthermore, Bachman and Palmer (1996) argue that language tests can be an important tool for providing information in terms of language teaching and monitoring the process of learning. Therefore, in order for lecturers and teachers to develop speaking proficiency of their students, they must include speaking tests in their on-going evaluation, even with large classes, communicative tests can still be applied (Cross, 1991).

Students are assessed after a course of study which shows only what they have learned at the end of the course they have taken (Oosterhof, 2001). Consequently, the assessments are just judgments of a student's performance and they do not display what the students can do and what they cannot do. They only present that some students are better than others using grades or percentages. The problem with grades or percentage

is that they do not provide any information or clue on how to help the students. Instead, they only indicate the overall result or achievement of the student (Oosterhof, 2001). For this reason, this study attempts to fill this gap by administering an ESL speaking proficiency test whereby language lecturers would be able to monitor their students' speaking proficiency improvement based on *assessment for learning*.

According to Brown (2004), assessment is perhaps the most significant aspect language lecturers and teachers can do to assist students to learn. Therefore, assessment for learning (formative assessment) is vital in teaching and learning languages in which we can expand the extent to which our assessment practices are more developmental, rather only judgmental. Furthermore, Brown (2004) argues that if we intend to change the nature of assessment and integrate assessment into learning, assessment should be at the heart of the process of learning. Hence, language lecturers should provide not only where their students have gone wrong, but what they need in improving their language learning.

In such a case, according to Masters (2015), the most practical and crucial type of assessment is formative assessment in which lecturers can provide information where their students are in their learning which it can be used to make decisions for future planning. Hence, data about where the students are, elucidates their current situation and helps in recognizing starting points for action to take. Moreover, Masters (2015) states that when assessment is intended to help and guide future planning, the crucial goal is to boost learning.

Therefore, an ESL speaking proficiency test would be crucially useful because language instructors would be able to identify and profile their students. Accordingly, this study is designed to develop an ESL speaking proficiency test, develop possible speaking proficiency descriptors, identify the ESL learners' speaking proficiency as well as profiling the students' speaking proficiency based on formative assessment.

## **1.2 Background of the Study**

Assessment has become increasingly significant in higher education over the past two decades, and higher education institutions have provided student assessment for a variety of purposes (Ewell, 2002). Assessment is the process of collecting and explaining data and information from different sources to develop a comprehensive understanding pertaining to what students are able to do, know and understand with their current knowledge as a result of their learning process in order to enhance and ensure continuous learning (Huba & Freed, 2000). They added that assessment is also utilized for progressive improvement in institutional quality for the purpose of promoting the process of learning and program review.

Powerful and effective assessment practices are those that are meant for the purpose of improvement and sustainment of educational programs and services (Banta, 2002). According to a study by Peterson, Einarson, Augustine, and Vaughan (1999), it was discovered that assessment data that was reported by institutions influenced the educational decision-making concerning academic services, academic programs, educational curricula and developing materials according to the students' needs.

Taking into consideration the importance of speaking proficiency for non-native speakers of English language and their speaking proficiency improvement, speaking proficiency assessment has become the most paramount aspect of language teaching and testing because speaking has been recognized as more central and crucially important in language learning, in particular after the spreading of communicative language teaching approach (Nakamura, 1993).

Information about where the students are in their learning is crucially significant to identify the starting points to take action (Masters, 2013). Summative assessment is unable to determine the strengths and weaknesses of students to assist enhance their betterment in the future because it is usually taken at the end of a course of study or semester with the purpose of grading (Sadler, 1989, cited in Costel, Simona, Ana, & Stefan, 2015). If summative assessment results are reported, it is a passive measure because it does not have any direct influence on learning. In contrast with summative assessment, there is formative assessment which is an opportunity to improve the process of learning because it shows language learners' difficulties and provides information to improve the tasks of learning (Costel et al., 2015). Unfortunately, the current practices of speaking assessment based on grading or percentage scores will not be able to pin down the strengths and weaknesses of the undergraduates in their learning. For instance, if a student gets a 'B' in a summative test, it only resembles that the student has learned some skills included in the syllabus and, tells the English command of the student as a whole. It does not indicate the speaking proficiency of the student. Therefore, lecturers might not be able to help their students much in improving their English-speaking proficiency. Thus, graduates will not be able to master their

speaking proficiency well which will negatively affect their employment in their future career after graduation.

It is clear that tests of spoken language proficiency are the most difficult compared to testing the other language skills (Underhill, 1987). This is further supported by O'Sullivan (2008) who argues that it is generally true that it is challenging to conduct speaking tests. However, great improvements in the area of speaking assessment have been achieved over the past few years. Additionally, it has to be noted that assessment of speaking proficiency, besides being difficult to carry out, made many researchers and scholars conduct different research, including developing different speaking assessments and criteria (Luoma, 2004).

According to a study by Alberola Colomar (2014), it was found that there is a close and important interrelation between assessment and teaching in terms of developing students' speaking proficiency outcomes. The study revealed that the assessment procedure helped in redesigning and improving the course syllabus and class materials for the betterment of speaking proficiency of the learners. Based on the students' speaking proficiency, the teachers were able to ascertain which parts of the program should be modified for the purpose of betterment of speaking proficiency of the language learners, focusing on the needs of the students based on their weaknesses.

As Burke (1992) mentions, one of the benefits of assessment is that the universities, language lecturers, and students obtain enough information about what the students can do and what they cannot do. Therefore, speaking assessment must be used



as a tool for improvement of speaking proficiency of language learners. She also states that it is also vital for language lecturers and teachers to take into their consideration what a score means in terms of what the students are able to do and not able to do in order for them to take necessary action to help improve students' ability in speaking the language meaningfully.

A review of published works and research shows that performance indicators, identifying, setting standards and continuous assessment have brought about development and achievement in the students' performance in different sorts of areas in education, specifically in language learning. Similarly, ongoing formative assessment and standards are significant because they provide lecturers and teachers with analytical information of what the learners can and cannot do. As such, lecturers will be able to know where their students are so they will be able to help boost their students' language learning. Therefore, it is believed that identifying and profiling students' performance is of paramount importance for the betterment of quality of learning and achievement in different areas of education.

Based on the above discussion, we can conclude that little research has been done to identify and profile Malaysian undergraduates' speaking proficiency. Hence, more research should be done on how identifying and profiling the undergraduates' speaking proficiency will help language lecturers to monitor their students' speaking proficiency improvement. Therefore, the prime focus throughout this study is to develop an ESL valid and reliable speaking proficiency test and possible speaking proficiency descriptors to identify the ESL language learners' current level of speaking proficiency.

In addition, to identifying and profile the undergraduates' speaking proficiency so that language lectures will be able to help and guide their students in improving their speaking proficiency based on their current speaking proficiency.

### **1.3 Statement of the Problem**

Although English language proficiency has been a requirement for both academic life and workplace, Malaysian graduates still lack the language proficiency, especially speaking (Azman & Razak, 2007; Lan, Khaun, & Singh, 2011). The biggest challenge faced by the ESL language learners is expressing themselves in English clearly and fluently (Liu & Jackson, 2008).

Speaking proficiency plays a central role in securing job employment for Malaysian graduates (Lan et al., 2011). Notwithstanding that students learn English for years at the secondary and tertiary levels, university leavers have yet to speak the English language fluently at the workplace (Hiew, 2012). The former Minister of Education of Malaysia, Tan Sri Muhyiddin Yassin, commented in a local newspaper that university graduates have a poor command of English speaking proficiency (Wong, 2014) and stated: "I am baffled about why our children, after completing pre-school, primary school, secondary school and tertiary education, still cannot converse in English" (Subramaniam, 2014).

This has raised the issue of the assessment of speaking proficiency before graduation. To improve speaking proficiency of students, language lecturers and instructors should include speaking assessment in a way that it can lead to improvement

(Cross, 1991). However, there is no specific test for assessing speaking proficiency of university students in Malaysia. Therefore, the development of an ESL speaking proficiency test for undergraduates is necessary.

Although the public English language tests, namely the International English Language Testing System (IELTS), Test of English as a Foreign Language (TOEFL) and the most relevant one to this study, the Malaysian University English Test (MUET), are valid and reliable tests internationally and locally, they provide no specific speaking proficiency descriptors of the ESL language learners. IELTS only offers the overall band of the test taker; it does not mention any description or information about the specific speaking proficiency of the test taker. For instance, Band Six of IELTS means the learner has generally good command of the English language, though with occasional inaccuracies, inappropriateness, and misunderstandings. Likewise, TOEFL iBT only offers scores of the test taker in the description of the scores. TOEFL iBT is based on the scale of 0-120 points and each section of language (listening, reading, writing and speaking) is based on the scale of 0-30. In speaking, if a test taker obtains between 0-9, he/she is considered weak, 10-17 is limited, 18-25 is fair and 26-30 is considered good. Accordingly, the test fails to gauge what the test taker can or cannot do in carrying out conversations. Instead, it only indicates the weak, limited, fair or good level of speaking proficiency. Hence, this has raised the concern for the development of descriptors for speaking proficiency.

To assess Malaysian undergraduates' levels of language proficiency, the Malaysian University English Test (MUET) has been conducted by the Malaysian

Examinations Council since 2003. MUET is aimed at helping stakeholders to assess the overall language level of candidates required (i.e., entrance requirement to public universities) to attain a particular band score among six bands (Malaysian Examinations Council, 2015). However, MUET only provides general and vague descriptions of bands. For example, Band 4 description of MUET indicates that candidates “lack the ability to convey the message accurately” but are at the same time “satisfactorily expressive and fluent . . . with occasional inaccuracies” (Malaysian Examinations Council, 2015, p.10). Other bands also have similar vague or contradictory descriptions. Therefore, the MUET band descriptors are of little help for differentiating between proficiency levels, provides no specific and clear descriptors for speaking proficiency of the language learners. The language lecturers thereby remain insufficient in tailoring their instructional materials to the needs of the language learners.

This drawback is not only peculiar to MUET/the local context, but also to band descriptors in international contexts. An outstanding example is the Common European Framework of Reference for Languages (CEFR), which is also used in non-European countries (Little, 2007). Band descriptors of CEFR have been criticized for ambiguities and inconsistencies about differentiating between proficiency levels (Alderson, 2007; Galaczi, 2013) and suitability for young learners (Hulstijn, 2010; Little, 2007). Similar critique is applicable to traditional assessments that are based on grades or percentages (i.e., only revealing who among students are better than others), which provides no insight or clue on how to improve language proficiency (Burke, 1992).

An obtained overall score or grade is not an indication of a combination of skills that a student has; therefore, “general” test scores are of little help for learners to enhance their speaking proficiency (Oosterhof, 2001). Relying on an overall score, language lecturers are also less likely to recognize students’ strengths and weaknesses in speaking proficiency. Like the other tests, MUET provides language instructors with no specific assessment of what the student can and cannot do in speaking proficiency. Therefore, language instructors remain unclear what the student is lacking with respect to speaking proficiency in order for them to help enhance their students’ speaking proficiency based on their current speaking proficiency. This implies that leading to speaking proficiency improvement based on the current practice of assessment and speaking descriptors have not been relatively comprehensive. To assess students’ speaking proficiency accurately or comprehensively, difficulty levels of test questions, from elementary to advanced need to be addressed.

As to identifying the current speaking proficiency levels of students, identifying speaking proficiency has been conducted by several researchers and language centers. However, there is a dearth of research on identifying ESL undergraduates’ speaking proficiency in Malaysia. Due to the lack of precise identifying, the ESL lecturers imprecisely monitor or determine the speaking proficiency of their students, thereby being unable to carry out necessary actions towards improving their students’ English-speaking proficiency. This calls for identifying the current level of students’ speaking proficiency.

Students at different levels of speaking proficiency have distinct profiles. To profile students with respect to their proficiency levels would facilitate lecturers provide students with instructional guidance and activities they need. Students' profile is crucial to take the necessary action such that lecturers are able to provide further assistance to help their students (Castejón, Gilar, Minano, & González, 2016), especially by profiling the undergraduates' speaking proficiency (Nopiah et al., 2011). However, profiling undergraduates of different speaking proficiency remains unaddressed in Malaysia. This suggests the need for research on profiling students according to their speaking proficiency.

To sum up, based on the above discussion and to the best knowledge of the researcher, determining the specific ESL speaking proficiency of undergraduates in the Malaysian context has remained unaddressed. To fill this gap, it is the prime focus of this study to develop a valid and reliable ESL speaking proficiency test and identify the ESL students' speaking proficiency. A set of descriptors has been developed to identify the undergraduates' specific speaking proficiency performance. The study has also identified the undergraduates' speaking proficiency. According to their proficiency level, the study has profiled them. Student profile would help lecturers determine instructional guidance and activities students need. The study hereby is intended to contribute to the body of literature on speaking proficiency assessment.

#### **1.4 Rationale for the Study**

First, identifying the undergraduates' speaking proficiency will help ESL language instructors to monitor the students' progress as well as to adapt their teaching

materials in the classroom based on the students' needs. Therefore, it is of paramount importance that language instructors should administer speaking tests in the classroom in order for them to suit their teaching materials according to the students' needs.

Another reason behind this study is that even though the language learners have studied English for several years, it seems that language learners are still not capable of conducting appropriate communications and daily conversations in English and still remains a problem, yet to be solved (Radzi, Hanadi, Azmin, Zolhani, & Abdul Latif 2007). Therefore, it is hoped that based on *assessment for learning* of the ESL learners by informing what the learners can do and what they cannot do, this study will help language instructors to assess their students' speaking proficiency regularly and develop their speaking proficiency teaching materials in order for them to help improve their students' speaking proficiency.

Last, identifying and profiling the undergraduates' speaking proficiency will provide the ESL lecturers with enough information to identify those students who are performing well and those who are failing in their speaking proficiency. As such, both the success and the failure will be reported. Therefore, language lecturers will be able to make practical decisions and help those language learners who are weak in speaking proficiency.

### **1.5 Objectives of the Study**

Based on the problem statement regarding providing the ESL lecturers with more information about the speaking proficiency, the lecturers will be able to assist their

students to improve their speaking. Therefore, the specific objectives of this study are:

1. Developing an ESL speaking proficiency test for undergraduates by establishing the following:
  - a. Selection of relevant sources for the construction of the ESL prototype speaking proficiency test (PSPT)
  - b. Determining the best combination of questions for the ESL speaking proficiency test.
  - c. Conducting pilot study to test the validity and reliability of ESL PSPT
  - d. Determining time allocated for the ESL speaking proficiency test
2. Developing a set of descriptors to identify the undergraduates' speaking proficiency performance by establishing:
  - a. Determining the cut scores for the bands
  - b. Identifying speaking skills undergraduates have acquired
3. Identifying the undergraduates' ESL speaking proficiency according to their performance in terms of:
  - a. Respondents' general ESL speaking proficiency
  - b. Gender
  - c. Ethnicity
  - d. University
  - e. School
4. Profiling the undergraduates' ESL speaking proficiency particularly in terms of:
  - a. Superior Performers
  - b. Advanced Performers
  - c. Intermediate Performers
  - d. Novice Performers



## 1.6 Research Questions

This study attempts to answer the following research questions:

1. How is an ESL speaking proficiency test for undergraduates developed?
  - a. What are the relevant sources for the construction of the ESL prototype speaking proficiency test (PSPT)?
  - b. What would be the best possible combination of ESL speaking proficiency questions selected for the ESL speaking proficiency test?
  - c. What is the validity and reliability of the ESL prototype speaking proficiency test?
  - d. What is the time allocated for the ESL speaking proficiency test?
2. How are the speaking proficiency descriptors developed to identify the undergraduates' speaking proficiency performance?
  - a. What are the most suitable cut scores for the performance bands?
  - b. What are the speaking skills undergraduates have acquired?
3. How do the undergraduates perform on the ESL speaking proficiency test?
  - a. What is the respondents' general ESL speaking proficiency?
  - b. What is the overall speaking proficiency according to gender, ethnicity, university, and school?
4. What are the profiles of the undergraduates' ESL speaking proficiency?
  - a. Who are the superior performers?
  - b. Who are the advanced performers?
  - c. Who are the intermediate performers?
  - d. Who are the novice performers?

## **1.7 Significance of the Study**

The primary goal of this current study is to contribute importantly to the existing body of literature on pedagogical assessment of speaking proficiency of undergraduates. The implications of the study will highly contribute in the area of teaching speaking proficiency as the performance of the undergraduates will provide English language lecturers with necessary data and information about the undergraduates' speaking proficiency. Additionally, its results and findings will further our understanding of a better way of assessing speaking proficiency of the undergraduates.

The findings will assist language lecturers to realize the vital role of 'assessment for learning' on enhancing the undergraduates' speaking proficiency as well as their strengths and weaknesses. This research study will help language instructors to recognize the undergraduates' current level of speaking proficiency. As a result, through the identification of the language learners' current level of speaking proficiency, the language instructors will be capable of adapting their instructional materials in the classroom that meet the undergraduates' needs to a great extent.

Likewise, identifying the undergraduates' speaking proficiency will present whether or not the undergraduates are performing well. Unfortunately, the current grades or scores only measure that some students are better than the others, they do not suggest what the students can do and what they cannot do. Thus, identifying provides a complete and comprehensive assessment of the undergraduates at the early stages of instruction to recognize those undergraduates who might not be making sufficient

progress. As a result, this will help the related institution to conduct measures and determine the needs of the undergraduates who are in need of improvement.

Finally, the information collected by the ESL language lecturers could be used by universities and the Ministry of Higher Education to plan what needs to be done to enhance the speaking proficiency of undergraduate students in Malaysia. Further, language lecturers can investigate the effects of any innovative method in the teaching and learning of speaking proficiency at the tertiary level that language lecturers can make. Such that, lecturers will be able to tailor their teaching materials to meet the undergraduate's current needs in pursuit of speaking proficiency improvement.

To conclude, all these will provide precious insights for language planners, curriculum developers, material designers, language instructors in order for them to make informed, instructive and comprehensive decisions with regard to identifying and profiling their students in order to help enhance language learners' level of speaking proficiency. Finally, the data will be gathered by the ESL instructors and teachers will help speaking courses to cater to the different levels of speaking proficiency.

## **1.8 Definition of Terms**

The following definitions are included to illuminate the terminology used in this study and how they are operationalized in this context.

### **Assessment**

Assessment involves the use of practical data on student learning to improve programs and enhance student learning (Allen, 2004). In other words, assessment is the process of gathering and discussing data from diverse sources to advance a deep understanding of what students know with their knowledge as a result of their learning experiences (Huba & Freed, 2000). In this study, however, assessment refers to formative assessment where students are assessed for the purpose of taking initial action based on students' performance. This is conducted to adapt teaching materials to meet the students' needs to promote and maximize learning. Besides, the focus of this study is on assessment for learning which refers to finding out the undergraduates' speaking proficiency.

### **Prototype Speaking Proficiency Test**

Prototype (Prototype, n.d.), as defined by the Merriam-Webster's Dictionary, means "an original model on which something is patterned or an original or first model of something from which other forms are copied or developed." Likewise, a prototype test is a test where revisions can be made to make certain about the test tasks and task type before it is used as a final version of a new test (Cumming, Grant, Mulcahy-Ernt, & Powers, 2004). In this study, however, a prototype speaking proficiency test refers to a test which is developed and tested for its reliability, validity and time allocated for

the test before the test is used as an ESL speaking proficiency test.

### Speaking Descriptors

Descriptor (Descriptor, n.d.), as defined by the Merriam-Webster's Dictionary, means "A word or phrase that serves to describe or identify an item in an information retrieval system." In addition, Davies *et al.* (1999) and Luoma (2004) define descriptors as an instrument consisting of a series of constructed levels along with written descriptions about test-takers' performance. In the context of this study, speaking descriptors consist of characteristics of student performance at each specific band. Besides, speaking descriptors are the detailed description of the specific speaking proficiency of undergraduates at each performance band. The speaking descriptors will be developed based on the respondents' speaking in the ESL speaking proficiency tests.

### Speaking Proficiency

Speaking proficiency refers to the ability of an individual to speak an acquired language and its form and meaning depend on the context in which it takes place (Burns & Joyce, 1997). Likewise, speaking proficiency can be referred to the speaking ability of an interlocutor measured by a particular test using rating scales (Davies et al., 1999). In this study, however, speaking proficiency is defined operationally as the undergraduates' speaking ability in speaking the English language in terms of communicative ability, fluency, and accuracy. It is also the undergraduates' speaking ability to answer fully-controlled, semi-controlled and free controlled communicative activities.

## **1.9 Summary**

This chapter highlights the crucially important impact of conducting a better way of speaking assessment which could probably be enhanced to provide comprehensive and complete information about the ESL undergraduates' speaking proficiency. The chapter presents the background of the study, the statement of the problem, the rationale for the study, objectives of the study, research questions, significance of the study, and finally the operational definition of terms. The next chapter will review the relevant studies conducted on speaking proficiency assessment of ESL students. Related theories and conceptual framework will also be presented.

## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **2.1 Introduction**

This chapter provides an overview of the previously conducted studies and reporting systems on speaking proficiency assessment, concerns of speaking proficiency descriptors of language learners, identifying and profiling the speakers. It begins with a concise elaboration on speaking and speaking proficiency, the importance of speaking proficiency in language teaching and learning as well as the assessment of speaking proficiency and its importance in the process of language learning. Because the main aim of this study is identifying and profiling undergraduates' speaking proficiency, the primary focus will be given to the development of an ESL speaking proficiency test and speaking proficiency descriptors of the undergraduates. Moreover, the chapter provides the background literature on summative assessment and formative assessment. In addition, the chapter presents the related theories and conceptual framework which serve to set the research objectives and research questions of the study within the broader context of existing knowledge. The chapter is concluded with a summary. What follows is the detailed explanation of the above-mentioned concerns.

#### **2.2 Speaking**

Speaking is perhaps the most challenging language skill to teach, learn and assess. According to Luoma (2004), speaking in a second or foreign language is a

difficult task and competence in speaking a new language might take a longer time to improve and develop. Speaking requires involving several capacities and it needs much effort on the part of language learners as it engages preparing to be capable of speaking the language in different real-life situations. Speaking in a new language is always problematic as it requires having linguistic knowledge along with the skills that necessitate speakers to know when and how to use it (Bachman, 1990).

During the last two decades, several scholars and experts in the field of language learning and teaching have attempted to define speaking. The speaking definitions have been presented so far look at speaking from different perspectives and viewpoints. In its simplest form, speaking is an interactive process of constructing communication and meaning which involves producing and receiving information (Burns & Joyce, 1997). Moreover, Florez (1999) and Howarth (2001) defined speaking as a two-way manner involving a true communication of ideas, information, and feelings among individuals. Further,

According to Nunan (2003), speaking is referred as one of the productive skills in language teaching and learning. It is defined as a process of constructing and sharing meaning via the use of language verbally or in oral form. In fact, speaking is basically an oral communication that offers information engaging two interlocutors and they can be referred as speakers who offer the message and listeners who receive the message.

Therefore, it is clear that the communication that takes place between two individuals involves the productive skill of speaking and the receptive skill of listening.



In a same vein, Nunan (1991) mentioned that speaking is a verbal interaction of presenting information, expressing our ideas and thoughts we possess in our mind. Hence, speaking is not only expressing our thoughts but also conveying new information to others.

To elaborate more on the nature of speaking, Burns and Joyce (1997) and Luoma (2004) defined speaking as an interactive manner creating meaning that includes producing, receiving and processing information; and its meaning depends on the context in which it takes place involving speakers, the physical environment and the purpose of speaking. Speaking is also defined as the language learner's capacity to express his/her ideas coherently, fluently and appropriately in given meaningful contexts.

Meanwhile, language scholars and language teaching experts frequently mentioned speaking as a technical term to refer to one of the four skills of language that language learners should learn and improve (Luoma, 2004). In addition, she claims that speaking is seen as a social activity since individuals speak and this constructs a part of the social activity of conversation. In a typical social interaction, two or more people speak to each other about topics of their common and mutual interest in the social interaction activity. They conduct the event together and this makes everyone in the activity a speaker and a listener at the same time (Luoma, 2004).

According to Nunan (2003), speaking consists of constructing systematic oral utterances to deliver meaning. He also mentioned that speaking is immediate and it

happens simultaneously. Most important, it requires that learners not solely need to know how to use language linguistically such as grammar, pronunciation, or vocabulary, they also need to understand when and in what way to use language in context.

Apart from the definitions provided for speaking, numerous definitions have also been presented for speaking proficiency. According to Iwashita, Brown, McNamara, and O'Hagan (2008), proficiency in a second language is one of the most fundamental concepts in language learning and accordingly its character is the subject of continuing and strong debate. Often times, this debate is about competing theories or models of second language proficiency and its development, as seen in the influential discussions by Canale and Swain (1980), and Bachman (1990).

One popular notion of speaking proficiency in a second or foreign language context is the ability to communicate the language and grow communicative competence (Breiner-Sanders, Lowe, Miles & Swender, 2000). Likewise, Burns and Joyce (1997) refer to speaking proficiency as the language learners' capacity to anticipate and produce the expected forms of specific language features. Proficiency is also their ability to manage features of conducting conversations such as turn-taking, providing feedback and closing conversations.

Additionally, as defined by the American Council for the Teaching of Foreign Languages, speaking proficiency is the ability of a language learner or individual to conduct conversations and communicate in appropriate ways (ACTFL, 2006). The

capacity to speak with confidence, accuracy, and clarity in a variety of communicative tasks; and to participate in most informal and some formal conversations on different familiar topics. It is also worth mentioning that the speaking proficiency tests require language learners to handle vocabulary, pronunciation, fluency, sociolinguistic functions and so on.

From the above-mentioned definitions provided by numerous scholars and experts in the field of language teaching about speaking and speaking proficiency, it can be concluded that speaking is a form to talk and express an idea or opinion in order to convey a message to make another person understand. Accordingly, speaking involves two or more people as speakers and listeners and speaking proficiency can refer to the ability of language learners to produce the target language fluently and appropriately in real life situations.

### **2.3 Importance of Speaking**

Mastering the art of speaking to many language learners is seen as the most important language aspect when learning a second or foreign language and success is judged based on how well the language learner can conduct conversations in the language he/she is learning (Nunan, 1991). Of all the four language skills (listening, speaking, writing and reading), speaking is automatically considered as the most important and essential skill for language learners to learn and improve (Ur, 1996). She also states that when people claim that they know a language, they undoubtedly refer to as speakers of that specific language, as if speaking comprises all the skills of a language. This indicates that many second or foreign language learners are mainly

interested in learning to speak the language fluently they are studying.

The desire to communicate with others, mainly face to face, encourages many language learners to try their best to learn and enhance their speaking proficiency, this makes speaking to be seen as the central language skill (Bailey & Savage, 1994). Therefore, speaking is a skill which deserves close attention as much as literacy skills in language learning.

Further to that, numerous language learners give priority to speaking proficiency as they believe if they master the speaking proficiency, they are considered as speakers of the language. Most important, the most frequent question that most language learners are asked about is “do you speak English?” not “do you write English?”. Therefore, the majority of language learners try to focus on learning and improving the speaking proficiency as they think this will help them in finding jobs in their future professions. It is in this regard that Baker and Westrup (2003) argue that learners who can speak English fluently, might have greater opportunities for employment. More important, speaking proficiency is crucial for both fluency and accuracy in expressing communicative intent (Rashid, Mohamed, Rahman, & Wan Shamsuddin, 2017).

Moreover, speaking is at the heart of second or foreign language learning. However, it has been neglected in teaching and testing (Egan, 1999). Egan also claims that despite its significance, speaking has not been given close attention in schools and universities due to massive focus on grammar, vocabulary. Clifford, 1987 (as cited in Egan, 1999) argued that speaking proficiency has also been absent from testing because

of the difficulty in assessing it objectively and the time it requires to conduct speaking tests. Likewise, in spite the importance of speaking for language learners to communicate, much attention was given to other elements of language such as reading and writing skills and speaking proficiency has been neglected enormously and the Grammar-Translation method is an example of this claim (Richards & Rodgers, 2001). It is generally true that the capacity to speak and communicate in a second language clearly contributes to the success of the language learner in university as well as later in every phase of life.

Malaysia is aiming to become a fully developed nation by the year 2020. Therefore, the English language is reasonably important for cross-cultural communication involving international participation. English, as the world lingua franca, is a means for social mobility and Malaysian graduates should have an awareness of its importance and develop a positive language attitude to master the language. It can be concluded that speaking is vital in learning and teaching a second or foreign language and thus language learners judge their success based on how well they have mastered the speaking proficiency of the language they are studying.

#### **2.4 Assessment**

In its simplest definition, assessment is defined as “the systematic collection of information about student learning, using the time, knowledge, expertise and resources available in order to inform the decision on how to improve learning” (Walvoord, 2004, P. 2-3). Assessment is crucially important in education, and it comes in different forms such as task-based performance, formative assessments (assessment for learning),

summative assessments (assessment of learning). In general, students do not learn what they are taught. It is this simple and thoughtful reality that means that assessment is possibly the fundamental process of effective instruction. If our students learned everything we taught, we would never want to assess (William, 2013). Therefore, it is vital to conduct ongoing assessments to make sure student progress.

In general, educational assessment has been either ‘summative’ assessment of learning or ‘formative’ assessment for learning. Formative assessment has been commonly used to assess and monitor students’ progress with instruction and as such, provides comprehensive feedback to the institution pertaining to the process of learning, as such, students’ prior knowledge is crucially important (Ambrose, Bridges, Dipietro, Lovett, & Norman, 2010). In contrast, summative assessment has been normally used to assess students’ progress at the end of a program or a semester to understand and determine what the students have achieved throughout the course of study (Bresciani, Gardner, & Hickmott, 2009).

Generally speaking, grades or scores are used to assess learners’ achievements, but a grade or a score provides little feedback on what students need and are lacking. Scores do not show the skills a student has. Instead, they merely show an overall indication of student achievement (Oosterhof, 2001). In contrast, Masters (2015) states that constructive and clear feedback is essential to report what students can do and what they cannot do, so their weaknesses can be considered for betterment and future planning. Therefore, it is central to inform about where the students are in which it helps in recognizing starting points for action to take. Moreover, when the assessment

is intended to help and guide future planning, the crucial goal is to boost learning (Masters, 2015).

If learning is the main goal, assessment should be more than just assigning grades or scores. As Kubiszyn and Borich (2010) stated, educational assessments can be either formative or summative, relying on whether they are applied to assess day-to-day improvement in learning (e.g., formative) or over a period of time, usually at the end of a term (e.g., summative). They also believe that assessment results can then be used to make decisions about students' progress, instruction, and teaching materials.

Therefore, effective formative assessment (assessment for learning) is at the focus of effective teaching, the teaching that benefits learners progress towards better performance. Formative assessment aids instructors and students define what is necessary to build toward this attainment. On the contrary, summative assessment can tell whether or not a student has met a standard, or is able in a performance that encompasses learning (Afflerbach, Cho, Kim, Crassas, & Doyle, 2013). As such, it is crucially essential if the assessment is aimed at gathering information that can be used to alter teaching and learning. Hence, this research will employ 'assessment for learning' to identify the undergraduates' speaking proficiency to decide where the learners are in their learning speaking, where they need to be and how best to get there.

#### **2.4.1 Speaking Assessment**

Generally speaking, assessing speaking proficiency is always considered to be the most difficult and challenging of all language elements to administer and score

because it is difficult to determine what and how to assess speaking (Madsen, 1983). He also argued that speaking requires the use of different abilities. Therefore, there are some criteria which should be considered in assessing the language learner's speaking proficiency (e.g., grammar, vocabulary, pronunciation, fluency, appropriateness, and communicative ability). Similarly, according to Luoma (2004), among all the skills of language, it has been largely recognized that speaking assessment is the most difficult language element to conduct. Further, a speaking test is time-taking and it takes much time to assess each student's speaking proficiency (Thornbury, 2005). Similarly, Morovat (2014) mentioned that among the four skills of language assessed in proficiency exams, speaking is said to be one of the most difficult and challenging ones. Another difficulty in speaking assessment lies in the fact that it is not an easy task to determine what speaking aspects should be assessed and what criteria should be used to test a language learner's speaking proficiency (Thornbury, 2005).

It is important to note that, there are two main methods of scoring scales when it comes to assessing speaking which they are holistic rating scale and analytical rating scale. According to Luoma (2004), a holistic scale is an overall impression of an examiner towards an examinee's speaking proficiency whereas in analytical rating scale, examiners are provided with a number of criteria, usually (3-5) with descriptors at different levels of the scale examine different aspects of speaking e.g., fluency, grammar, and communicative ability. Likewise, Thornbury (2005) stated that there are two major ways during the speaking assessment. The first scoring method is known as holistic scoring where it gives a single score on the basis of an overall impression. Holistic rating score is adventurous in the sense that it is quicker and it is perhaps



sufficient for informal testing progress. The second scoring method is called analytical rating scale. It gives a separate score for different aspects of the speaking task. Although analytical rating scoring requires more time, it compels students to take a variety of speaking categories into their consideration. It is worth mentioning that if the specific speaking categories are well-chosen, analytical rating score is undoubtedly fairer and more reliable (Thornbury, 2005).

However, this study considers the analytical rating scale to be used as the scoring method during marking the undergraduates' speaking proficiency. This scoring method is used because it shows the strengths and weaknesses of language learners in detail which assist both learners and lecturers to take the learner's lacking speaking skills into consideration for later improvement (Underhill, 1987). Underhill also feels that the use of analytical rating scale helps spoken tests easier and more consistent for the assessor mentally during grading students' speaking proficiency. Moreover, the use of analytical rubric is substantial to influence the assessment test, thereby providing feedback for improvement (Liao & Hsu, 2014). As such, this leads to increasing the awareness of adjusting teaching materials suited for the current level of the students based on the analytical rating scoring (Fulcher, 2012).

#### **2.4.2 Summative Assessment**

In a general sense, summative assessments take place in the end of a unit, a term or a year in the learning process and help to inform about and document an individual's learning in a particular subject (Sadler, 1989, cited in Costel et al., 2015). Further, summative assessment is known as 'assessment of learning' and is distinguished with

formative assessment, which is ‘assessment for learning’. In addition, any assessment that happens at the end of learning or instruction with the purpose of overall achievement and knowledge can be known as summative assessment (Leahy, Lyon, Thompson, & William, 2005).

Summative assessment occurs after instruction and the results of any summative assessment provide too late information for instructors to take any action for the benefit of the current class of learners (Leahy et al., 2005). This kind of testing is usually conducted with the only purpose of grading or evaluation progress of the learning process (Earl, 2003).

Another way to look at summative assessment is the end result of the teaching and learning process or quality control, whereas formative assessment identifies the weaknesses along the way and modifies instruction and teaching to report those weaknesses in order to provide feedback for the sole purpose of improvement (Leahy et al., 2005). In general, assessment is done traditionally which is assessing after a period of time but recently, there have been changes in emphasis from summative to formative assessment, from making judgmental only to that of establishing descriptions of what students can do and what they cannot do which can be useful to address students’ strengths and weaknesses in further development of learning (Earl, 2003). The next Section Addresses how formative assessment guarantees this quality in the learning and teaching process.

### **2.4.3 Formative Assessment**

Formative assessment is a continuous assessment which is usually in line with the actual learning (Yin et al., 2008). In formative assessment, educators and lecturers use the information they obtain from their students' assessment to alter their teaching and instruction in order to adapt their teaching to meet the students' needs to promote and maximize learning (Wang, Wang, & Wang, 2006). Likewise, the ultimate goal of formative assessment is to help students to learn and teachers to teach (William, 2011) which is the most appropriate aim of teaching and learning in line with modern learning theories that recognize students' central role in their learning (Penuel & Shepard, 2016).

In addition, Bell and Cowie (2001) defined formative assessment as 'the process used by teachers and students to recognize and respond to student learning in order to enhance that learning, during the learning' (p. 537). Further, continuous assessment assists lecturers and teachers to provide immediate feedback which is useful to promote student learning (Sadler, 1989, cited in Jacoby, Heugh, Bax, & Branford-White, 2014). Research has shown that students' language learning process can be improved by using appropriate formative assessments (Bransford, Brown, & Cocking 1999; Henly 2003; Velan, Kumar, Dziegielewski, & Wakefield, 2002; Wang et al., 2006). Formative assessment has been acknowledged by researchers as a central element in conducting learning activities for promoting student learning success (Bell & Cowie, 2001). Most important, it has been seen that formative assessment has been used as an excellent means of improving student performance and to ensure the maximum benefits for students, especially the achievement of lower-performing students (Dunn & Mulvenon, 2009).

Briton (2011) claimed that not all assessments need to be conducted in the final stage of learning, they can be helpful and contributive if they are taken to determine students' improvement. According to Boston (2002), when teachers are aware of how their students are progressing and where they have trouble in their learning, they can use the information they achieve from on-going assessments to make necessary adjustments, alternative teaching approaches or altering teaching materials to assist students to improve their learning. As such, the learners will be able to realize where they are in their learning, what they can do, what they cannot do and what they have to do to get where they are supposed to be.

It is worth mentioning that when assessments are taken for the purpose of diagnosing where the students are in their learning, it is known as assessment for learning. Assessment for learning is about informing learners of their improvement to enable them to take the necessary action to progress their performance (Pead, 2008). As such, assessment for learning is about supporting classroom learning effectively. Thus, it connects assessment and learning outcomes. Further, assessment for learning involves frequent and on-going assessments of student improvement to find out learning needs and adapting teaching materials to help the learners reach the target level of learning (Clark, 2008; Penuel & Shepard, 2016).

Likewise, Earl (2003) claimed that assessment for learning helps educators and teachers to alter their instructional materials because it shifts assessment from summative to formative for the purpose of benefiting individuals in their learning from making judgments to establishing descriptions that can be utilized to serve the learning

process. Therefore, for the context of this study, assessment for learning will be proposed to assess the undergraduates' speaking proficiency in order to identify where the undergraduates are, what trouble they have and what they need to do to get where they need to be.

#### **2.4.4 Validation of Test Instrument**

In order to ensure that the assessment is fair and just, certain underlying procedures of assessment with regard to validity and reliability should be taken into account in advance. Validity refers to the appropriateness, meaningfulness, and usefulness of a test and inferences a researcher makes (Fraenkel & Wallen, 2009). In addition, validity is defined as the extent to which a measuring instrument or an assessment measures what purports to measure. Moreover, validity is the most significant notion to take into consideration when selecting an instrument to be used. Therefore, researchers more than anything else want the inferences they make via the use of an instrument to serve their intended purpose (Fraenkel & Wallen, 2009). As such, an assessment which is not valid is considered to be inappropriate and meaningless.

Validity is often divided into three interconnected categories namely; construct-related evidence of validity, criterion-related evidence of validity and content-related evidence of validity. First, construct-related validity constructs whether or not the student performance to be observed shows an authentic indicator of the capacity the teacher or assessor hopes to assess (Oosterhof, 2001). Any information that lets the assessor know if the results correspond to what they expect based on their knowledge

about what is being measured, resembles the construct validity of a test or instrument (Kubiszyn & Borich, 2000). Further, construct validity is the most essential form of validity as it brings the fundamental issue if a test measures the real construct. Second, criterion-related validity is constructed when scores from a new test and an established test are compared (Kubiszyn & Borich, 2000). In determining criterion-related validity, an important question has to be considered-that is how well performance on the new test matches performance on an existed well-established test. Therefore, in establishing criterion-related evidence of validity, both the new test and an external criterion (existed established test) are administered to a group of respondents, then finding the correlation between the scores from both sets of tests. If the same group of students score similarly on both tests, shown by a high correlation, then the new test is said to have criterion-related evidence of validity. Finally, content-related validity which is established by comparing the test items with the objectives to determine if the test items measure or match the objectives of the study. By doing such an examination, it can be said that whether or not a test has content validity. Content-related validity usually involves collecting the judgments of experts who can judge the test items. As such, no correlation coefficient is computed. Instead, only human judgment is involved (Kubiszyn & Borich, 2000).

Different procedures have been taken to determine the validity of speaking tests. For example, Sak (2008) carried out a research on investigating the validity and reliability of an English-speaking exam at a Turkish university. In her study, she used face validity, content validity as well as construct validity to validate the speaking exam. She found that the exam had satisfactory face validity and the quality of content

validity. She also concluded that the construct validity of the test was relatively low.

In another study, Zhao (2013) conducted a study on developing a procedural framework for the development and validation of diagnostic speaking tests. He claimed that test validation is a crucial phase in developing any test since it assists the researcher to enhance the test and the quality of test items, as well as to investigate the workability of the test. The framework is composed of four major phases: needs analysis, test design, test piloting and administration and validation.

It is worth mentioning that criterion-related validity was not considered to be conducted in this current study due to having no any external criterion (existed established test) to compare the speaking proficiency performance of the undergraduates on both the new test and the existed established test. Therefore, it is clear that content validity and construct validity were conducted to validate the test instrument of this study.

In line with the validity matters, reliability is the most important issue to consider when preparing an instrument for use (Fraenkel & Wallen, 2009). There are a number of ways to estimate the reliability of a test or any measuring research instrument, but the three basic and fundamental methods most frequently used are called *test-retest (stability)*, *Parallel-Forms Reliability (alternative forms)* and *internal consistency*. As the name implies, test-retest is a method of estimating reliability that is given twice and the correlation between the two sets of scores are considered. Likewise, Parallel-Forms Reliability is a method that requires administering two or more forms of the same

assessment to the same group of students under the condition as nearly as possible (Kubiszyn & Borich, 2010; Linn & Miller, 2005). Parallel-Forms Reliability uses two or more different but equivalent forms of an assessment. Linn and Miller (2005) have mentioned that equivalent forms (parallel forms) are constructed to the same set of specifications (e.g., test content and difficulty) but are built independently. They have also pointed out that the forms of an assessment are administered to the same group of individuals in close sequence and the scores are then correlated. This correlation provides a measure of the degree to which generalizations about student performance from one assessment to another is justified.

As such, the correlation tells the degree to which the forms of an assessment are measuring the same aspect of behavior. The most appropriate way to judge whether an assessment measures a domain, speaking, for example, is to construct more than a version of the assessment that is said to cover the same student skill and correlate the results (Linn & Miller, 2005).

Whereas, internal consistency requires one single administration of the research instrument. Internal consistency is a most widely-used method of checking reliability (Gay & Airasian, 2003). The internal consistency methods of estimating reliability include split-half, Kuder-Richardson and Cronbach's Alpha coefficient of reliability and they provide information about the consistency among the items in a single test (Fraenkel, Wallen, & Hyun, 2012). More important, internal consistency shows to what extent the items of a test assess the same skill or quality and testers generally prefer internal consistency to evaluate the reliability of a test (Brown, 2005). A coefficient of



0.00 tells the complete absence of a relationship, therefore no reliability, whereas a coefficient of 1.00 shows the complete relationship among the items of a single test.

Previous studies have utilized different methods to determine the reliability of speaking tests. For example, in a study, Sak (2008) investigated the interrater reliability of a speaking exam at a Turkish university. In her study, she only utilized interrater reliability to judge the reliability of the speaking exam. She found that the interrater reliability of the exam was not satisfactory. Likewise, Halleck (1996) investigated the interrater reliability of proficiency level judgments of graduate student trainee raters on oral proficiency interviews. It was found that interviews of the higher level of interviewees resulted in higher level of interrater reliability. In fact, determining the reliability of a speaking test through solely interrater reliability is not enough. Therefore, to make sure that the test will be used in this study is reliable, three basic principles of reliability were taken into consideration which are parallel-forms reliability, Cronbach's Alpha, and inter-rater reliability.

In addition to student assessment, language instructors and lecturers should be equipped with descriptors as the basis of their student assessment to determine if the learning outcomes are according to the standards specified. Therefore, the next section will address the issue of the speaking proficiency performance descriptors.

## **2.5 Speaking Proficiency Descriptors**

Descriptors are generally defined in terms of what a student is expected to know, understand and be able to do as a result of learning (New Zealand Qualifications

Authority, 2013). In the context of monitoring, assessment and this study, a descriptor is a metric which offers information to monitor and assess performance and achievement. Simply put, descriptors are standardized measures that permit for comparisons with the passage of time in pursuit of improvement.

One of the fundamental issues in identifying the undergraduates' speaking proficiency concerns the development of the speaking proficiency descriptors. In this study, the researcher will develop the speaking proficiency descriptors to show the students' speaking proficiency based on the performance bands. The cut scores for the performance bands were calculated based on z-scores as discussed in chapter five.

The ESL lecturers can use the speaking proficiency descriptors as the basis for their assessment whether the learning outcomes have been met at each expected level. As such, the lecturers can refer to the speaking proficiency descriptors to recognize the undergraduates' speaking proficiency.

With regards to language proficiency in general and speaking proficiency specifically, different proficiency descriptors have been developed by different researchers and institutions so far such as MUET speaking descriptors, IELTS speaking band descriptors, ACTFL proficiency guidelines-speaking (Breiner-Sanders, Lowe, Miles, & Swender, 2000), Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) (Council of Europe, 2001), just to name a few. However, these descriptors are not specific speaking proficiency e.g., MUET speaking criteria are too general in terms of what students can do and what they cannot

do. Therefore, in this study, the speaking proficiency descriptors will be developed based on the ability of the students to perform linguistic competence and sociolinguistic competence based on the results of the ESL speaking proficiency test.

In addition, Rafieyan (2014) developed a set of descriptors for undergraduates' pragmatic comprehension. In his study, he utilized z-scores to determine four categories of performers based on cut scores. However, the study did not determine the performance bands so the descriptors could be developed in accordance with the performance bands. In contrast, the current study develops the speaking proficiency descriptors based on the performance bands for the undergraduates.

In the context of this study, speaking proficiency descriptors are developed to state what students are expected to do, what they can do and what they cannot do at each specific level so that the ESL lecturers will be able to help their students improve their English-speaking proficiency with the passage of time. Therefore, the speaking proficiency descriptors were developed to find out the undergraduates' speaking proficiency based on the results of the ESL speaking proficiency test.

By using descriptors, the ESL instructors will be able to identify the students' weaknesses and needs in speaking proficiency based on the assessments they carry out. As such, this helps the ESL instructors to provide intervention strategies for the students who are not performing well. Next, the issue of identifying the ESL undergraduates' speaking proficiency will be discussed.

## **2.6 Identifying Respondents' Speaking Proficiency**

Language assessment is a valuable tool for providing adequate information regarding teaching. Given the important role of assessment, there are crucial issues to be investigated, in particular in pursuit of improvement. For the last two decades, significant changes have been identified in the body of literature on how assessment has been changed in higher education (White, 2009). For example, assessment for learning as an approach for providing feedback for improvement has been given due attention (i.e., giving constructive feedback rather than just awarding grades). This change is to consider assessment as a major portion of the teaching process rather than an activity in the end of a teaching term (White 2009).

Over the last decades, different factors have been identified in the literature in playing their role in students' speaking proficiency. For example, gender has been connected with the issue of language learning. In a study by Hunter, Gambell, and Randhawa (2005), gap differences between female and male students' performance in their speaking proficiency was studied. The study concluded that female students performed better than male students in their speaking proficiency performance.

In another study by Gorjian, Moosavian and Shahramiri (2011), the effect of gender on the speaking skill was investigated. They studied the effect of oral summary of short stories on male and female learners' speaking proficiency. The results of their study showed no significant difference between male and female students, although females slightly outperformed males. Further, Koosha, Ketabi, and Kassaian (2011) studied gender differences in their speaking proficiency. The study did not find any

statistically significant differences between male and female students in their speaking proficiency performance. It has to be mentioned that at this point few studies have been conducted to show the effect of gender on speaking proficiency of language learners. Therefore, evidence for gender differences in speaking proficiency performance is not convincing. As such, it is of paramount significance to explore gender in relation to speaking proficiency.

In the same vein, in the literature ethnicity has been considered in relation with language learning, especially speaking proficiency performance. A number of studies have been conducted investigating ethnicity differences in relation to speaking proficiency performance. For example, in a study by Mahyuddin et al. (2006), gender differences between different races of Malay, Chinese and Indian students were studied. The study concluded that Indian students performed better than Chinese and Malay students. Likewise, Renganathan and Chong (2007) studied performance differences of Indian, Chinese as well as Malay students in relation to their speaking proficiency performance. The study showed that Indian students outperformed Chinese and Malay students. Previous research has suggested that the Indian students are able to speak English more fluently compared to the other races in Malaysia, Especially Chinese and Malay students. This is due to the fact that Indians perceive English to have a higher status than their own mother language. Renganathan and Chong (2007) also stated that Chinese students are more likely to speak English more frequently in their daily life. This leads them to be better English speakers, and they are more confident in using English language.

In addition, in a project, Mewald, Gassner and Sigott (2009), set the specifications of a speaking assessment test, E8 Speaking Test Specifications, in Australia. The aim of the E8 speaking test is to recognize the language learners' strengths and weaknesses in speaking skill and communicating meaningfully that can present as closely as feasible to the real-life language environment. The outcome of this speaking test is also important to the language learners, their teachers as well as principals. This aids students to screen their improvement and identify their strengths and weaknesses. Moreover, since the aim of the assessment is analysis, the most important competencies needed for speaking naturally and meaningfully are to be recognized for assessment purposes. As such, language naturalness of speech, communicative competence, vocabulary, and grammar are assessed in pursuit of the students' speaking proficiency improvement.

Hence, identifying students' speaking proficiency is used in this study to identify the undergraduates' speaking proficiency in pursuit of improvement as a more in-depth and detailed analysis of weaknesses and needs of the undergraduates. Therefore, it can be stated that identifying undergraduates' speaking proficiency will be able to equip the language instructors and lecturers with detailed information regarding the undergraduates' speaking proficiency.

## **2.7 Profiling Undergraduate Students' Speaking Proficiency**

The last step after identifying was profiling the undergraduate students' speaking proficiency. This is crucially essential for language instructors to categorize who the students are and what they are able to do and not able to do. In fact, recognizing and

classification of the students is crucially vital for future actions. Instructors need to be aware that students may not be able to improve for a variety of reasons. That is, it would be inappropriate to group or categorize students under one similar category. That is, categorizing and profiling students' performance is significant to recognize students' strengths and weaknesses (Nopiah et al., 2011). Identifying the profiles of students would enable instructors to place the students into relevant categories. As such, categorizing undergraduate students into superior performers, advanced performers, intermediate performers and novice performers would be able to help the ESL language instructors with detailed information with regard to the undergraduates' speaking proficiency.

To determine the strengths and weaknesses of students in speaking, students' speaking assessment has been a recurring issue since appropriate assessment guarantees student's betterment in speaking proficiency (Luoma, 2004). The level of students' speaking proficiency in previous studies has been utilized to categorize students into relevant categories in pursuit of improvement. For example, in an attempt, The American Council for the Teaching of Foreign Languages (ACTFL, 1999) was devised. The level descriptors contain ten levels whereby it mentions the speaking tasks the learners can cope with and the language activities they can conduct (Luoma, 2004). The ACTFL Proficiency Guidelines-Speaking Levels are presented in Table 2.1 below.

Table 2.1

*The ACTFL Proficiency Guidelines-Speaking*

| Superior                    | Advanced                    | Intermediate                    | Novice                    |
|-----------------------------|-----------------------------|---------------------------------|---------------------------|
| Superior-level speakers are | Advanced-level speakers are | Intermediate-level speakers are | Novice-level speakers are |

|   |  |  |  |
|---|--|--|--|
| <p>characterized by the ability to:</p> <ul style="list-style-type: none"> <li>• participate fully and effectively in conversations in the formal and informal settings on topics related to practical needs and areas of professional and/or scholarly interests</li> <li>• provide a structured argument to explain and defend opinions and develop effective hypotheses within extended discourse</li> <li>• discuss topics concretely and abstractly</li> <li>• deal with a linguistically unfamiliar situation</li> <li>• maintain a high degree of linguistic accuracy</li> <li>• satisfy the linguistic demands of professional and/or scholarly life</li> </ul> | <p>characterized by the ability to:</p> <ul style="list-style-type: none"> <li>• participate actively in conversations in most informal and some formal settings on topics of personal and public interest</li> <li>• narrate and describe in major time frames with good control of aspect</li> <li>• deal effectively with unanticipated complications through a variety of communicative devices</li> <li>• sustain communication by using, with suitable accuracy and confidence, the connected discourse of paragraph length and substance</li> <li>• satisfy the demands of work and/or school situations</li> </ul> | <p>characterized by the ability to:</p> <ul style="list-style-type: none"> <li>• participate in simple, direct conversations on generally predictable topics related to daily activities and personal environment</li> <li>• create with the language and communicate personal meaning to sympathetic interlocutors by combining language elements in discrete sentences and strings of sentences</li> <li>• obtain and give information by asking and answering questions</li> <li>• sustain and bring to a close a number of basic, uncomplicated communicative exchanges, often in a reactive mode</li> <li>• satisfy simple personal needs and social demands to survive in the target language culture</li> </ul> | <p>characterized by the ability to:</p> <ul style="list-style-type: none"> <li>• respond to simple questions on the most common features of daily life</li> <li>• convey minimal meaning to interlocutors experienced with dealing with foreigners by using isolated words, lists of words, memorized phrases and some personalized recombinations of words and phrases</li> <li>• satisfy a very limited number of immediate needs</li> </ul> |
|---|--|--|--|

In another similar vein, The Common European Framework (CEF) is designed as a reference for curriculum guidelines, language syllabuses, improvement as well as assessment. The framework explains in a comprehensive way what language learners need to learn in order to use a language for communication (Council of Europe, 2001). Moreover, the framework discusses scales and levels of speaking to describe students'



level as shown in Table 2.2.

Table 2.2

*Common Reference Levels: Speaking*

|                  |    |   |
|------------------|----|---|
| Proficiency User | C2 | Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of Proficient meaning even in more complex situations.   |
|                  | C1 | Can understand a wide range of demanding, longer texts and recognize implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing the controlled use of organizational patterns, connectors, and cohesive devices.   |
| Independent User | B2 | Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialization. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and Independent disadvantages of various options. |
|                  | B1 | Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, and leisure. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes, and ambitions and briefly give reasons and explanations for opinions and plans.                               |
| Basic User       | A2 | Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment, and matters in areas of immediate need.                                     |
|                  | A1 | Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.   |

Source: (CEF, 2001)

Moreover, in a study, Littlewood and Liu (1996) carried out a study on profiling students entering universities. The study was a two-year study profiling student who were entering universities for their tertiary education so that enough information about students can be collected for future instruction. Likewise, Evans and Green (2007)

conducted a study on profiling tertiary students to identify their current needs of English language so that future intervention can be made. In this study, profiling students is established so that groups of students can be recognized for the purpose of future planning in pursuit of improvement.

## **2.8 Theories Related to Speaking Proficiency**

In general, this study has been based on two theories: Littlewood's (1981) Methodological Framework and Long's (1981) Interaction Hypothesis. Littlewood's Methodological Framework focuses on the distribution of speaking activities into different levels while Interaction Hypothesis deals with the necessity of engaging interaction among language learners. That is, language instructors and lecturers can start teaching from pre-communicative activities to free communicative activities in the classroom. Since there is a strong relationship between teaching and assessment, the two theories have been used to conduct the current study (Jaya, 2003).

### **2.8.1 Littlewood's Methodological Framework**

According to Littlewood's (1981) Methodological Framework, teachers can teach communicative activities at different levels of speaking and communication. Littlewood (1981) proposed a methodological framework for teaching language communicatively whereby language instructors can start teaching from pre-communicative activities to free communicative activities. Based on the framework, pre-communicative activities are activities whereby language learners merely practise the new language structure with their lecturers. As such, practising the overall speaking proficiency is not learned at this point.

Pre-communicative activities include the majority of the tasks and activities appear in teaching materials and formal environment of teaching such as question-and-answer, form and function practice. Basically, these types of communication activities can only help language learners to practise the linguistic system rather than requiring them to conduct communicative activities (Littlewood, 1981). Therefore, the purpose of these activities is solely to produce language which is acceptable and appropriate and they help prepare language learners for a higher level of language learning at a later stage.

Whereas in free communicative activities, learners must engage in communicative activities where they have to involve in communications and free discussion. They are now totally engaged in practising the total skill of speaking (Littlewood, 1981). The methodological framework of Littlewood has been adapted and represented as follows:

Pre-communicative activities  Structural and functional activities

Communicative activities  Social interaction activities

This methodological framework is also supported by Brown (2001) where he claims that any lesson should contain a series of activities that allow learners to practise the language they are learning. Therefore, it is important to follow the order of the activities, considering when and why they will be practised to promise the class progresses efficiently from easy to difficult activities (Brown, 2001).

As such, language activities must begin with fully-controlled activities where the lecturer knows the answer and there might be one possible response. As familiarity with the new language structure grows, partially-controlled activities can be the alternative to practise whereby there is, to some degree, an increased amount of freedom for the learner. Following partially-controlled activities, free communicative activities should involve at the last stage of the learning process. At this stage, learners have complete freedom in the language they produce. Therefore, their lecturer will not be capable of predicting what will be uttered in the activity (Gavilán, 2008). The methodological framework proposed by Littlewood (1981) is useful for speaking lecturers who are interested in the order of teaching and testing of speaking proficiency. Accordingly, Littlewood's Methodological Framework was taken into consideration in constructing the conceptual framework of this study.

### **2.8.2 Interaction Hypothesis**

One fundamental aspect in developing the ESL prototype speaking proficiency test which cannot be neglected is Long's (1981) Interaction Hypothesis. The Interaction Hypothesis is a theory of second-language acquisition which shapes that the development of language proficiency is encouraged by face-to-face interaction and communication. For the role of interaction in second language learning, Long (1981) introduced the Interaction Hypothesis. Long (1983) shows that though the comprehensible input is necessary, it alone cannot ensure acquisition. Ellis (1995, P.) defines interaction as "the participants of equal status that share similar need, make an effort to understand each other. If role relationship is asymmetrical, meaning negotiation is inhibited".

Through the interactional discourse, the second language learners will reach a better level of comprehension when they interact with their interlocutors. Moreover, through the negotiation of meaning in the interaction, the learners correct each other and are more directed towards the target language (Long, 1983).

The Long's (1981) Interaction Hypothesis as one of the elements of the PSPT was taken into consideration during the development of the test. Because the last section of the PSPT is a group discussion of four individuals, Interaction Hypothesis should be considered. In this section of the test, the respondents engage in a group discussion in a way that each respondent plays a role where he/she has to present his/her own ideas about the topic of discussion and then they continue discussing their points of view on the given topic.

Further, According to Long's Interaction Hypothesis (1996), students' speaking proficiency is enhanced through interaction where students can engage in authentic conversations with their peers. According to Long (1981), the interaction between language learners creates acute language learning and improvement where learners learn via negotiation of meaning. Furthermore, research has proved that interaction among language learners is effective in helping language learning (Ellis, 1999; Wang, & Castro, 2010).

More important, speaking tasks involve two or more than two people using language for interactional purposes because much of our daily communication remains interactional (Shumin, 2002, P. 208). Moreover, it is generally believed that the

Interaction Hypothesis in language learning facilitates learning through interaction or active use of the language (Miller, 1998). Besides, it is necessary to note that there is a general consensus that assessment should be aligned with what it is taught. In this study, therefore, group discussion as an interactional activity was considered in the assessment developed.

## **2.9 English Language Syllabus Offered by the School of Languages, Literacies, and Translation**

The syllabus of the English language courses offered at the School of Languages, Literacies, and Translation is a document in which the language contents are designed according to the plans which are taken from familiar context such as home, business, discussion, the community and so on. These contents provide the context through which the language skills are to be taught and learned in the appropriate manner. According to Jaya (2003), there is a strong relationship between the syllabus and assessment and types of questions given in the assessment. Therefore, it is of vital importance that the syllabus is analyzed when an assessment is developed. Hence, certain aspects appeared in the syllabus would be taken into account as the objectives of the assessment.

The teaching of English at the School of Languages, Literacies, and Translation is basically to help all USM-leavers to use the English language in everyday life conversations and conducting conversations at the workplace after graduation. In order to achieve this goal, the syllabus must provide related content and teaching materials which help the undergraduates communicate with one another.

The English language syllabus of the aforementioned school is basically to help USM undergraduates to improve all the four language skills of speaking, reading, writing and listening. However, this study focuses only on the speaking proficiency of the undergraduates. In learning a new language, speaking proficiency is said to be the most important language element of all the four language skills for the learners (Ur, 1996). Therefore, it is of central concern to ensure that the learners are able to improve their speaking proficiency in order for them to be able to converse well in English after graduation.

In the context of this research, students from four English language courses offered by the School of Languages, Literacies and Translation namely; LMT 100: Preparatory English, LSP 300: Academic English, LSP 403: Business and Communication English and LHP 456: Spoken English are taken into account. In each of these courses, students are taught different speaking activities such as ask and answer questions, discussion groups, topic discussion, role play, debates, and job interviews. These activities were analyzed and then utilized for the purpose of developing the ESL prototype speaking proficiency test.

## **2.10 Kathleen Bardovi-Harlig's Operationalizing Conversation Speech Acts**

Another aspect of constructing the PSPT is the speech acts used in conversations by Bardovi-Harlig (2015). The researcher of the current study adapted 10 speech acts utilized in conversation in Section A. As such, the researcher constructed 10 easy questions whereby the respondents need to write down appropriate answers based on the given scenarios, these questions would construct Section A of the PSPT. Spoken

language can sometimes be gained through written form of language. Written questions can be a means for assessing spoken language in the form of a written production measure and they are called “written-for-oral” tasks (Cohen & Shively, 2007, p.196). Written-for-oral tasks require students to produce language in writing what they would say in conversation. Written-for-oral tasks can be scenarios to which students respond in writing (Eslami & Liu, 2013, p. 71).

### **2.11 Public Exams: IELTS and MUET Past Year Examinations**

One final aspect in developing the ESL prototype speaking proficiency test which cannot be neglected is the public examinations of IELTS and MUET. In constructing the PSPT, ideas from part B of IELTS and Task B of MUET examination were taken into account. An analysis was conducted to identify the related information that could be taken from part B of IELTS and Task B of MUET speaking test to construct the prototype speaking proficiency test.

As it was discussed earlier, the PSPT is divided into three different levels of speaking tasks based on Littlewood’s (1981) Methodological Framework which is also supported by Brown (2001). Therefore, MUET past year examination speaking tasks and ideas from IELTS were considered to construct the PSPT. This analysis was carried out to develop an ESL prototype speaking proficiency test to identify the undergraduates’ speaking proficiency.



## **2.12 Conceptual Framework**

Conceptual framework serves to set the research objectives and research questions of the study and theories that underline the study within the broader context of existing knowledge. In other words, the conceptual framework helps the researcher to plan to conduct the study using the existing body of literature as a starting point of learning conduct a new research project (Ravitch & Riggan, 2012, as cited in Damro, 2015).

In this study, the conceptual framework is constructed based on the notion that a grade or score by itself does not tell much about the specific speaking proficiency of the students (Oosterhof, 2001). Based on this principle, identifying and profiling the ESL undergraduates' speaking proficiency is essential in order to assist the ESL lecturers to arrange their goals and make necessary plans to maximize their students' speaking proficiency through improving the speaking lessons and plan different types of teaching instruction. By diagnosing what the students are able to do, the ESL lecturers will be able of planning their teaching materials according to their students' current needs.

To conduct the current study, four issues are of central concern. The first issue concerns the development of an ESL prototype speaking proficiency test. In developing the prototype speaking proficiency test, the speaking activities specified in the syllabus of the English language courses taken by the USM undergraduates in line with four other major sources of Littlewood's (1981) Methodological Framework, Long's Interaction Hypothesis, Kathleen Bardovi-Harlig's (2015) Operationalizing

Conversation Speech Acts and public exams e.g., IELTS and MUET were analyzed and taken into account. It is of vital importance that the measuring instrument prepared to assess the undergraduates' speaking proficiency is valid and related. Therefore, a careful procedure with regards to validity and reliability issues were taken into account.

The second issue involves the development of speaking proficiency descriptors. In developing the speaking proficiency descriptors, the researcher has developed the number of bands for the speaking proficiency performance based on cut scores for each band. The cut scores have been developed based on the z-scoring system.

The third issue relates to identifying the undergraduates' speaking proficiency. Generally speaking, identifying is used in education and other fields in different ways to improve quality. Nonetheless, in this study, assessment for learning is applied to substitute the traditional method of assessing students for the purpose of improving the students' English-speaking proficiency. Lastly, the fourth issue is profiling the undergraduates' speaking proficiency.

It is worth noting that all these four issues are closely interrelated. They are used as guiding concerns in the forming of the conceptual framework of this study. A representation of the conceptual framework of the study has been illustrated in Figure 2.1 below.

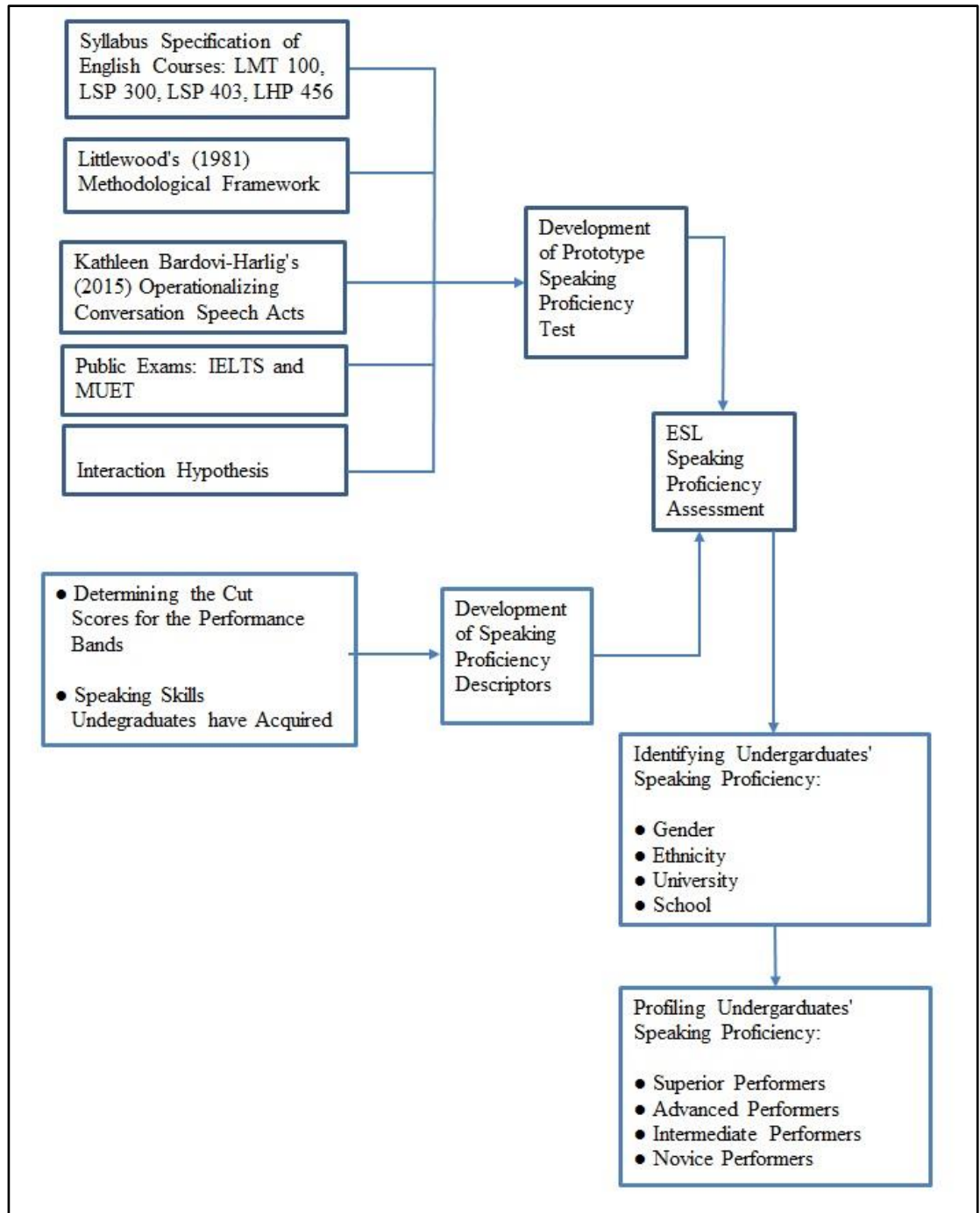


Figure 2.1: Representation of conceptual framework of study

### **2.13 Conclusion**

The current chapter starts off with a discussion on speaking and the importance of speaking proficiency in language teaching and learning for language learners. The chapter also sheds some light on assessment during language learning, speaking assessment and types of assessment such as summative assessment, formative assessment. In addition, it elaborates on the issue of test instrument validation in terms of validity and reliability. The chapter also presented the conceptual framework of the study. The next chapter will address the necessary methodology used in this study.

## **CHAPTER 3**

### **METHODOLOGY**

#### **3.1 Introduction**

Basically, this chapter provides the overview of the specific methodology used in the study in order to assist the current study find the answers to the research questions raised in the first chapter. In this chapter, the researcher briefly elaborates on how the research instrument was constructed. The chapter begins with the research design which guides the research approach. Following this, the chapter presents the research procedure, reliability, and validity, in short, scoring procedure, respondents of the study and sampling procedure. This chapter also describes the necessary data collection procedure and data analysis for identifying and profiling purposes. What follows is the detailed explanation of each section.

#### **3.2 Research Design**

One of the backbones of a research lies in its research design. In other words, the research design is the researcher's plan to answer the research questions guiding the research. A good research design will combine quantitative and qualitative methods that highlight the issue in a complementary way (Hinkel, 2005). In addition, Creswell (2012) claims that the use of quantitative and qualitative methods in combination provides a better insight and understanding of the research problem than either research method alone. Therefore, to conduct this study, a mixed methods research design is used.

A mixed methods research design is used because it helps the researcher to collect the desired data necessary to answer the research questions. In adopting a mixed methods research approach, this study employed a sequential explanatory design, consisted of two distinct phases: quantitative followed by qualitative. The collection of the quantitative data involved data collected from an ESL speaking proficiency test and biographical information of the respondents of the study while the collection of the qualitative data involved the use of follow-up structured interviews with 9 students to explain the results of the quantitative data. Therefore, the researcher views this study in two phases, quantitative phase followed by qualitative phase. The necessary procedure to collect the necessary data and information for the study follows a sequential order as it is illustrated in the research design in Figure 3.1 below.

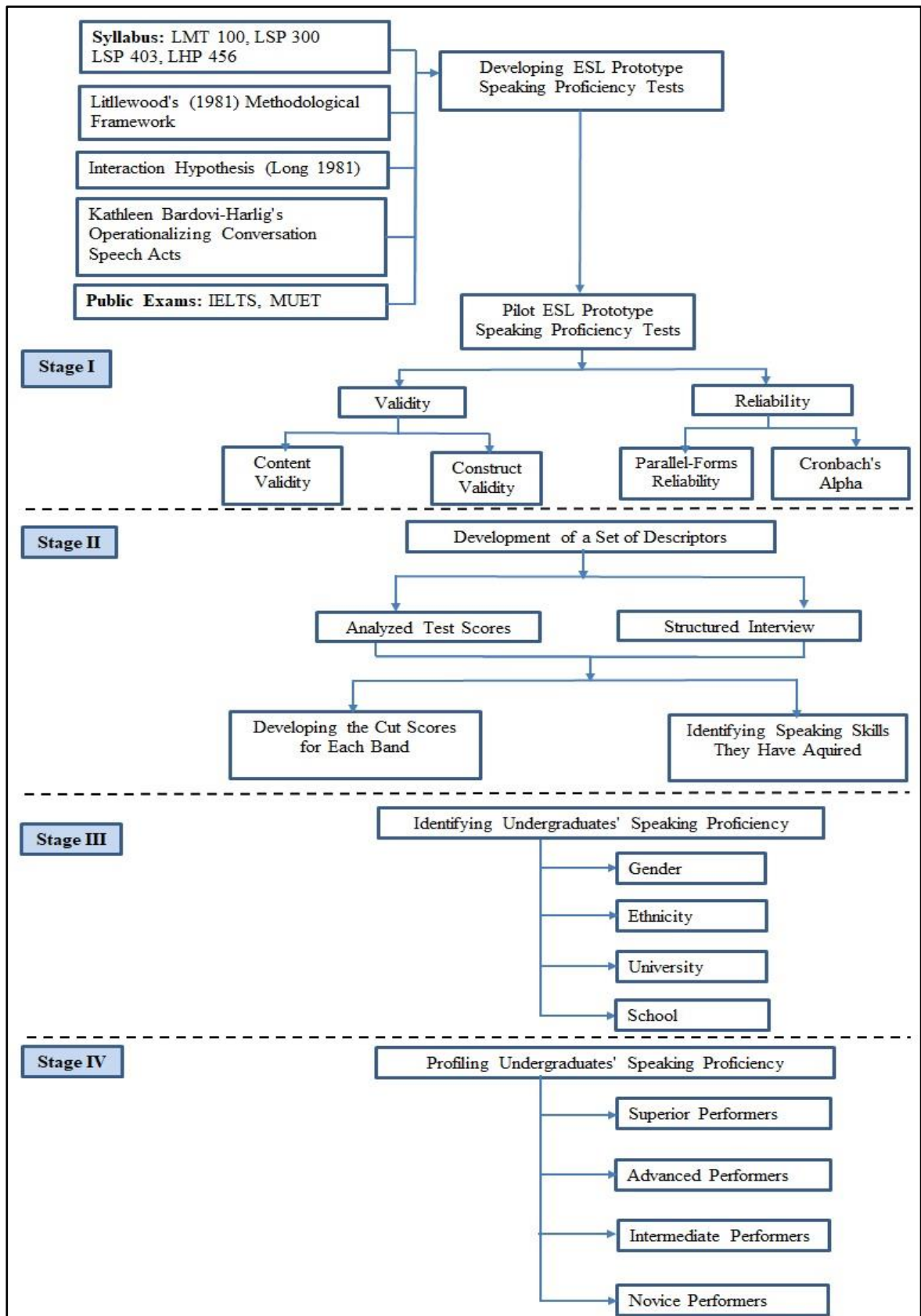


Figure 3.1: Research design

### **3.3 Procedure of Study**

This section presents the necessary procedures for the study. In order to construct the undergraduates' speaking proficiency standards, a number of processes were taken throughout the study. In this section, the processes that have been taken place were put into stages as explicated below. In this study, in order to construct the undergraduate's speaking proficiency standards, the specific following stages were conducted as it is also illustrated in Figure 3.1.

- a. Stage 1: Development of ESL prototype speaking proficiency tests and then Piloted to test validity and reliability,
- b. Stage 2: Development of a set of descriptors to identify the undergraduates' speaking proficiency,
- c. Stage 3: Administering questionnaire and ESL speaking proficiency test,
- d. Stage 4: Gathering data for analysis,
- e. Stage 5: Analyzing data for identifying students' performance,
- f. Stage 6: Identifying the undergraduates' speaking proficiency,
- g. Stage 7: Profiling the undergraduates' speaking proficiency

### **3.4 Development of ESL Prototype Speaking Proficiency Tests (PSPT)**

This section provides a brief explanation of the development of the research instrument used in the study. The process of the development of the instrument will be discussed in great length in chapter 4. However, for the purpose of this methodology chapter, I will highlight certain parts of the development in a nutshell. Therefore, the next chapter will be dedicated solely to explain the development of the instrument from the beginning right to the usage.



In developing the PSPT, five important documented materials were carefully taken into account and analyzed. First, the syllabus of the English language courses offered at the School of Languages, Literacies, and Translation at USM was taken. In this study, the different sections of the ESL speaking proficiency test are constructed in line with the tasks appear in the different English courses offered at the School of Languages, Literacies and Translation, e.g., question and answer practice, job interview practice and group discussion. Second, Littlewood's Methodological Framework was studied. Littlewood's (1981) Methodological Framework arranges speaking proficiency teaching and speaking tests from pre-communicative activities to communicative activities. Therefore, in developing the PSPT Littlewood's Methodological Framework was taken into consideration.

Third, the speech acts used in conversations by Bardovi-Harlig (2015) were studied and taken. The researcher constructed 10 questions at the elementary level whereby the respondents have to write down the most appropriate answers based on the given scenarios. Spoken language can be gained through written form of language. According to Cohen and Shively (2007), written questions can be a means of assessing spoken language in the form of a written production measure, which is called "written-for-oral" tasks. Fourth, in constructing the PSPT, ideas from part B of IELTS and Task B of MUET examinations were taken into consideration. An analysis was carried out to identify the relevant information that could be extracted from part B of IELTS and Task B of MUET speaking test to develop the PSPT.

Last, The Long's (1981) Interaction Hypothesis as one of the elements of the PSPT was taken into consideration during the development of the PSPT. As the last Section of the PSPT is a group discussion of four individuals, Interaction Hypothesis was considered as it is truly communicative. In essence, a group discussion was chosen to be used as the last Section of the ESL speaking proficiency test.

### **3.5 Validity**

Because the validity issue in this study is an aspect of the process of developing the ESL prototype speaking proficiency test, it is discussed in great length in the next chapter. However, for the purpose of this methodology chapter, the validity issue is discussed briefly. To ensure that the PSPT is measuring what it is supposed to measure, the two sets of the PSPT were tested for content validity and construct validity. To evaluate the content validity, the researcher invited two experts to validate the developed test. The two experts examined the content of the test items. They examined the two sets of the prototype speaking proficiency test items based on a scale of 1 (least appropriate) to 5 (most appropriate). The developed test was examined in terms of content appropriateness in relation to the research objectives of the study as well as the vocabulary used in the test.

The results of the judgments obtained from the two experts showed the suitability of questions of the prototype speaking proficiency test. This indicates that the content validity of the research instrument is appropriate and the test can be used for the intended purpose. Moreover, amendments and improvement were made after the comments from the content experts were gathered.

To make sure the validity of the developed tests, construct validity was also established in the current study. Ninety-six mixed ability students were taken to determine the construct validity of the developed tests. This was conducted to judge to what extent the two tests mirror the construct they are purported to measure. To judge the construct validity of each set of the prototype speaking proficiency test, 48 undergraduate students among students from higher to lower language proficiency levels were taken. As the PSPT consists of three Sections, the mean score of students' performance for each Section was taken. Next, a comparison was conducted to identify the respondents' speaking proficiency at the different English proficiency levels.

The results showed that the content of the developed tests is suitable for the targeted respondents. That is, students from a higher level of English language proficiency performed better than students from the lower level of English language proficiency. For example, LMT students (lower level of English language proficiency) scored the mean score of 47.32 out of 100 in Section A in Set One of the test whereas the mean score of LSP 300 students (higher level of English language proficiency) was 63.32 out of 100. Like Set One of the test, another example of construct validity can be seen in Set Two of the developed test. For instance, the mean score of LMT100 respondents during answering Section A questions was 49.23 out of 100, while the mean score for LSP 300 respondents was 62 out of 100. That is, the tests can distinguish between students of different language proficiency levels. In essence, this shows that the two sets of the prototype speaking proficiency test have construct validity.

### 3.6 Reliability

The reliability in this study is explained in great details in the following chapter as it is an aspect of the process of the ESL prototype speaking proficiency test development. However, for the purpose of this section in the present chapter, I briefly highlight the reliability of the developed test. To determine the reliability of the ESL prototype speaking proficiency test used in this study, a pilot study was conducted before conducting the actual study. In doing so, 96 students who were undergraduate students at the time of the study sat for the developed ESL speaking proficiency tests. As such, parallel-forms reliability and internal consistency using Cronbach's Alpha for the two sets were generated to make sure the reliability of the two sets.

The parallel-forms reliability is usually used in tests that might have two or more versions available and it determines if the two versions are equivalent (Linn & Miller, 2005). In this study, a parallel forms reliability for the two sets of the prototype speaking proficiency tests was tested through the Pearson product-moment correlation analysis. The correlation between Set One and Set Two was positive and strong ( $r = 0.81, p < .01$ ). This result shows that the participants' speaking proficiency in both sets is consistent.

In addition to Parallel Forms Reliability, the Cronbach's Alphas was generated. Because the ESL speaking proficiency test in this study is a test where there is no wrong or correct answer, the Cronbach's Alpha is said to be more suitable to be utilized to judge the internal consistency reliability of the two sets. Therefore, the prototype speaking proficiency tests were tested for internal consistency, using the Cronbach's Alpha method. The reliability coefficient for Set One of the prototype speaking

proficiency test through Cronbach's Alpha was 0.83 while for Set Two of the prototype speaking proficiency test was 0.82.

### **3.7 Scoring Rubrics**

One of the biggest challenges of developing the speaking proficiency tests was the need to design a speaking proficiency test and evaluation rubrics that could precisely evaluate the sociolinguistic competence and linguistic competence of the language learners. Therefore, the researcher developed the analytical rating (analytical rubric) scale in such a way to accurately measure the undergraduates' sociolinguistic competence and linguistic competence. The speaking proficiency test is constructed in three different difficulty levels, ranging from elementary to advanced levels. This is further supported by Mok (2000). As such, the distribution of the scoring is in line with Mok's view of the difficulty of test items, that is 25% of the weighting scoring is given to easy questions, 50% average and 25% difficult.

The researcher specifically developed the evaluation rating scale based on the communicative approach to second language teaching and testing by Canale and Swain (1980). The important work of Canale and Swain maintained that communicative approach of language comprises of four areas of knowledge and proficiency: linguistic competence, sociolinguistic competence, strategic competence and discourse competence. Linguistic competence belongs to the mastery of knowledge of the language itself. Whereas sociolinguistic competence refers to the mastery of the use of the language and discourse. That is, the focus is given to the appropriateness and naturalness of the speech or discourse. Strategic competence refers to the mastery of verbal and non-verbal strategies to make up

communication breakdowns. Finally, discourse competence belongs to the capacity to use the rules and conventions of combining grammatical forms and meanings.

However, it is not always essential to assess all the aspects of communicative approach mentioned above. Canale and Swain (1980) and Canale (1983) suggested that their theoretical framework can be used as a guide to develop or use appropriate assessment criteria for a specific purpose. Most important, Chambers and Richards (1992) argued that it is always unlikely to assess all the different components at a time by a given task or given equal importance. As such, the current study focuses only on linguistic competence and sociolinguistic competence to assess the undergraduates' speaking proficiency. As such, the scope of the analytical rating scale assessment covered appropriateness of speech, grammar, communicative ability, managing a discussion, fluency, pronunciation, and vocabulary (see Table 3.1).

It has to be mentioned that after the scoring rubric was finally constructed for the purpose of the current study, it was sent to the same two content experts who validated the speaking proficiency tests developed in this study. That is, the scoring rubric was checked and validated in relation to the research objectives to make sure the scoring rubric is able to accurately assess students' speaking proficiency.

There are two main methods to score speaking tests namely; holistic rating score and analytical rating score (Thornbury, 2005; Taylor, 2011). Holistic scales, as Luoma (2004), elicits "overall impression of an examinees' ability in one score" (p. 61). Whereas analytic scales normally consist of "a number of criteria, 3-5, each having

descriptors at the different levels of the scale” (Luoma, 2004, p. 68). More important, in analytical rating scoring, a single score is given for different language aspects, for example, vocabulary, communicative activity, pronunciation, and fluency. It has to be mentioned that holistic rating scoring is faster and it is mostly used to know an overall impression of student progress. Although analytical rating scoring is more time-consuming, it is more reliable and fairer if the task is well-scored. More important, the analytical rating score gives the opportunity to know the learner’s strengths and weaknesses in detail which helps the learner and the teacher as well to take the student’s weaknesses into consideration for later improvement. Further, it is important to use the analytical rating scale for scoring to promote the reliability of the assessment, which can provide clear information pertaining to teaching and learning improvement (Srikaew, Tangdhanakanond, & Kanjanawasee, 2015).

Likewise, Fulcher (2003) mentioned that a rating scale, sometimes known as a scoring rubric is widely used to assess learner performance on authentic language speaking assessments. Therefore, in this study, analytical rating score was used to mark the students’ speaking proficiency. Interestingly, the use of analytical rating scale makes oral tests mentally easier and more consistent for the assessor during marking (Underhill, 1987).

In the evaluation rubrics, the primary attention is given to the communicative aspects and secondary attention to the accuracy aspects of speaking. Therefore, a larger portion of the mark is allocated to the sociolinguistic competence than to the linguistic competence. When marks are given differently for different language aspects during

assessment, a weighting system is preferred to be used. A weighting system is a procedure in which scores are given out of the same total initially for different aspects and then these marks are multiplied up by different factors to give them more or less influence in the total score (Underhill, 1987). It is certainly difficult to mark one category out of ten, another one out of twenty and a third one out of thirty, at the same time. Hence, marks are given out of the same total initially and then multiplied the marks to obtain a weighted score. The rating scale was used in this study was particularly created to accurately assess the students' speaking proficiency.

Analytical rating scales are frequently used to assess language learners' proficiency within one single modality, for example, speaking. This method of scoring can also help in providing more accurate information about learners' performance and it can guide continuous feedback about progress (Stiggins, 2001, cited in Reddy & Andrade, 2010). Moreover, using analytical rating score is considered important in the sense that it gives the potential to capture necessary information about learners' responses and thus more information learner's areas of strengths and weaknesses, than does giving a single right/wrong score (Bachman & Palmer, 1996). This is further supported by Shepard (2000) as well as Brookhart (2003).

To score the respondents' responses in the current study, scale levels which are also known as criterion-referenced were used. That is, criterion-referenced from zero (no evidence of knowledge) to five (complete evidence of knowledge) was used. Responses were not scored right or wrong, they were scored in terms of degrees of correctness which is also called partial-credit scoring (Bachman & Palmer, 1996). Given



numbers of 0 (no evidence of knowledge) to 5 (complete evidence of knowledge) were used to avoid complexity or inconsistency. Further to that, it would be easy to create rating scales with fifteen ability levels, but it is unlikely that raters are able to make distinctions between so many rating abilities with any kind of consistency. Therefore, ability levels of 0 to 5 are preferred and most appropriate (Bachman & Palmer, 1996).

As such, the rating scale gives the raters prepared descriptions and he/she then chooses the one which best suits the learner. Table 3.1 illustrates the speaking proficiency assessment rubrics and rating scales.

Table 3. 1

*Assessment Rubrics for Speaking Proficiency*

| Levels of ability/mastery                              | Description  |                              |
|--|--|------------------------------|
| 0. None  | No evidence of knowledge   |                              |
| 1. Very limited  | Evidence of very limited knowledge   |                              |
| 2. Limited   | Evidence of limited knowledge  |                              |
| 3. Moderate  | Evidence of moderate knowledge   |                              |
| 4. Extensive   | Evidence of extensive knowledge  |                              |
| 5. Good  | Evidence of complete knowledge   |                              |
| <b>Section A:</b>                                      |  |                              |
| Aspects to assess                                      | Proficiency  | Score achieved               |
| Appropriateness:<br>(Sociolinguistic competence)       | <ul style="list-style-type: none"> <li>He/she answers appropriately in given context for the intended purpose/ good command of form and function.</li> </ul>   | 0/1/2/3/4/5----×3=---<br>/15 |
| Grammar:<br>(Linguistic competence)                    | <ul style="list-style-type: none"> <li>He/she uses accurate and correct grammar.</li> </ul>  | 0/1/2/3/4/5----×2=---<br>/10 |
| Total: ----/25=----                                    |  |                              |
| <b>Section B:</b>                                      |  |                              |
| Appropriateness:<br>(Sociolinguistic competence)       | <ul style="list-style-type: none"> <li>He/she speaks appropriately in given context for the intended purpose.</li> </ul>   | 0/1/2/3/4/5----×3=---<br>/15 |
| Communicative ability:<br>(Sociolinguistic competence) | <ul style="list-style-type: none"> <li>He/she is able to answer questions meaningfully.</li> <li>He/she is able to demonstrate well in conveying his/her message.</li> </ul>                               | 0/1/2/3/4/5----×3=---<br>/15 |
| Fluency and Pronunciation:<br>(Linguistic competence)  | <ul style="list-style-type: none"> <li>He/she speaks fluently and smoothly.</li> <li>He/she speaks without any pausing for too long.</li> <li>He/she pronounces the individual words correctly.</li> </ul> | 0/1/2/3/4/5----×2=---<br>/10 |

|  |  |                              |
|--|--|------------------------------|
| Grammar and vocabulary:<br>(Linguistic competence) | <ul style="list-style-type: none"> <li>• He/she is able to express stress and intonation correctly.</li> <li>• He/she uses accurate and correct grammar.</li> <li>• He/she uses a range of correct grammatical sentences.</li> <li>• He/she uses a wide range of vocabulary effectively.</li> <li>• He/she uses appropriate vocabulary.</li> </ul> | 0/1/2/3/4/5=---×2=---<br>/10 |
|--|--|------------------------------|

Total: ----/50=----

---

Section C:

|  |  |                    |
|--|--|--------------------|
| Appropriateness:<br>(Sociolinguistic competence)       | <ul style="list-style-type: none"> <li>• He/she speaks appropriately in given context for the intended people.</li> </ul>  | 0/1/2/3/4/5=--- /5 |
| Communicative ability:<br>(Sociolinguistic ability)    | <ul style="list-style-type: none"> <li>• He/she is able to communicate effectively with the other candidates.</li> <li>• He/she is able to demonstrate good interactive ability in carrying out the discussion.</li> <li>• He/she is able to maintain eye contact with the other candidates skillfully.</li> </ul> | 0/1/2/3/4/5=--- /5 |
| Managing a discussion:<br>(Sociolinguistic competence) | <ul style="list-style-type: none"> <li>• He/she is able to develop a discussion and manage it in terms of:               <ul style="list-style-type: none"> <li>➤ Initiating</li> <li>➤ Turn-taking</li> <li>➤ Interrupting</li> <li>➤ Negotiating</li> <li>➤ Closing</li> </ul> </li> </ul>                       | 0/1/2/3/4/5=--- /5 |
| Fluency and pronunciation:<br>(Linguistic competence)  | <ul style="list-style-type: none"> <li>• He/she speaks fluently and smoothly.</li> <li>• He/she speaks without any pausing for too long.</li> <li>• He/she pronounces the individual words correctly.</li> <li>• He/she is able to express stress and intonation correctly.</li> </ul>                             | 0/1/2/3/4/5=--- /5 |
| Grammar and vocabulary:<br>(Linguistic competence)     | <ul style="list-style-type: none"> <li>• He/she uses accurate and correct grammar.</li> <li>• He/she uses a range of correct grammatical sentences.</li> <li>• He/she uses a wide range of vocabulary effectively.</li> <li>• He/she uses appropriate vocabulary.</li> </ul>                                       | 0/1/2/3/4/5=--- /5 |

Total: ----/25=----

Total score of Sections A+B+C= ----/100=----

---

### 3.8 Interview Protocols

The qualitative data for the current study came from interview. According to Wellington (2015), interviews can assist researchers to investigate and prompt things

that cannot be observed by other methods of research. In addition, interviewees' thoughts, values, perceptions, and views can be discovered through conducting interviews. In this study, structured interviews were utilized to confirm the results of the ESL speaking proficiency test. The interviews were also used to strengthen in developing the speaking proficiency descriptors. To analyze the qualitative data, the transcribed data went through content analysis.

The respondents of the structured interview are nine respondents who took the ESL prototype speaking proficiency test to elicit their specific opinion on the test; what difficulties they have encountered, what they can do and what they cannot do with regards to speaking proficiency. The respondents were contacted again by the researcher to participate in the qualitative data after their performance band was recognized. The 9 respondents have been selected randomly from different bands. Three respondents from each band (Band Two, Band Three and Band Four) who took part in the PSPT. However, the qualitative data did not include Band One since none of the respondents was categorized as Band One. The specific answers from the respondents were used to strengthen develop the speaking proficiency descriptors.

In general, there are four types of interviews namely; structured, semi-structured, informal and retrospective. Structured and semi-structured interviews are verbal questionnaires containing a sequence of questions to obtain specific answers from respondents. Whereas informal interviews are similar to casual conversations which don't include any specific type of questioning. Finally, a retrospective interview is the

one which gets a respondent to recall and retrieve from memory something that has happened previously (Fraenkel et al., 2012).

However, the interview session for the present study followed a structured type of interview and selected respondents are asked same questions in the same order. Interview questions in the current study were developed by the researcher to elicit ideas from respondents about the speaking proficiency test and what they could do and could not do (See Appendix A). After the structured interview was developed, it was checked for its credibility and validity in relation to the research objectives by the two content experts who validated the prototype speaking proficiency test.

### **3.9 Administering Questionnaire and ESL Speaking Proficiency Test**

The respondents' biodata such as gender, ethnicity, university, and school are obtained through questionnaire (See Appendix B). To identify the respondents' speaking proficiency, the ESL speaking proficiency test was administered. After that, the necessary data were coded and keyed into SPSS, version 23 for descriptive analysis using frequencies and percentages.

#### **3.9.1 Population and Sampling Procedure**

This section introduces the participants who were involved in the current study to collect the necessary data. A sample is a portion of the research population selected to participate in a study or a group on which data is attained, representing the population where the study is conducted (Fraenkel & Wallen, 2009). One hundred and forty students from two public universities in northern Malaysia served as participants of the present

study. The participants of the current study are 80 first-year undergraduates of 20 different schools at university A as well as 60 first-year undergraduates of 15 different schools at university B. To identify and profile the undergraduates' speaking proficiency, the study included respondents from year one from different schools of both universities.

This study employed the stratified sampling to gather the data necessary for the study. "Stratified sampling is a process in which certain sub-groups or strata are selected for the sample in the same proportion as they exist in the population" (Fraenkel & Wallen, 2009, p. 94). They also believe that using stratified sampling is central to increasing the likelihood of representativeness. Stratified sampling also ensures the key characteristics and certain kinds of individuals will be included in the sample of the population where the study is conducted.

Stratified sampling was used to ascertain that there would be the participation of students from a variety of different schools of both universities, male and female students and also students from different ethnic groups. A simple sample of the human population may be stratified by sex, ethnicity, geographical places and so on (Thompson, 2012). Choosing this type of sampling is the intent of the researcher to have respondents from different schools of both universities and respondents from different groups of Malay, Chinese and Indian ethnicities. Table 3.2 illustrates the respondents of the study.

Table 3.2

*Respondents of Study*

| University   | Background       | Frequency (f) | Percentage (%) |
|--------------|------------------|---------------|----------------|
| University A | <u>Gender</u>    |               |                |
|              | Male:            | 27            | 33.7           |
|              | Female:          | 53            | 66.3           |
|              | <u>Ethnicity</u> |               |                |
|              | Malay:           | 47            | 58.8           |
|              | Chinese:         | 22            | 27.5           |
| University B | Indian:          | 11            | 13.7           |
|              | <u>Gender</u>    |               |                |
|              | Male:            | 26            | 43.3           |
|              | Female:          | 34            | 56.7           |
|              | <u>Ethnicity</u> |               |                |
|              | Malay:           | 41            | 68.3           |
| Total        | Chinese:         | 12            | 20             |
|              | Indian:          | 7             | 11.7           |
|              | <u>Gender</u>    |               |                |
|              | Male:            | 53            | 37.9           |
|              | Female:          | 87            | 62.1           |
|              | <u>Ethnicity</u> |               |                |
| Malay:       | 88               | 62.9          |                |
| Chinese:     | 34               | 24.3          |                |
| Indian       | 18               | 12.8          |                |

**3.9.2 Gathering Data for Analysis**

To collect the necessary data for the present study, the researcher first showed a permission letter which was provided by the School of Educational Studies, Universiti Sains Malaysia to the undergraduates who were tested. The researcher personally visited the respective respondents' schools at both universities in order for him to approach the undergraduates and, groups of four undergraduates were chosen because Section C of the ESL test requires four individuals together at a time. Four students were chosen based on their gender and ethnicity to ascertain the participation of both male and female students as well as students from different ethnicities.

First, each respondent got an identification number (ID) in order to identify their speaking proficiency during marking later. Second, the respondents were given Section A of the test to complete and are given 11 minutes to complete. Later, the researcher sat with the respondents individually to conduct Section B of the speaking test and a proper voice recorder was used to record the answers. Section B lasts 2.5 minutes. Finally, in groups of four, the respondents got together and sat at a table to conduct Section C of the speaking test and a video recorder was used this time. This Section lasts 8 minutes. This way the researcher continued until the necessary data was gathered.

### **3.10 Data Analysis**

The present study is a mixed methods research study. Therefore, it included quantitative and qualitative methods to collect the desired data to conduct the study. What follows is the detailed explanation of data analysis for both quantitative and qualitative data.

#### **3.10.1 Quantitative Data**

In order to get the necessary data to identify the respondents' speaking proficiency, an ESL speaking proficiency test was administered among 140 first-year undergraduates of each respective school at both public universities in northern Malaysia. The quantitative data collected from the speaking proficiency test was keyed into the Statistical Package of Social Sciences (SPSS) version 23. Descriptive statistics, utilizing percentages were utilized to describe the respondents' speaking proficiency performance which was achieved from the speaking proficiency assessment.

### **3.10.2 Qualitative Data**

In this study, the qualitative data came from interview. Interview data helped to validate the data from the ESL speaking proficiency test. The data gained from the structured interview questions were transcribed using the verbatim procedure and then went through content analysis. As such, the identification of themes (categories) that emerged from the analysis has been taken into consideration for the purpose of the analysis.

The content analysis which is a widely-used method in qualitative studies is a method that assists researchers to study human behavior in an indirect way via the analysis of communication. It is the analysis of an individual's beliefs, attitudes, and ideas which are often addressed in the form of communication (Fraenkel et al., 2012). Via content analysis, condensing words into fewer content-related categories is possible. It is believed that when classified into categories, words, phrases and the similar deliver the same meaning (Cavanagh, 1997) and the aim of content analysis is to gain a distilled description of a phenomenon (Elo & Kyngas, 2008).

Current applications of content analysis reveal three different approaches of content analysis namely; conventional, directed and summative (Hsieh & Shannon, 2005). These approaches are used to interpret meaning from the content of text data. In the conventional content analysis, themes are taken directly from the text data. With the directed approach, analysis begins with a theory or related research results and findings as guidance for initial codes. While in a summative approach content analysis, counting and comparisons are involved, usually in the form of keywords and quantifying words,



followed by the interpretation of what they reveal (Hsieh & Shannon, 2005). However, the current study used conventional approach of content analysis to identify themes in organizing content and arriving at a description of data derived from the structured interviews.

The first step of analyzing the interview data was precise reading and defining of the aspects of the content of the interview transcripts. During the process of reading and marking the transcripts, the researcher started to label the passages that are interesting. With consequent to that, themes that were related to the analysis of the content of the interviews were formulated. After establishing the themes form the data, the researcher invited two colleagues for checking and revising the extracted themes (Creswell, 2012). This was conducted to ensure the credibility of the analysis conducted. As such, through the repetitive analysis, any irrelevant themes were discarded and others that provided additional information were also added.

As for labels of reference used in chapter 5, they were created to show that discussion was substantial with the data. Hence, labels were followed to refer to participants of the study. Table 3.3 shows the labels of reference of the qualitative data.

Table 3.3

*Labels of Codes of Participants*

| Bands      | Code of Participants |
|------------|----------------------|
| Band Two   | B2-1-22/10/2016      |
|            | B2-2-22/10/2016      |
|            | B2-3-22/10/2016      |
| Band Three | B3-1-25/10/2016      |
|            | B3-2-25/10/2016      |
|            | B3-3-25/10/2016      |
| Band Four  | B4-1-29/10/2016      |
|            | B4-2-29/10/2016      |
|            | B4-3-29/10/2016      |

### **3.11 Identifying the Undergraduates' Speaking Proficiency**

In this study, assessing students' speaking proficiency is used to identify the undergraduates' speaking proficiency as an in-depth analysis of the undergraduates' proficiency in speaking. To identify the undergraduates' speaking proficiency, the performance of the undergraduates on the ESL speaking proficiency test is used. Besides, the undergraduates' speaking proficiency is identified by referring the test scores to the speaking performance bands and descriptors.

### **3.12 Profiling the ESL Speakers' Speaking Proficiency**

Profiling the undergraduate students' speaking proficiency is used to profile the individual groups. As such, in this study, the respondents were categorized into four categories e.g., superior performers, advanced performers, intermediate performers and novice performers.

### **3.13 Summary**

This chapter specifically illustrates the methodology used in the study. The chapter also briefly explained how the research instrument was developed, how it was administered and how the necessary data for the study was collected, stored and finally analyzed. This chapter also explained the procedure of the study and the specific stages needed for the study. Table 3.4 presents a summary of the objectives of the study, research questions, source of data and technique of data collection and method of data analysis. The next chapter will highlight the process of developing the prototype speaking proficiency test in details.

Table 3.4

*Research Matrix*

| Research Objectives   | Research Questions   | Source of Data  | Technique of Data Gathering  | Method of Data Analysis   |
|---|--|---|--|---|
| <p>1) To develop an ESL speaking proficiency test for undergraduates by establishing the following:</p> <ul style="list-style-type: none"> <li>• Selection of relevant sources for the construction of the ESL prototype speaking proficiency test</li> </ul> | <p>1) How is an ESL speaking proficiency test for undergraduates developed?</p> <ul style="list-style-type: none"> <li>• What are the relevant sources for the construction of the ESL prototype speaking proficiency test?</li> </ul> | <ul style="list-style-type: none"> <li>• Syllabus specifications: (School of Languages, Literacies, and Translation): LMT 100, LSP 300, LSP 403 and LHP 456</li> <li>• Kathleen Bardovi-Harlig's Operationalizing Conversation Speech Acts (2015)</li> <li>• Littlewood's Methodological Framework (1981)</li> <li>• Public examinations: IELTS and MUET</li> <li>• Interaction Hypothesis (Long 1981)</li> </ul> | <ul style="list-style-type: none"> <li>• Document gathering</li> </ul>   | <ul style="list-style-type: none"> <li>• Document analysis</li> </ul>   |
| <ul style="list-style-type: none"> <li>• Determining the best combination of questions for the ESL speaking proficiency test</li> </ul>   | <ul style="list-style-type: none"> <li>• What would be the best possible combination of ESL speaking proficiency questions selected for the ESL speaking proficiency test?</li> </ul>  | <ul style="list-style-type: none"> <li>• conversation speech acts, IELTS and MUET</li> </ul>  | <ul style="list-style-type: none"> <li>• Document gathering</li> </ul>   | <ul style="list-style-type: none"> <li>• Document analysis</li> </ul>   |
| <ul style="list-style-type: none"> <li>• Conducting pilot study to test the validity and reliability of the ESL PSPT</li> </ul>   | <ul style="list-style-type: none"> <li>• What is the validity and reliability of the ESL prototype speaking proficiency test?</li> </ul>   | <ul style="list-style-type: none"> <li>• Content experts</li> <li>• Respondents' tests scores</li> </ul>  | <ul style="list-style-type: none"> <li>• Document gathering</li> <li>• Prototype speaking proficiency tests</li> </ul> | <ul style="list-style-type: none"> <li>• Analyzing content experts' comments</li> <li>• Descriptive statistics</li> <li>• Cronbach's Alpha</li> <li>• Parallel-Forms Reliability</li> </ul> |
| <ul style="list-style-type: none"> <li>• Determining time allocated for the ESL speaking proficiency test</li> </ul>  | <ul style="list-style-type: none"> <li>• What is the time allocated for the ESL speaking proficiency test?</li> </ul>  | <ul style="list-style-type: none"> <li>• Observation of time taken</li> </ul>   | <ul style="list-style-type: none"> <li>• Average time taken to answer prototype speaking proficiency tests</li> </ul>  | <ul style="list-style-type: none"> <li>• Analysis of time taken</li> </ul>  |

|  |   |  |   |   |
|--|---|--|---|---|
| <p>2) Developing a set of descriptors to identify the undergraduates' speaking proficiency performance by establishing:</p> <ul style="list-style-type: none"> <li>• Determining the cut scores for the bands</li> <li>• Identifying speaking Skills undergraduates have acquired</li> </ul> | <p>2) How are the speaking proficiency descriptors developed to identify the undergraduates' speaking proficiency performance?</p> <ul style="list-style-type: none"> <li>• What are the most suitable cut scores for the performance bands?</li> <li>• What are the speaking skills undergraduates have acquired?</li> </ul> | <ul style="list-style-type: none"> <li>• Respondents' test scores</li> <li>• Respondents' test scores</li> <li>• Respondents' test scores</li> <li>• Interview protocol</li> </ul>   | <ul style="list-style-type: none"> <li>• Prototype speaking proficiency test</li> <li>• Prototype speaking proficiency test</li> <li>• Prototype speaking proficiency test</li> <li>• Structured interview</li> </ul> | <ul style="list-style-type: none"> <li>• Descriptive statistics</li> <li>• Descriptive statistics</li> <li>• Z-Scores, standard deviation and mean</li> <li>• Descriptive statistics</li> <li>• Content analysis</li> </ul> |
| <p>3) Identifying the undergraduates' speaking proficiency according to their performance in terms of:</p> <ul style="list-style-type: none"> <li>• Respondents' general ESL speaking proficiency</li> <li>• Gender</li> <li>• Ethnicity</li> <li>• University</li> <li>• School</li> </ul>  | <p>3) How do the undergraduates perform on the ESL speaking proficiency test?</p> <ul style="list-style-type: none"> <li>• What is the respondents' general ESL speaking proficiency?</li> <li>• What is the overall ESL speaking proficiency performance according to gender, ethnicity, university, and school?</li> </ul>  | <ul style="list-style-type: none"> <li>• Analyzed respondents' test scores</li> <li>• Analyzed respondents' test scores</li> <li>• Analyzed respondents' test scores</li> </ul>  | <ul style="list-style-type: none"> <li>• ESL speaking proficiency test</li> </ul>   | <ul style="list-style-type: none"> <li>• Descriptive statistics: Percentages</li> </ul>   |
| <p>4) Profiling the undergraduates' ESL speaking proficiency particularly in terms of:</p> <ul style="list-style-type: none"> <li>• Superior performers</li> <li>• Advanced performers</li> <li>• Intermediate performers</li> <li>• Novice performers</li> </ul>                            | <p>4) What are the profiles of the undergraduates' ESL speaking proficiency?</p> <ul style="list-style-type: none"> <li>• Who are the superior performers?</li> <li>• Who are the advanced performers?</li> <li>• Who are the intermediate performers?</li> <li>• Who are the novice performers?</li> </ul>                   | <ul style="list-style-type: none"> <li>• Analyzed respondents' test scores</li> <li>• Analyzed respondents' test scores</li> <li>• Analyzed respondents' test scores</li> <li>• Analyzed respondents' test scores</li> </ul> | <ul style="list-style-type: none"> <li>• ESL speaking proficiency test</li> <li>• ESL speaking proficiency test</li> <li>• ESL speaking proficiency test</li> <li>• ESL speaking proficiency test</li> </ul>          | <ul style="list-style-type: none"> <li>• Descriptive statistics: Percentage</li> </ul>  |

## **CHAPTER 4**

### **DEVELOPMENT OF ESL SPEAKING PROFICIENCY TEST**

#### **4.1 Introduction**

The current chapter highlights the process of developing the ESL speaking proficiency test used in the present study. In other words, the chapter introduces the subject of instrumentation in general. It basically answers research question 1 of the study. Therefore, this chapter includes the process of the development of the ESL speaking proficiency test and the relevant sources for the construction of the ESL speaking proficiency test. It also explains the best possible combination of ESL speaking proficiency questions selected for the ESL speaking proficiency test. Following that, the chapter presents the pilot study to test validity and reliability of the ESL prototype speaking proficiency test. In addition, the chapter describes the time allocated for the ESL speaking proficiency test. Finally, it introduces the developed ESL speaking proficiency test used as the main instrument of the study. What follows is the detailed explanation of each section.

#### **4.2 Development of ESL Prototype Speaking Proficiency Test**

This section introduces the process of the development of the ESL speaking proficiency test and how it was developed to find the answers to the research questions raised in chapter one. In addition, the Section Also explicates why the test is designed in such a way. In essence, the speaking proficiency of the undergraduate students can be measured accurately by administering the ESL speaking proficiency test.

As discussed in chapter one, there is a dearth of research done on the assessment of speaking proficiency of undergraduates. Similarly, there is also a dearth of research on identifying and profiling of undergraduates' speaking proficiency in Malaysia. Therefore, an accurate ESL speaking proficiency test that can accurately evaluate the undergraduates' speaking proficiency is needed. In essence, an ESL speaking proficiency test was devised so that the undergraduates' speaking proficiency can be accurately evaluated. Administering this test will help language lecturers to know the present level of their students' speaking proficiency. Having enough information about the students' speaking proficiency will assist the language instructors to be able of adapting their teaching materials based on the current needs of the undergraduates so they can enhance their speaking proficiency accordingly. At this stage of the study, a number of activities were carried out as follows:

- I. Relevant sources of information were selected to be analyzed for the construction of the ESL PSPT.
- II. Determining the best possible combination of ESL speaking proficiency questions selected for the ESL speaking proficiency test.
- III. A pilot study was carried out to test validity and reliability of the ESL prototype speaking proficiency test.
- IV. The time allocated for the Speaking proficiency test was determined.
- V. The ESL speaking proficiency test was finally developed.

#### **4.2.1 Selection of Relevant Sources for the Construction of the ESL Prototype Speaking Proficiency Test**

The following documented materials were selected for analyzing as relevant sources of data needed to develop the ESL prototype speaking proficiency test:

- a. The syllabus of the English language courses offered at the School of Languages, Literacies, and Translation at USM
- b. Littlewood's Methodological Framework (1981)
- c. Kathleen Bardovi-Harlig's Operationalizing Conversation Speech Acts (2015)
- d. Public Examinations: IELTS Part B Speaking Test and MUET past year examinations
- e. Long's Interaction Hypothesis (1981)

#### **a. The Syllabus of the English Language Courses Offered at the School of Languages, Literacies, and Translation at USM**

In the development of the ESL prototype speaking proficiency test, the vital role of the syllabus cannot be disregarded and it is therefore considered as the main source in this regard. According to Rabinni (2002), a syllabus is "... an expression of opinion on the nature of language and learning; it acts as a guide for both teacher and learner by providing some goals to be attained." According to Jaya (2003), the syllabus is a set of subjects for the lecturer to complete in a course of study or a term. As such, important aspects found in the syllabus would be considered for analyzing as the objectives of the assessment considered in pursuit of improvement.

Moreover, Jaya (2003) states that there is a strong relationship between the syllabus and evaluation and types of questions appear in the evaluation. Therefore, the syllabus of the English language courses offered at the School of Languages, Literacies and Translation at USM which aim at improving undergraduates' overall English language proficiency to meet their needs of English language in academic settings, everyday life, and their future workplace after graduation was taken into consideration. Table 4.1 illustrates some of the activities (speaking tasks) found in the syllabus of the English courses offered at School of Languages, Literacies. and Translation.

Table 4.1

*Speaking Tasks Found in the Syllabus of the English Courses at the School of Languages, Literacies, and Translation (SOLLAT)*

| SOLLAT English Courses                         | Speaking Tasks   |
|--|--|
| LMT 100:<br>Preparatory English                | <ul style="list-style-type: none"> <li>• Ask and answer questions</li> <li>• Topic and texts discussion among students</li> <li>• Group discussion</li> </ul>  |
| LSP 300:<br>Academic English                   | <ul style="list-style-type: none"> <li>• Conversation activities such as pair work-practice and role play</li> <li>• Group discussion of four individuals.</li> <li>• Conversation activities Expressing agreement and disagreement</li> </ul>                     |
| LSP 403:<br>Business and Communication English | <ul style="list-style-type: none"> <li>• Group discussion: students are given topics to discuss among themselves</li> <li>• Oral presentations</li> <li>• Job interview practice</li> </ul>  |
| LHP 456:<br>Spoken English                     | <ul style="list-style-type: none"> <li>• Oral presentations</li> <li>• Panel discussion: Groups of four individuals discuss topics.</li> <li>• Debates: In groups of four, they explain social topics. They explain, agree, disagree and given reasons.</li> </ul> |

Consequently, the researcher in this study constructed the ESL speaking proficiency test in such a way that it is line with the speaking tasks found in the different courses of the English language at SOLLAT. For instance, practicing question and



answer, conducting job interviews as well as group discussion.

**b. Littlewood's Methodological Framework (1981)**

Another significant aspect of the development of the ESL PSPT is the Littlewood's Methodological Framework. Littlewood (1981) posited a methodological framework for language teaching in a communicative way, starting from pre-communicative to communicative tasks. In pre-communicative tasks, language learners only practise the new language structure, thus practising the total speaking proficiency is not acquired at the pre-communicative stage. This category includes the majority of learning activities found in textbooks; e.g., question-and-answer and form and function practice. These kinds of tasks essentially assist learners to practise the command of the linguistic system, but less so getting them to utilize this system for communication purposes.

Accordingly, the language learners' main aim is to produce language which is acceptable and appropriate rather than to communicate meanings effectively and powerfully. These activities and teaching tasks are merely to prepare language learners for a higher level of language production which is communication (Littlewood, 1981).

On the contrary, in communicative tasks, the learner usually integrates pre-communicative competence knowledge to use them for actual communications. Therefore, language learners are now fully engaged in practising the whole skill of speaking (Littlewood, 1981). As such, in developing the ESL PSPT, Littlewood's Methodological Framework was taken into the researcher's account. The

Methodological Framework of Littlewood has been adapted and represented as follows in Table 4.2.

Table 4.2

*Littlewood's Methodological Framework*

| Stages                       | Types of Activities                  |
|------------------------------|--------------------------------------|
| Pre-communicative activities | Structural and functional activities |
| Communicative activities     | Social interaction activities        |

Hence, the constructed tasks in the PSPT also follow the Littlewood's Methodological Framework in which he argues that tests should be constructed from pre-communicative activities to free communicative activities (Littlewood, 1981). This is further supported by Brown (2001) whereby he argues that any lesson should involve a series of activities that permit language learners to practise the language they are learning. Therefore, it is important to follow the order of the activities, considering when and why they will be practised to guarantee the class progresses smoothly from easy to difficult activities (Brown, 2001). As such, language activities should start with fully-controlled activities whereby the lecturer knows the answer and there might be one possible response. Fully-controlled activities allow language learners to focus merely on the new language structure, not meaningful communications yet. As familiarity with the new language structure increases, semi-controlled (partially-controlled) activities can be the option to practise in which there is somewhat an increased amount of freedom for the language learner. Following partially-controlled activities, free communicative activities should come last in any lesson. At this stage, language learners have absolute freedom in the language they produce and as such their lecturer will not be able to predict what will be said in the activity (Gavilán, 2008).

In this study, the Sections of the PSPT are in line with Littlewood's Methodological Framework in the way that the PSPT begins with question and answer practice which is considered as pre-communicative tasks and ends with group discussion which is considered as free communicative activities according to Littlewood's Methodological Framework.

**c. Kathleen Bardovi-Harlig's Operationalizing Conversation Speech Acts (2015)**

Another crucially significant aspect of constructing the PSPT is the speech acts utilized in conversations by Bardovi-Harlig (2015). The researcher adapted 10 speech acts used in conversations as Section A of the PSPT. The researcher constructed 10 questions whereby students need to write down the most suitable responses based on the scenarios provided. These questions construct Section A of the PSPT. In fact, spoken language can be achieved via a written form of language. Written questions can be a means for measuring spoken language in the form of a written production measure which is known as "written-for-oral" tasks (Cohen & Shively, 2007). Written-for-oral tasks require language learners to produce language in writing what they would say in conversation. Written-for-oral tasks can be scenarios to which respondents respond in writing (Eslami & Liu, 2013).

**d. Public Examinations: IELTS Part B Speaking Test and MUET Past Year Examinations**

In designing the ESL speaking proficiency test, ideas for the ESL speaking tasks from part B of IELTS and Task B of MUET examinations were studied and considered.

In doing so, an analysis was conducted to find out the related information that could be taken from part B of IELTS and Task B of MUET speaking tests in constructing the ESL speaking proficiency test.

As it was discussed previously, the PSPT is categorized into three different levels of speaking tasks based on Littlewood's Methodological Framework which is also supported by Brown (2001). Littlewood (1981) proposed a methodological framework for teaching language communicatively, starting from pre-communicative activities to communicative activities. This category of activities is also supported by Brown (2001) and Gavilán (2008) who argue that language activities should start with fully-controlled activities, semi-controlled activities, and free controlled activities. As such, the ESL speaking proficiency test is built in three different difficulty levels, ranging from elementary to advanced levels. As such, MUET past year examination speaking tasks and ideas from IELTS were considered to develop the PSPT. This analysis was conducted to develop a prototype speaking proficiency test to identify undergraduates' speaking proficiency.

**e. Long's Interaction Hypothesis (1981)**

The Long's (1981) Interaction Hypothesis as one of the crucial aspects of the PSPT was taken during constructing the PSPT. Including a group discussion task as the last section (Section C) of the ESL prototype speaking proficiency test was based on the Interaction Hypothesis. In this section, the respondents engage in a group discussion in a way that each respondent plays a role whereby he/she has to present his/her own ideas about the topic of discussion and then they continue discussing their

points of view on the given topic. According to Long (1981), the interaction between language learners creates acute language learning and improvement where learners learn via negotiation of meaning. Furthermore, research has proved that interaction among language learners is effective in helping language learning (Ellis, 1999; Wang & Castro, 2010).

As such, the Interaction Hypothesis plays a vital role in learning via conversation and interaction in improving language learners' speaking proficiency. This is further supported by Cross (1991) who states that if your target is to develop speaking proficiency, you need to embrace group discussion which is an excellent way to assess the speaking proficiency of language learners. It is crucial to note that Section C of the ESL speaking proficiency test is a fully free communicative task since all the interlocutors engage in a free group discussion. Therefore, a group discussion has been suggested to be included as Section C in the PSPT. Finally, the five major sources that were selected for constructing the ESL speaking proficiency test can be summarised in Table 4.3.

Table 4.3

*Summary of Sources of Data for PSPT*

| Sources of Data   | English Language Courses and Tasks  |
|---|---|
| I) The syllabus of the English language courses offered at the School of Languages, Literacies and Translation at USM | <ul style="list-style-type: none"> <li>• LMT 100: Preparatory English</li> <li>• LSP 300: Academic English</li> <li>• LSP 403: Business and Communication English</li> <li>• LHP 456: Spoken English</li> </ul> |
| II) Littlewood's Methodological Framework   | <ul style="list-style-type: none"> <li>• Pre-communicative</li> <li>• Communicative</li> </ul>  |
| III) Kathleen Bardovi-Harlig's Operationalizing Conversation Speech Acts  | Adapted speech acts   |
| IV) Public Tests: <ul style="list-style-type: none"> <li>• IELTS Part B Speaking Test</li> </ul>                      | <ul style="list-style-type: none"> <li>• Stimulus-response</li> <li>• Group discussion</li> </ul>   |

---

|                                  |                                     |
|----------------------------------|-------------------------------------|
| • MUET past year examinations    |                                     |
| V) Long's Interaction Hypothesis | Interaction among language learners |

---

#### **4.2.2 Determining the Best Possible Combination of ESL Speaking Proficiency**

##### **Questions Selected for the ESL Prototype Speaking Proficiency Test**

The ESL prototype speaking proficiency test consists of three sections namely; Section A, Section B, and Section C. The test is constructed in three different difficulty levels, ranging from elementary to advanced levels as proposed by Brown (2001). He argues that any lesson should involve a series of activities that allow language learners to practise the language. As such, class activities may start with fully-controlled activities, partially-controlled activities and finally, free communicative activities. Accordingly, language assessments should also follow the order from elementary to advanced levels.

This is further supported by Littlewood's (1981) Methodological Framework in which he argues that tests should be constructed from pre-communicative activities to free communicative activities. More important, the three categories of questions were built in three different difficulty levels which consist of different types of questions.

As for the PSPT, it was decided to be constructed in such a way that comprises three different difficulty levels which are elementary to advanced levels (Mok, 2000). Section A consists of ten questions in which the respondents have to write down the most appropriate answers based on the scenarios given. While in Section B the respondents are given newspaper cuttings such as a job application or teaching brochure (a flyer in the form of a stimulus or an interview) to read and then they are

asked five questions about the flyer in which they have to answer the questions verbally. The last Section of the test is a group discussion of four individuals. In this Section, the respondents engage in a group discussion in a way that each respondent plays a role whereby he/she has to present his/her own ideas about the topic of discussion and then they continue discussing their points of view on the given topic.

The speaking proficiency tests are divided into three sections namely; Section A, Section B, and Section C and it is built in three different difficulty levels (Refer to Table 4.4), ranging from elementary to advanced levels as suggested by Brown (2001). Section A constitutes 25% of the test, whereas Section B constitutes 50% of the test and finally Section C which constitutes 25% of the test as proposed by Mok (2000). The distribution of the difficulty of a test should be 25% easy, 50% average and 25% difficult (Mok, 2000).

In the light of this category, the researcher constructed 10 questions whereby the respondents have to write down the most appropriate answers based on the given scenarios, these questions constitute Section A of the PSPT. As a matter of fact, spoken language can be measured through written form of language. As Cohen and Shively (2007) mention, written questions can be a means of assessing spoken language which is called “written-to-oral” tasks (p.196). Written-for-oral tasks require learners to produce language in writing what they would say in conversation and they could be scenarios that respondents can respond in writing (Eslami & Liu, 2013, P.71). Likewise, in assessing speaking proficiency the technique of writing as if speaking can be utilized and in doing that technique, certain desired speech acts can be considered

(Norris, 1991). In this method, respondents hear a statement or read a situation-description and then write what they think they would most likely say. Further to that, Sindermann and Horsella (1989) found that students seem to signal in their writing the same strategy markers recognized in oral production such as hesitation, pausing, repetition, self-repair and crossing out.

In line with Section B of IELTS, newspaper cuttings were used as a stimulus for speaking to be given to the respondent as Section B of the PSPT. Five questions are asked based on the newspaper cutting in order for the respondent to respond verbally. The newspaper cutting is an authentic job application that most students will be in dire need of conducting job interviews after graduation. “An authentic task is one which resembles very closely something which we actually do in everyday life” (Underhill, 1987, p. 8). Semi-controlled tasks can be a stimulus which is a task that is built to motivate the language learner to speak, usually by providing some information or a subject to talk about. As such, it can be a photo to describe, a text or a particular topic (Underhill, 1987). Underhill argues that although there is somewhat freedom of expression in semi-controlled activities, the topic of discussion is still around the stimulus. He also stipulates that before the test begins, the learner is given a picture or an object to look at. Then the interviewer asks the learner questions about the stimulus. This is a good and common way of leading into a discussion which is a higher level of difficulty.

In line with Section B of MUET, the researcher used a group discussion as Section C of the PSPT. In this Section, the respondents engage in a group discussion



in a way that each respondent plays a role that he/she has to present his/her own ideas about the topic of discussion and then they continue discussing their points of view on the given topic. According to Long (1981), the interaction between language learners creates acute language learning and improvement whereby learners learn via negotiation of meaning. Furthermore, research has proven that interaction among language learners is effective in helping language learning (Ellis, 1999; Wang & Castro, 2010). As such, the Interaction Hypothesis plays a vital role in learning via conversation and interaction in improving language learners' speaking proficiency. This is further supported by Cross (1991), who states that if improving speaking proficiency is aimed, then you need to embrace group discussions which is an excellent way to assess the speaking proficiency of language learners. This approach is truly communicative since all the members engage and interact in a discussion.

In the light of the above discussion, a speaking task involves a group discussion has been preferred to be used in the PSPT as Section C. The Section item has been chosen from the MUET year 2010. The reason behind choosing the item from the previous years is that the respondent has not been exposed to the question. The first-year undergraduate students would not have sat for the MUET examination in 2010. Instead, the first-year students would have sat for the MUET examination in 2014 or 2015. Therefore, it would be appropriate to use MUET questions from previous years. Table 4.4 illustrates the three different sections of the test, the number of different types of speaking questions and references.

Table 4.4

*Sections of ESL Prototype Speaking Proficiency Test*

| Sections                                  | Types of questions                     | References  |
|---|--|---|
| Section A:<br>25% of test<br>Elementary   | 10 speech acts                         | Adapted from Kathleen Bardovi-Harlig's study on operationalizing conversation |
| Section B:<br>50% of test<br>Intermediate | Stimulus given followed by 5 questions | Idea was taken from part B of IELTS   |
| Section C:<br>25% of test<br>Advanced     | Group discussion                       | Adopted from MUET past year examinations                                      |

The next section will be specifically about the fundamental issue of the pilot study conducted to test validity and reliability of the prototype speaking proficiency test.

#### **4.2.3 Pilot Study to Test Validity and Reliability of the ESL Prototype Speaking Proficiency Test**

In general, a pilot study can be used on a small scale of respondents in preparation for a main study. The term of pilot study is defined as “a small scale of a test of methods and procedures to be used on a large scale...” (Porta, 2008, cited in Hazzi & Maldaon, 2015). In other words, a pilot study or tryout is an examination which takes place before the research instrument is used for the main study (Alderson, Clapham & Wall, 1995). Further, the main aim of a pilot study is to collect enough information about the usefulness of the instrument in order for the researcher to make revisions in the instrument, rather than to make decisions about the individuals (Bachman & Palmer, 1996). Besides, one of the advantages of doing a pilot study is that it helps the researcher in offering advice about where the actual study may fail, whether or not the proposed instrument is appropriate, so it can be said that a pilot study is an essential

element of a good study design (Van Teijlingen & Hundley, 2001). The significance of the pilot study also lies in promoting the quality and the efficiency of the actual study (Hazzi & Maldaon, 2015).

However well designed and edited a test may be, it is difficult to know how the test will work until it is tried out. It is in this regard that it is crucial to conduct the pilot study in order for the researcher to make necessary revisions by selecting the best items and make essential improvements in the test (Alderson et al., 1995). More important, a pilot study is like a mini version of the actual study; and it assists the researcher to put the study in an appropriate focus. Likewise, the pilot test might show unexpected surprises that may turn the main research into trouble.

As such, the researcher of the present study conducted a pilot study before embarking on the main study to determine the validity and reliability of the instrument used in this study. The pilot study was conducted at a quiet office at the School of Educational Studies, Universiti Sains Malaysia in the first week of March 2016.

The sample of the pilot study consisted of the undergraduate learners who enrolled in differing English courses based on their MUET score, at the School of Languages, Literacies, and Translation, Universiti Sains Malaysia. To conduct the pilot study, the researcher included a total number of 96 respondents who were undergraduates of USM, enrolled in four English courses organized by the School of Languages, Literacies and Translation, USM, 24 respondents in each course based on their MUET results to improve their overall English language competency.

The respondents consisted of students from different schools of Universiti Sains Malaysia. They were at different levels of academic study including Year One (36 respondents), Year Two (24 respondents), Year Three (20 respondents) and Year Four (16 respondents). Among the 96 respondents, 72 were females and 24 were males.

The respondents who volunteered to participate in the pilot study were students from four different groups. Firstly, group A consists of students who score Band Two or Band Three in MUET taking LMT 100: Preparatory English. Group B comprises of students with Band Four in MUET taking LSP 300: Academic English. Group C comprises of students with Band Five in MUET taking LSP 403: Business and Communication English and lastly, group D comprises students with Band Six in MUET taking LHP 456: Spoken English, which is the highest English language course level based on the school's English language courses. The pilot study was conducted comprising participants from different English proficiency levels to identify their speaking proficiency. Table 4.5 presents the respondents of the pilot study.

Table 4.5

*Respondents of the Pilot Study*

| No | Course name | MUET result         | Number of respondents | Total |
|----|-------------|---------------------|-----------------------|-------|
| 1  | LMT100      | Bands Two and Three | 24                    |       |
| 2  | LSP300      | Band Four           | 24                    |       |
| 3  | LSP403      | Band Five           | 24                    |       |
| 4  | LHP456      | Band Six            | 24                    | 96    |

The pilot study was administered to 96 respondents in four days, 24 respondents each day. Set One of the PSPT was administered on the first two days and Set Two on the second two days. Both sets of tests were administered to avoid possible influence

of exposure among participants of their performance for those who performed and those who were yet to perform tasks.

First, students were all placed in a room to sit for Section A of the test. After Section A was conducted, they were quarantined in the room before they sat for Sections B and C of the speaking test. While conducting Section B, others were waiting in the quarantine room for their turn.

It has to be mentioned that the results of the pilot study are not included in the results of the actual study as there might be modifications and improvements that are necessary to be made in the research instrument in light of the findings of the pilot study (Peat, 2001). Similarly, the participants of the pilot study are not included in the actual study as it might affect the results of the study because the tests are no longer novel for them since they have already been exposed to the tests. Therefore, they may respond differently from those participants who have not formerly experienced the tests. More important, a constructed test should be administered to a group of respondents similar to the target group (Birjandi, Bagheridoust, & Mossalanejad, 2004).

As such, this section of the study presents the results of the pilot study which was conducted to try out the ESL prototype speaking proficiency tests in advance. It is worth mentioning that prior to conducting the pilot study, a permission letter from the deputy dean of School of Educational Studies to the deputy dean of the School of Languages, Literacies, and Translation, requesting assistance was provided (See

Appendix C). The main topics of this section include the validity of ESL prototype speaking proficiency test and reliability of ESL prototype speaking proficiency test. A detailed explanation of each topic is provided as follows.

**a. Validity of ESL Prototype Speaking Proficiency Test**

Nothing could be more crucial and significant to a good test than its validity. Its capacity to measure what it is supposed to measure (Kubiszyn & Borich, 2000). They also claim that a test is valid if it does the job it was designed to. Further, the quality of any test highly depends on its validity.

Therefore, it is essential if the measurement is valid. If a measurement does not measure what it is purported to measure, it is useless. Hence, validity is the most crucial quality of any test in educational measurements (Oosterhof, 2001). As such, to make sure that the PSPT is measuring what it is supposed to measure, the two sets of the PSPT were tested for content validity and construct validity.

**i. Content Validity**

To evaluate the content-related evidence of validity, the researcher wrote down a description of what he intended to measure and gave the description along with the research instrument to two content experts. They looked at the description provided by the researcher and then went through the items of the tests.

In this study, the two experts checked and examined the content of the two sets of the ESL prototype speaking proficiency test items. The two experts were

experienced researchers and university lecturers in the field of TESOL and curriculum for over 10 years. They checked both sets of the prototype speaking proficiency test items based on a scale of 1 (least appropriate) to 5 (most appropriate). The appropriateness of the prototype speaking proficiency test questions depends on the ratings of the content experts. As such, agreement on items among the content experts is crucially significant to ensure the prototype speaking proficiency tests have content validity. The content experts examined the two sets (Set One and Set Two) with regard to the appropriateness of the questions and the vocabulary used in the two sets of the test to judge the content validity of the instrument which was used to collect the necessary data for the study. The two sets were of the same format and same level of difficulty. As such, the content experts assessed both sets of the PSPT (See Appendix D). It is necessary to mention that both sets of the prototype speaking proficiency tests were used in the pilot study so that different groups of students receive different sets of questions.

After the two sets of the prototype speaking proficiency tests were judged by the content experts, the researcher went through the items again and wrote down new items based on the content experts' judgments. The outcomes of the judgments received from the two experts indicated the appropriateness of questions of the prototype speaking proficiency tests. This tells that the content validity of the research instrument is appropriate and promising. Hence, the items of both sets were suitable. Table 4.6 illustrates the results of the content experts for the prototype speaking proficiency tests.

Table 4.6

*Ratings of Content Experts*

| Section | Item | Set One    |            | Average of agreement | Comment                                   | Set Two    |            | Average of agreement | Comment  |
|---------|------|------------|------------|----------------------|---|------------|------------|----------------------|--|
|         |      | Expert one | Expert two |                      |   | Expert one | Expert two |                      |  |
| A       | A 1  | 3          | 4          | 3.5                  | -   | 4          | 4          | 4                    |  |
|         | A 2  | 4          | 4          | 4                    | -   | 4          | 4          | 4                    |  |
|         | A 3  | 4          | 4          | 4                    | -   | 3          | 4          | 3.5                  |  |
|         | A 4  | 3          | 3          | 3                    | Change item because it is not appropriate | 3          | 4          | 3.5                  |  |
|         | A 5  | 3          | 4          | 3.5                  |   | 4          | 3          | 3.5                  |  |
|         | A 6  | 3          | 4          | 3.5                  | Add 'so' after 'you loved the movie'      | 4          | 4          | 4                    |  |
|         | A 7  | 4          | 4          | 4                    |   | 3          | 4          | 3.5                  | Add 'heavily and' after 'raining'                                    |
|         | A 8  | 3          | 3          | 3                    | Correct the grammatical mistakes          | 4          | 3          | 3.5                  | Add 'included in the exam' after 'topics'                            |
|         | A 9  | 3          | 4          | 3.5                  |   | 4          | 4          | 4                    |  |
|         | A 10 | 3          | 4          | 3.5                  |   | 4          | 4          | 4                    |  |
| B       |      | 4          | 5          | 4.5                  |   | 2          | 2          | 2                    | Change of stimulus due to inappropriateness and grammatical mistakes |
|         | B1   | 4          | 4          | 4                    |   | 4          | 4          | 4                    |  |
|         | B2   | 4          | 4          | 4                    |   | 4          | 4          | 4                    |  |
|         | B3   | 4          | 4          | 4                    |   | 4          | 4          | 4                    |  |
|         | B4   | 4          | 4          | 4                    |   | 4          | 4          | 4                    |  |
|         | B5   | 4          | 4          | 4                    |   | 4          | 4          | 4                    |  |
| C       |      | 5          | 5          | 5                    |   | 5          | 5          | 5                    |  |

Revisions and amendments were made after the comments from the content experts were collected. After the assessment of the two sets of the prototype speaking proficiency tests by the content experts, question 4 in Set One of Section A was changed due to inappropriateness. The word 'so' was added to question 6 in Section A. Likewise, based on the content experts' comments, the grammatical mistakes in question 8 of Section A were corrected. As for Set Two, 'heavily and' was added to question 7 of Section A. Similarly, 'included in the exam' was added after 'topics' in question 8 of Section A. In addition to Section A, the stimulus in Section B of Set Two



was changed into a better and more informative stimulus that can better measure respondents' speaking proficiency. After all the revisions were made, the prototype speaking proficiency tests were administered and piloted to the sample group of the study.

## **ii. Construct Validity**

In addition to content-related evidence of validity, construct-related evidence of validity was also considered to ensure the suitability of the tests for the intended purpose, which is measuring the English-speaking proficiency of undergraduates. A total number of 96 respondents consisting of mixed speaking proficiency (24 respondents from each of the four English courses) were taken to determine the construct validity of the prototype speaking proficiency tests. This is to determine to what extent the two tests represent the construct they are supposed to measure. Hence, a pilot study was conducted to determine the construct validity of the two sets (Set One and Set Two) of the PSPT, taking 48 undergraduate students for each of the two sets.

In this study, the researcher designed the PSPT in three different difficulty levels, ranging from elementary to advanced levels based on the classification of Littlewood (1981) in which speaking activities are categorized as pre-communicative and communicative learning activities. That is, speaking activities are classified from easy, intermediate to difficult activities. In pre-communicative activities, language learners are only expected to produce language which is acceptable (appropriate) rather than to communicate meaningfully. These types of activities are found in textbooks, such as different types of drills or question-and-answer practice. Whereas in communicative

activities, the learner has to activate his pre-communicative knowledge in order to use them to communicate meaningfully. Therefore, he/she is now involved in practicing communication, such as different types of group or social interactions and discussions (Littlewood, 1981).

This is rather supported by Lemus (2014) who classifies learning activities into fully-controlled, semi-controlled (partially-controlled) and free communicative activities. In fully-controlled activities, the answer is usually predictable by the teacher. Controlled activities allow learners to focus on the target language structure and the target language becomes familiar. As familiarity with the target language increases, the teacher may choose more semi-controlled activities in which there is a somewhat amount of freedom for the learner to maintain. Brainstorming activities, describing a picture or a flyer with information are examples of semi-controlled activities. Free communicative activities come last in a lesson whereby learners have extreme freedom in the language they study. In these kinds of activities, the teacher cannot predict what the students will be saying, for instance, a group discussion which is held by four or five individuals (Lemus, 2014). This is also in line with Mok's (2000) view of the classification of the difficulty level of test questions from easy to difficult which is 25% easy, 50% average and 25% difficult.

To judge if the prototype speaking proficiency tests have construct validity, a pilot study consisting of 96 respondents was conducted to show whether respondents at a higher level of English proficiency (LHP 456) scored higher than respondents at the lower level of English proficiency (LSP 403). Likewise, respondents at the (LSP

300) scored higher than respondents at the lower level of English proficiency (LMT 100) when answering the speaking proficiency tests. 24 mixed speaking ability respondents from each level of the four English proficiency levels from the School of Languages, Literacies, and Translation, USM were taken. To determine the construct validity of the speaking proficiency tests, the data collected from the respondents' speaking proficiency test performance was keyed into Excel to obtain the mean scores of the respondents' speaking proficiency based on their English proficiency level. Since the ESL PSPT consists of three sections, the mean scores of the respondents' performance of each section for each English proficiency level were taken. A comparison was made to realize the respondents' speaking proficiency at their different English proficiency levels and their performance on each section of the PSPT. Table 4.7 presents the mean scores of the respondents' speaking proficiency based on their English proficiency levels. The mean for all the four English proficiency levels was taken and standardized by transferring all the scores for each Section By dividing the total scores by the number of the respondents.

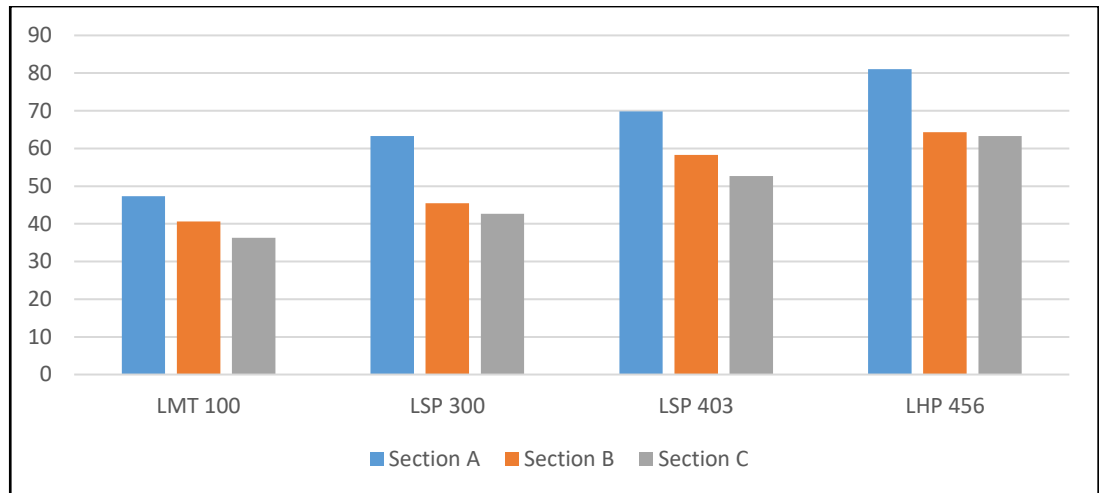
Table 4.7

*Comparison of the Mean Scores of Respondents' Speaking Proficiency Based on Their English Proficiency for Set One and Set Two*

| Set One          |           |           |           | Set Two          |           |           |           |
|------------------|-----------|-----------|-----------|------------------|-----------|-----------|-----------|
| English course   | Mean (%)  |           |           | English course   | Mean (%)  |           |           |
|                  | Section A | Section B | Section C |                  | Section A | Section B | Section C |
| Group A: LMT 100 | 47.3      | 40.6      | 36.3      | Group A: LMT 100 | 49.2      | 37.3      | 33.3      |
| Group B: LSP 300 | 63.3      | 45.5      | 42.6      | Group B: LSP 300 | 62.0      | 44.0      | 37.3      |
| Group C: LSP 403 | 69.8      | 58.3      | 52.6      | Group C: LSP 403 | 71.0      | 58.8      | 46.0      |
| Group D: LHP 456 | 81.0      | 64.3      | 63.3      | Group D: LHP 456 | 81.0      | 66.0      | 61.6      |

From the results of the pilot study, it was found that the content of the PSPT is appropriate for the target respondents who were tested. That is, students with higher band score performed better than students with lower band score. For instance, the mean score of Set One for LMT 100 respondents during responding Section A questions, was 47.3 out of 100, while the mean score for LSP 300 respondents was 63.3 out of 100, for LSP 403 respondents was 69.8 out of 100 and for LHP 456 respondents was 81 out of 100 during answering Section A questions. Likewise, the mean score of Set Two for LMT100 respondents during answering Section A questions was 49.2 out of 100, while the mean score for LSP 300 respondents was 62 out of 100, for LSP 403 respondents was 71 out of 100 and for LHP 456 respondents was 81 out of 100 during responding Section A questions. Hence, from this pilot study, it was noticed that respondents at a higher English proficiency level were able to respond the questions of the PSPT better than respondents at the lower English proficiency levels as it is obvious in the above table.

As it is evident from the above table, this gradual increase can also be seen during answering sections B and C of the PSPT based on the different English proficiency levels. This shows that respondents at the higher English proficiency levels scored higher when responding the questions of the PSPT. In this study, Section A questions were considered easy whereas Section B was considered intermediate and Section C was considered difficult. As such, it can be stated that the PSPT can differentiate between respondents at lower and higher English language proficiency levels as it can be seen in Figure 4.1 below.



*Figure 4.1: Relationship between respondents of English language proficiency levels and their speaking proficiency*

#### **b. Reliability of ESL Prototype Speaking Proficiency Test**

As a matter of fact, nothing could be more fundamental and essential to a good test than its reliability (Kubiszyn & Borich, 2000). The reliability of a test refers to the consistency with which it yields the same rank an individual taking the test several times. In other words, “a test (or any measuring instrument) is reliable if it consistently yields the same or nearly the same ranks over repeated administrations during which we would not expect the trait being measured to have changed” (Kubiszyn & Borich, 2000, p. 311). In addition, a research instrument is reliable if the scores achieved for a respondent on the same instrument at two different times or on two different parts of the same instrument are consistent (Fraenkel et al., 2012). By a simple definition, reliability is the degree to which an instrument measures something consistently.

According to Oosterhof (2001), a test or any measuring instrument can be reliable (measuring something consistently) without being valid (measuring what it is supposed

to measure). Furthermore, reliability can be an important quality of a test or any measuring instrument, whether that test is a written test, a performance assessment or an informal observation. As such, to ensure that the two sets of the PSPT are reliable, parallel-forms reliability, internal consistency using Cronbach's Alpha for the two sets were taken. Likewise, interrater reliability of test scores of the 96 respondents of the pilot study was taken into consideration to guarantee the reliability of the test scores.

#### **i. Parallel-Forms Reliability**

The parallel-forms method of judging reliability is commonly conducted in tests which have more than one or more versions and it judges if two versions are identical (Linn & Miller, 2005). Further, it is largely suitable to utilize the parallel-forms reliability if two or more forms or versions of a test are developed and used interchangeably (Oosterhof, 2001). Most important, the two versions are identical in every way except for the actual items included in both versions. This verifies that the different forms are measuring the same thing. Both versions of the test are administered to the same group of students (12 respondents) in close succession and the correlation between the two sets of scores is determined (Kubiszyn & Borich, 2010). Moreover, the parallel forms reliability can be tested through the Pearson Product Moment correlation analysis (Linn & Miller, 2005). A high coefficient would indicate a strong evidence of reliability that the two forms are measuring the same construct (Fraenkel & Walen, 2009, P. 156).

In the present study, the parallel forms reliability for the two sets (Set One and Set Two) of the prototype speaking proficiency tests was generated through the Pearson

Product-Moment Correlation Analysis over 12 respondents (See Appendix E). The correlation between both sets was positive and high ( $r = 0.81, p <.01$ ). That is, the performance of the respondents on both administrations were close to one another. As such, the two sets of the developed speaking test were identical and therefore can be used interchangeably.

## **ii. Internal Consistency Reliability**

Since the prototype speaking proficiency test in the current study is a test where there is no absolute answer (wrong or correct), the Cronbach's Alpha is considered to be more appropriate to determine the internal consistency reliability of the two sets of the prototype test developed. Cronbach's Alpha is computing the internal consistency to obtain the coefficient of reliability of the research instrument (Linn & Miller, 2005). Therefore, the ESL speaking proficiency test in this study was tested for internal consistency by generating the Cronbach's Alpha method of estimating reliability through SPSS.

This is also supported by Gay and Airasian (2003) who mentioned that if items have more than two scores (e.g., 0, 1, 2, 3), the Cronbach's Alpha is said to be used. On the contrary, Kuder-Richardson method of estimating of reliability is not used as it is more appropriate for estimating of the reliability of an instrument which consists of dichotomous items in which, are scored right versus wrong.

Before the actual data collection, a pilot study was conducted to assess the reliability of the ESL prototype speaking proficiency tests. Ninety-six undergraduate

learners of English taking different English courses based on their MUET at the School of Languages, Literacies and Translation at Universiti Sains Malaysia were involved in the pilot study. The reliability for Set One of the prototype speaking proficiency test was 0.83 (Appendix F) whereas the reliability coefficient for Set Two of the prototype speaking proficiency test through Cronbach's Alpha was 0.824 (Appendix G). The reliability of the prototype speaking proficiency tests has been tabulated in Table 4.8 which shows a strong reliability of the prototype speaking proficiency tests.

Table 4.8

*Reliability of the Prototype Standardized Speaking Proficiency Test*

| Reliability Statistics           |            |                                  |            |
|----------------------------------|------------|----------------------------------|------------|
| Set One                          |            | Set Two                          |            |
| Cronbach's Alpha                 | N of Items | Cronbach's Alpha                 | N of Items |
| .83                              | 12         | .824                             | 12         |
| Item-Total Statistics            |            | Item-Total Statistics            |            |
| Cronbach's Alpha if Item Deleted |            | Cronbach's Alpha if Item Deleted |            |
| A1                               | .814       | A1                               | .819       |
| A2                               | .830       | A2                               | .804       |
| A3                               | .812       | A3                               | .795       |
| A4                               | .828       | A4                               | .797       |
| A5                               | .812       | A5                               | .810       |
| A6                               | .810       | A6                               | .814       |
| A7                               | .812       | A7                               | .824       |
| A8                               | .814       | A8                               | .798       |
| A9                               | .805       | A9                               | .804       |
| A10                              | .827       | A10                              | .809       |
| B                                | .821       | B                                | .820       |
| C                                | .819       | C                                | .831       |

Note: Section A: 10 items, Section B: 1 item and Section C: 1 item

**iii. Interrater Reliability of Tests Scores (Pilot Study: 96 Respondents)**

In addition to parallel-forms reliability and internal consistency of the PSPT, inter-rater reliability was also conducted. In order for scores reported within a language modality (speaking, for example) based on analytical rating scales to be useful for an



intended purpose, ratings given by raters should be correct and reliable and decisions made based on such ratings should be dependable for analytical ratings (Sawaki, 2007).

When humans are a part of any assessment, especially in assessing speaking, the reliability and the consistency of the results of the raters should be considered and when assessing respondents' productive skills (speaking and writing) or subjectively scored assessments, raters are of paramount significance to take into consideration to guarantee the assessment and scoring procedure (Brown, 2005, cited in Sak, 2008).

Further to that, during assessing speaking, it is advisable to have more than one rater (examiner) in order to ensure the scores given to the respondents are correct and reliable. As such, for the purpose of the pilot study, two examiners have studied how the speaking tasks are scored based on the speaking rubrics utilized in the present study. Therefore, two raters scored the respondents' speaking performance independently. The raters were experienced English language and TESOL lecturers. The raters have been working in the field of teaching the English language for several years at Salahaddin University-Erbil in Kurdistan region who were Ph.D. graduates of Universiti Sains Malaysia.

Inter-rater reliability is evaluated by examining the scores of two or more raters given independently and calculating the correlation coefficient between the two sets of the scores using intraclass correlation coefficient in SPSS (Linn & Miller, 2005). The results of the correlation coefficient would tell the consistency of the scores given by

the raters (Sak, 2008). Inter-rater reliability measures the consistency between different raters or examiners which is the degree of correlation between two or more raters for the purpose of determining whether or not they score respondents' performance similarly (Hadley & Mort, 1999).

The assessment of interrater consistency is essential to ensure that students do not achieve high scores as the result of rater generosity while others receive low scores as the result of rater stringency (Linn & Miller, 2005). Therefore, to make sure the reliability of the scores given to the respondents by the two raters, a test of inter-rater reliability was used in this present study to test how consistent the two raters are to one another.

Any value above 0.70 is acceptable, above 0.80 is good and above 0.90 is excellent. As the prototype tests consist of three sections, the two raters scored each section independently. As such, the assessment of interrater consistency for scoring of each section of the PSPT was conducted separately. The interrater reliability of the scores for the three sections (sections A, B, and C) of the PSPT by determining the average measures were respectively 0.80 (See Appendix H), 0.79 (See Appendix I), 0.82 (See Appendix J). The interrater reliability given by the two raters is tabulated in Table 4.9 which shows that there is a high interrater reliability between the two raters.

Table 4.9

*Interrater Reliability of the Prototype Speaking Proficiency Tests*

| Section | Average Measures (Interrater Reliability) |
|---------|---|
| A       | 0.80                                      |
| B       | 0.79                                      |
| C       | 0.82                                      |

#### **4.2.4 Time Allocated for the ESL Speaking Proficiency Test**

Generally speaking, there is no a specific time for speaking tests to be completed. Time allocated for speaking tests totally depends on test specifications. For example, an IELTS speaking test takes up 11 to 14 minutes whereas the speaking section of the TOEIC takes approximately 20 minutes to complete. In this study, a pilot study was conducted to find out the time taken to complete the Prototype Speaking Proficiency Test of 96 respondents. The respondents consisted of different English language proficiency levels. Therefore, the time needed to complete the PSPT was expected to be different among the selected respondents sat for the tests in the pilot study. As such, it was necessary to record the time taken by respondents of different English language levels. It was observed that LMT 100 and LSP 300 respondents who are low-performing respondents took 12 minutes to answer Section A, 2 minutes to answer Section B and 7 minutes to answer Section C whereas LSP403 and LHP 456 respondents who are higher performing respondents took 10 minutes to respond Section A, 3 minutes to respond Section B and 9 minutes to respond Section C. Finally, the decision for the amount of time allocated to complete the speaking test was made based on the average time taken by respondents from different English proficiency levels. That is, respondents are given 11 minutes to complete Section A, 2.5 minutes to complete Section B and 8 minutes to complete Section C. In total, 21.5 minutes will be given for any respondent to respond the PSPT. Table 4.10 shows the time allocated for the ESL speaking proficiency test.

Table 4.10

*Pilot Test to Find Out Time Taken to Complete PSPT*

| English Proficiency Level         | Time taken  |             |           |
|-----------------------------------|-------------|-------------|-----------|
|                                   | Section A   | Section B   | Section C |
| LMT 100                           | 12 minutes  | 2 minutes   | 7 minutes |
| LSP 300                           | 12 minutes  | 2 minutes   | 7 minutes |
| LSP 403                           | 10 minutes  | 3 minutes   | 9 minutes |
| LHP 456                           | 10 minutes  | 3 minutes   | 9 minutes |
| Average time taken                | 11 minutes  | 2.5 minutes | 8 minutes |
| Total time taken to complete test | 21.5 minute |             |           |

### 4.3 Development of ESL Speaking Proficiency Test

Before the ESL prototype speaking proficiency tests were used as the ESL speaking proficiency tests, revisions and some changes were made based on the feedback given by the content experts as it is explained previously in content validity section. In addition to that, the results of the piloted PSPT revealed that the two sets of the PSPT do not show any problem with regard to the following issues:

- Content validity is high.
- Reliability is high.
- The two sets of the PSPT are able to distinguish between higher and lower English proficiency levels. Hence, the tests are construct valid.
- The difficulty of sections is appropriate for different English proficiency levels.

However, there were some grammatical and spelling mistakes as well as inconsistency in the spacing of sentences and mistakes and they were duly corrected. Based on the content experts' feedback, the content of the PSPT for sections A, B and C were carefully selected. With these corrections made to the ESL prototype speaking proficiency tests, the ESL Speaking Proficiency Tests were developed (See Appendix

K). It is worth mentioning that two sets of the ESL SPT were built as a battery of tests for the questions bank. Both sets of the PSPT were piloted and calibrated so that both sets are of equal difficulty and accuracy.

The ESL Speaking Proficiency tests comprise three different difficulty sections namely; Section A, B and C. In Section A, respondents respond to 10 questions whereas, in Section B, respondents are given a newspaper cutting (a stimulus) whereby they read and are then asked 5 questions related to the stimulus. In the last section, respondents engage in a group discussion. The time allocated to complete the three sections of the ESL Speaking Proficiency Test is 21.5 minutes. The score achieved by the respondents is determined by their speaking proficiency on the three sections appear on the ESL SPT. Cut scores for the different performance bands (Band One to Band Four) were established to indicate different degrees of speaking proficiency. It is important to note that Band One shows that the student is at a very limited speaking proficiency, while Band Four indicates that the student is at a high level in speaking proficiency.

#### **4.4 Summary**

This chapter specifically introduced the process of the development of the ESL Speaking Proficiency Test. In other words, the chapter illustrated the manner in which the research instrument was developed. The SPT was specifically developed to accurately assess the undergraduates' speaking proficiency in a way that it can help language lecturers to identify their students' strengths and weaknesses. Based on the students' current level, language lecturers will be able to tailor their teaching materials

needed to help the students enhance their speaking proficiency. The reason behind developing the SPT in the present study is that the developed test consists of three different difficulty question levels; ranging from elementary to advanced levels. Therefore, the SPT is able to tell enough information about students' speaking problems which can be used for the purpose of improvement by lecturers. The findings and results of this study are revealed in the next chapter.

## **CHAPTER 5**

### **DATA ANALYSIS AND FINDINGS**

#### **5.1 INTRODUCTION**

This chapter presents the results of the current mixed-methods research on identifying and profiling undergraduates' speaking proficiency. The chapter answers the research questions 2, 3 and 4. The gathered data are analyzed to develop speaking proficiency descriptors as well as identifying and profiling the undergraduates' English-speaking proficiency. The objectives and research questions of the present study were specifically developed to meet the purpose of identifying and profiling the undergraduates' English-speaking proficiency. In order to identify and profile the undergraduates' speaking proficiency, a number of stages were conducted and discussed below. The stages are chronological as each single stage would lead to the next stage.

The write up in the current chapter is divided into three main sections to facilitate the comprehension of the chapter. The first section illustrates establishing the undergraduates' speaking proficiency descriptors based on the respondents' speaking proficiency and their responses on the structured interview questions. The second section of the chapter deals with identifying the undergraduates' English -speaking proficiency which comprises respondents' general speaking proficiency followed by the overall speaking proficiency performance based on gender, ethnicity, university,

and school. Last, the third section describes the profiles of the Superior Performers, Advanced Performers, Intermediate Performers and Novice Performers. In this chapter, the findings are reported in terms of the research questions (Research questions 2, 3 and 4) raised earlier in this study.

## **5.2 Developing a Set of Descriptors to Identify the Undergraduates' ESL**

### **Speaking Proficiency**

This section intends to answer research question 2 posed in the introductory chapter. Generally, the section deals with developing the speaking proficiency descriptors for undergraduates.

Research Question 2:

How are the speaking proficiency descriptors developed to identify the undergraduates' speaking proficiency performance?

- a. What are the most suitable cut scores for the performance bands?
- b. What are the speaking skills undergraduates have acquired?

### **5.2.1 Determining the Cut Scores for the Bands**

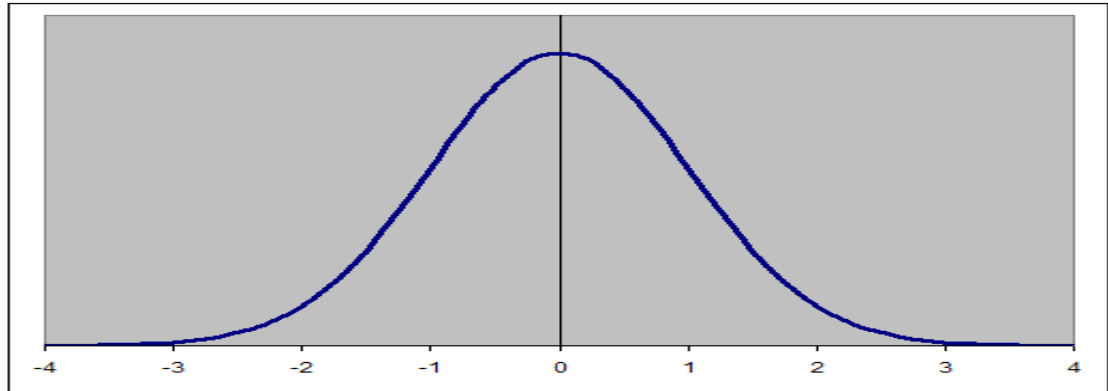
To answer research question 2, the researcher of the study first had to determine the cut scores for the bands. After the ESL prototype speaking proficiency test was administered and the scores were gathered, the scores were categorized into bands in order to determine the cut scores for the performance bands by utilizing the z-scores.



In order to compare raw scores in different distributions, raw scores need to be converted into z-scores (Birjandi et al., 2004). As a matter of fact, a score by itself does not give much. For example, if a student obtains a score of 80 on a test, this might be good if other students obtain 50, but less so if many students obtain 95. Put simply, a score is meaningful only relative to the mean of the sample.

A z-score is also called a standard score and it has a mean of 0 and a variance of 1. A z-score shows the precise location of a data value within a distribution. It is obtained by converting a raw score into a signed number (+ or -) and this signifies whether the score is above the mean (positive) or below the mean (negative). The signed number tells the distance between the score and the mean by counting the number of standard deviations and the result will be the standard normal distribution (Gravetter & Wallnau, 2013).

According to Gravetter and Wallnau (2013), a z-score of +1.00 corresponds to a position 1 standard deviation above the mean and a z-score of +2.00 is placed 2 standard deviations above the mean. A z-score of -1.00 corresponds to a place precisely 1 standard deviation below the mean and a z-score of -2.00 is exactly located 2 standard deviations below the mean. Figure 5.1 illustrates the relationship between z-score and location in a standard distribution.



*Figure 5.1:* The normal curve relationship between z-score and location in a standard distribution

After the prototype speaking proficiency test was administered and the scores were collected, the scores were broken up into groups in order to determine the performance bands by using the z-scores (standard scores). “Standard scores compare a respondent’s performance to that of other respondents at the same grade level” (Kubiszyn & Borich, 2000, p. 357). They also claim that the best type of scores is the z-score and all other types of standard scores are derived from z-scores. Z-scores are also crucial since they permit testers to make comparisons of scores within respondents.

The cut scores for the performance bands were developed based on the respondents’ performance on the prototype speaking proficiency test. The respondents for the prototype speaking proficiency test was USM undergraduate students from MUET Band Two to MUET Band Six. Therefore, the performance bands are based on the performance of those respondents who score Band Two to Band Six in MUET. It has to be noted that for the different proficiency levels (from MUET Band Two to MUET Band Six), the cut scores for each level will be different because the descriptors of respondents’ speaking proficiency level will be different.

To establish the cut scores for the bands (Band One to Band Four), a pilot study was conducted. Ninety-six respondents from different English language proficiency levels participated in the prototype speaking proficiency tests. The 96 respondents were from different English language proficiency levels (MUET Band Two to MUET Band Six) taking English courses at the School of Languages, Literacies and Translation, at Universiti Sains Malaysia.

In order to construct the cut scores for the bands, a comparison of the speaking proficiency of the high performers and low performers (from MUET Band Two to MUET band Six) was conducted. The data derived from descriptive analysis of the results obtained from the performance of the 96 respondents participated in the pilot study on the prototype speaking proficiency test has been shown in Table 5.1. Descriptive statistics presented in Table 5.1 include the number of the respondents, the minimum and maximum score achieved by the respondents, mean and standard deviation. To calculate the cut scores of the prototype speaking proficiency test, the mean and standard deviation were transformed into z-scores to categorize the respondents into the 4 bands.

Table 5.1

*Descriptive Statistics for Pilot Study*

|                           | No | Min (%) | Max (%) | Mean  | SD   |
|---------------------------|----|---------|---------|-------|------|
| Speaking Proficiency Test | 96 | 35      | 82      | 54.12 | 11.7 |

The cut scores for Band One to Band Four were developed based on the z-score which is the most standard type of scores. Z-scores provide a convenient way to

transform a raw score into a standardized score that provides useful information about how far the raw score relates to the mean, e.g., whether the score is above or below the mean and how far from the mean is the score for example, how many standard deviations the raw score is from the mean (Gravetter & Wallnau, 2013). A z-score of 0.00 shows that a respondent scored zero standard deviation (sd). That is, the respondent scored exactly the same as the mean. When a score is exactly on the mean, it corresponds to the z-score of 0.00 standard deviation and negative z-scores indicate that the respondent scored below the mean, whereas positive z-scores indicate that the respondent scored above the mean.

The mean and the standard deviation obtained from the performance of the respondents on the prototype speaking proficiency test were respectively 54.12 and 11.7. Hence, the z-score of +2.00 standard deviations which is located at a position exactly two standard deviations from the mean (54.12) would be 77.52 and the z-score of -2.00 standard deviations which corresponds to a position exactly two standard deviations below the mean would be 30.72. To simplify the calculation, the obtained z-scores were then rounded up to the nearest whole numbers. Therefore, the mean score was rounded up to 54 and the standard deviation was rounded up to 12. As such, the z-score of +2.00 standard deviations was rounded up to 78 because it was two standard deviations above the mean and the z-score of -2.00 standard deviations was rounded up to 30 because it was two standard deviations below the mean. Figure 5.2 illustrates how the cut scores were developed based on z-scores.

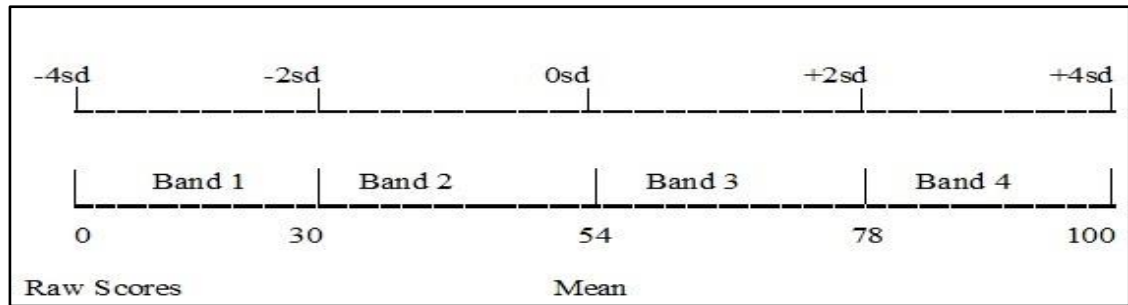


Figure 5.2: Cut scores based on z-scores

As it can be seen in the above figure, respondents were categorized into the four performance bands e.g., Band One to Band Four according to their speaking proficiency on the prototype speaking proficiency test. The z-score for Band Four should be between +2.00 sd to +4.00 sd showing that the respondents scored between +2.00 sd to +4.00 sd above the mean score. Likewise, the z-score for Band Three should be between 0.00 sd to +2.00 sd showing that the respondents scored between 0.00 sd to +2.00 sd above the mean score. In contrast, the z-score for Band Two should be between 0.00 sd to -2.00 sd showing that the respondents scored between 0.00 sd to -2.00 sd below the mean score. Similarly, the z-score for Band One should be between -2.00 sd to -4.00 sd showing that the respondents scored between -2.00 sd to -4.00 sd below the mean score. The scores for the different bands can be calculated based on z-score as it is shown in Table 5.2.

Table 5.2

*Establishing Scores for Bands*

| Bands      | Raw Scores       |
|------------|------------------|
| Band One   | Between (0-29)   |
| Band Two   | Between (30-53)  |
| Band Three | Between (54-77)  |
| Band Four  | Between (78-100) |

The respondents could be categorized into 4 bands based on their speaking proficiency. By calculating the raw scores from the z-score, it was found that for the z-score of 0.00 sd, the raw score was 54. Consequently, the range of scores for Band One should be within 0-29 and the range of scores for Band Two should be within 30-53. In contrast, the range of scores for Band Three should be within 54-77 and the range of scores for Band Four should be within the range of 78-100.

In general, only 4 respondents out of the total 96 respondents participated in the pilot study (4.16 percent of respondents) was categorized as Band Four performers, 48 respondents (50 percent of respondents) were categorized as Band Three performers, 44 respondents (45.84 percent of respondents) were categorized as Band Two performers and none of the respondents was categorized as Band One performers. Table 5.3 illustrates the classification of cut scores for each performance band with the number and percentage of respondents participating in the pilot study classified in each performance band comprising Band One to Band Four.

Table 5.3

*Cut scores and Categories of Performers*

| Categories of Performers | Cut scores | Number of Respondents | Percentage of Respondents |
|--------------------------|------------|-----------------------|---------------------------|
| Band One                 | (0-29)     | 0                     | 0%                        |
| Band Two                 | (30-53)    | 44                    | 45.84%                    |
| Band Three               | (54-77)    | 48                    | 50%                       |
| Band Four                | (78-100)   | 4                     | 4.16%                     |

## 5.2.2 Identifying Speaking Skills Undergraduates have Acquired

This section deals with the results of the respondents' speaking proficiency in different sub-skills at different sections of the speaking proficiency test for the purpose of developing the speaking proficiency descriptors. For that, the findings are analyzed to find out how the respondents perform in the different sub-skills of the speaking proficiency test. To answer research question 2, the raw scores obtained from the ESL prototype speaking proficiency test were analyzed through z-scores, by using the standard deviation and mean of the respondents' speaking proficiency performance on the prototype speaking proficiency test.

The mean score of the 96 respondents' speaking proficiency performance participated in the pilot study at each component of each section of the prototype speaking proficiency test has been presented to find out how the respondents perform in each section of the test as it is presented in Table 5.4 below.

Table 5.4

*Mean Score of Respondents' Speaking Proficiency Performance at Different Sections of PSPT*

| Difficulty Levels       | Sub-skills of Speaking    | Bands (%) |      |      |      |
|-------------------------|---------------------------|-----------|------|------|------|
|                         |                           | 1         | 2    | 3    | 4    |
| Section A: Elementary   | Appropriateness           | -         | 47.2 | 72.0 | 88.0 |
|                         | Grammar                   | -         | 52.6 | 76.3 | 89.0 |
| Section B: Intermediate | Appropriateness           | -         | 35.0 | 57.4 | 81.0 |
|                         | Communicative Ability     | -         | 35.4 | 57.1 | 79.0 |
|                         | Fluency and Pronunciation | -         | 46.3 | 62.8 | 84.5 |
|                         | Grammar and Vocabulary    | -         | 47.9 | 64.0 | 84.0 |
| Section C: Advanced     | Appropriateness           | -         | 37.4 | 55.0 | 80.0 |
|                         | Communicative Ability     | -         | 40.4 | 55.4 | 80.0 |
|                         | Managing a Discussion     | -         | 38.0 | 54.2 | 79.0 |
|                         | Fluency and Pronunciation | -         | 37.2 | 54.0 | 80.0 |
|                         | Grammar and Vocabulary    | -         | 38.6 | 54.8 | 80.0 |

The respondents were categorized into 4 bands based on their speaking proficiency performance and z-score. A total number of 96 respondents involved in the prototype speaking proficiency test and the calculation of the percentages was based on the number of respondents in each band as it is depicted in Table 5.5 below.

Table 5.5

*Frequency and Percentages of Respondents in Performance Bands*

| Band  | Respondents (f) | Percentage (%) |
|-------|-----------------|----------------|
| One   | -               | 0              |
| Two   | 44              | 45.84          |
| Three | 48              | 50             |
| Four  | 4               | 4.16           |
| Total | 96              |                |

In this section, the analysis of the respondents' speaking proficiency performance who obtained various scores in the different sub-skills at each section of the prototype speaking proficiency test is illustrated. The mean score of the respondents' performance has been taken into account to derive the terms used in this study. The analysis was tabulated to show the respondents' speaking proficiency performance on the prototype speaking proficiency test as shown previously in Table 5.4.

Basically, this section explains the development of the speaking proficiency descriptors. Therefore, the data in Table 5.4 were specifically studied to develop the speaking proficiency descriptors. The mean score of the respondents' performance who obtained different scores in the different sub-skills of speaking at each section of the PSPT in each band was analyzed and then used to derive the terms to describe the



respondents' speaking proficiency. Hence, great effort has been taken to make sure that the respondents' speaking proficiency is described specifically and accurately.

To arrive at the terms used in the speaking proficiency descriptors describing the respondents' specific speaking proficiency such as 'hardly', 'has difficulty', 'satisfactorily' and 'very well', the range of the mean score of the respondents' performance in different sections of the speaking test was considered to determine the suitable terms to be utilized to describe the respondents' speaking proficiency in Band One through Band Four. Based on the range of the mean score of the respondents' speaking proficiency performance, it indicates that respondents whose mean score is between 78 to 100 perform very well, respondents whose mean score is between 54 to 77 perform 'satisfactorily', while respondents whose mean score is between the range of 30 to 53 have difficulty in performing speaking proficiency on the prototype speaking proficiency test. Last, respondents whose mean score is between 0 to 29 can hardly speak English. In doing so, "very well" is used to describe respondents' performance whose range of the mean of respondents' performance is within 78 to 100 which is in line with the cut scores of the respondents in Band Four which is within 78 to 100. Likewise, "satisfactorily" is used to describe respondents' performance whose range of the mean of respondents' performance is within 54 to 77. Similarly, "has difficulty" is used to describe respondents' performance whose range of the mean of respondents' performance is within 30 to 53. Last, "hardly" is used to describe respondents' performance whose range of the mean of respondents' performance is within 0 to 29 which is in line with the cut scores of the respondents in Band One which

is within 0 to 29. Terms used in the speaking proficiency descriptors to describe the respondents' specific speaking proficiency have been depicted in Table 5.6.

Table 5.6

*Terms Used in Speaking Proficiency Descriptors*

| Range of the Mean of Respondents' performance | Terms Used in the Speaking Descriptors Based on the Range of the Mean Score |
|---|---|
| 78-100  | very well   |
| 54-77   | satisfactorily  |
| 30-53   | has difficulty  |
| 0-29  | hardly  |

The speaking proficiency descriptors would indicate that respondents whose mean score is between 78 to 100 'are able to answer in given context for the intended purpose and audience very well and possess a very good command of form and function' when answering Section A of the test. Similarly, by looking at Table 5.4, the findings show that respondents whose mean score is between 78 to 100 'can perform on grammar very well' when answering Section A of the speaking test which is 89%. Therefore, these respondents' performance was classified as 'very well' e.g., their performance in grammar.

The speaking proficiency descriptors also show that respondents whose mean score is between 54 to 77 'are able to answer in given context for the intended purpose and audience satisfactorily and possess a satisfactory command of form and function' when answering Section A of the test. By looking at Table 5.4, the findings reveal that respondents whose mean score is between 54 to 77 'can perform on grammar satisfactorily' when answering Section A of the speaking test which is 76.3%.

Therefore, these respondents' performance was classified as 'satisfactorily' e.g., their performance in grammar.

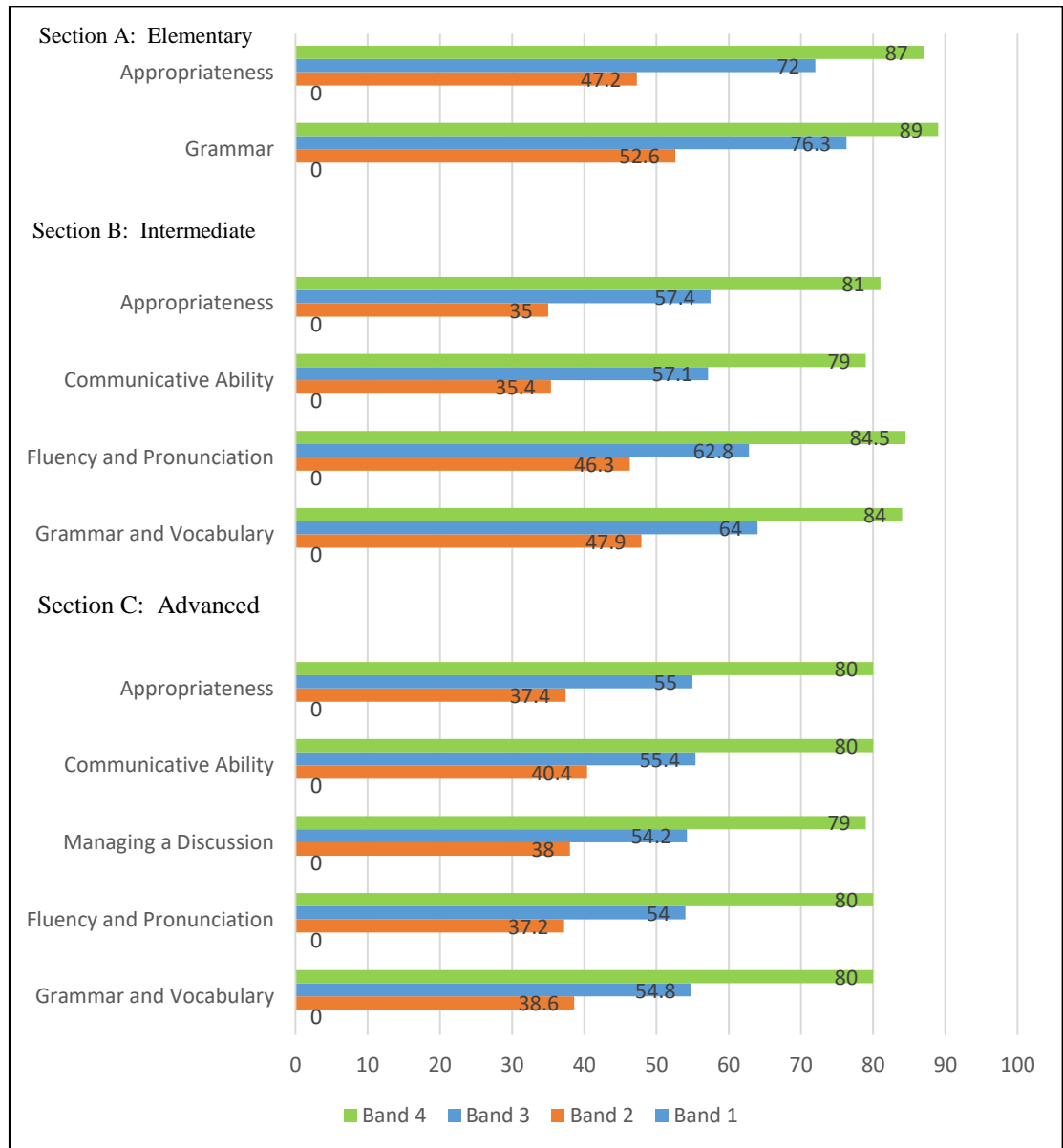
In parallel with that, another example to clarify the use of 'has difficulty' can be seen in Table 5.6. By referring to Table 5.4, it was clear that respondents whose mean score is between the range of 30 to 53 'have difficulty in answering questions meaningfully as well as demonstrating well in conveying their message' in communicative ability when answering Section B of the test which is 35.4%. Therefore, these respondents' performance was classified as 'has difficulty' e.g., 'has difficulty in their performance in communicative ability'.

Last, explaining the use of 'hardly' in Band One could be seen in Table 5.6. As such, from Table 5.4, it was evident that respondents whose mean score is between the range of 0 to 29 'can hardly answer questions meaningfully as well as demonstrating well in conveying their message' in communicative ability when answering Section B of the test. As a result, these respondents' performance was classified as 'hardly' e.g., 'can hardly perform in communicative ability'.

In this study, it is necessary to describe the development of the speaking proficiency descriptors since the speaking proficiency performance bands on their own are not specific to show the respondents' specific speaking proficiency. The specific speaking proficiency performance descriptors were developed based on the respondents' speaking performance on the prototype speaking proficiency test as well as the respondents' responses on the structured interview questions pertaining to the

prototype speaking test. As such, the speaking descriptors were developed based their performance in each component in the different sections of the prototype speaking proficiency test.

In this study, the quantitative analysis of the respondents based on their performance in the prototype speaking proficiency test was conducted to find out the respondents' speaking proficiency performance. To be precise, the main purpose of this analysis was to realize the respondents' speaking proficiency at each sub-speaking skill of the three different sections of the PSPT. Their speaking proficiency performance was analyzed and interpreted based on quantitative data. The graphical presentation of the speaking proficiency performance of the respondents in each different speaking skill at different sections of the speaking proficiency test has been depicted in Figure 5.3.



*Figure 5.3: Respondents' acquired speaking skills*

As it is evident in Figure 5.3, the respondents in Band Four can perform very well on sections A, B and C of the speaking proficiency test. For example, the respondents in Band Four appeared to score the average mark of 88 out of 100 in providing appropriate answers to questions in given context for the intended purpose. The respondents can answer questions appropriately in given context for the intended

purpose very well. Similarly, respondents in Band Four are able to score the average mark of 89 out of 100 in grammar. They can use accurate and correct grammar very well during speaking.

In a same vein, the same respondents scored the average score of 84.5 out of 100 in fluency and pronunciation. That is, the respondents are able to speak fluently and smoothly very well. They can speak without pausing for too long very well. Besides, they can pronounce individual words correctly very well. They are also able to express stress and intonation very well. Furthermore, the respondents in Band Four can also perform very well in grammar and vocabulary, communicative ability and managing discussions. That is, they are capable of using correct grammar and a wide range of correct grammatical sentences very well. They are also able to use a wide range of vocabulary effectively and appropriately very well. Besides, Band Four respondents can communicate and demonstrate very well in conveying their message. Additionally, the respondents can develop and maintain a discussion and manage it very well in terms of initiating, interrupting, closing and so forth.

By referring to Figure 5.3, it is clear that the respondents in Band Three were able to perform satisfactorily in speaking skills in the different sections of the speaking test. For instance, Band Three respondents were able to score the average mark of 72 out of 100 in providing appropriate answers to questions in given context for the intended purpose. The respondents are able to answer questions appropriately in given context for the intended purpose satisfactorily. In the same way, respondents in Band Three could score the average mark of 76 out of 100 in grammar in Section A which

shows that they are able to use accurate and correct grammar satisfactorily during speaking.

In the same vein, the same respondents scored the average score of 62.8 out of 100 in fluency and pronunciation. The respondents are able to speak fluently and smoothly satisfactorily. They can speak without pausing for too long satisfactorily. Besides, they can pronounce words correctly and express stress as well as intonation satisfactorily. Furthermore, the respondents in Band Three can also perform satisfactorily in grammar and vocabulary, communicative ability and managing discussions. They are capable of using correct grammar and a wide range of correct grammatical sentences satisfactorily. They are also able to use a wide range of vocabulary effectively and appropriately in a satisfactory way. Besides, Band Three respondents can communicate and demonstrate satisfactorily in conveying their message. Additionally, the respondents can develop and maintain a discussion and manage it satisfactorily in terms of initiating, interrupting and so forth.

By looking at Figure 5.3, it is seen that respondents in Band Two have difficulty in speaking skills in the different sections of the speaking test. For instance, Band Two respondents could score the average mark of 47 out of 100 in providing appropriate answers to questions in given context for the intended purpose. That shows that the respondents have difficulty in answering questions appropriately in given context for the intended purpose. Similarly, respondents in Band Two can score the average mark of 52 out of 100 in grammar in Section A. Obviously, they have difficulty in using accurate and correct grammar during speaking English.

Moreover, the same respondents scored the average score of 46.3 out of 100 in fluency and pronunciation. That indicates that the respondents have difficulty to speak fluently and smoothly. They have difficulty to speak without pausing for too long. In addition, they also have difficulty in pronouncing individual words. The Band Two respondents have also difficulty in expressing stress and intonation correctly. Further, the respondents in Band Two also have difficulty to perform well in grammar and vocabulary, communicative ability and managing discussions. That is, they have difficulty in using correct grammar and a wide range of correct grammatical sentences. They also have difficulty to use a wide range of vocabulary effectively and appropriately. Additionally, Band Two respondents have difficulty to communicate and demonstrate well in conveying their message. Furthermore, the respondents have difficulty to develop and maintain a discussion and manage it in terms of initiating, interrupting and so forth.

Based on the cut scores developed in the current study, any respondent scores between 0-29 is categorized as Band One. By looking at the speaking proficiency performance bands developed specifically for the purpose of the current study, we can see that respondents in Band One can hardly speak English. They can hardly answer questions, conduct interviews as well as engage in group discussions. As the speaking proficiency performance bands show, they can hardly answer questions in given context for the intended purpose. Likewise, they can hardly provide accurate and correct grammatical sentences during speaking. As this takes place, they can hardly speak fluently and smoothly without pausing for too long. Having said that,



respondents in Band One can hardly perform in pronunciation, vocabulary, communicative ability and managing a discussion.

They can hardly pronounce individual words correctly as well as express intonation and stress. The respondents can hardly use a wide range of vocabulary effectively. In addition, Band One respondents can hardly communicate and demonstrate well in conveying their message. The respondents can hardly develop and maintain a discussion and manage it in terms of initiating, interrupting and so forth. With that, we can notice that respondents in Band One can hardly speak English.

In developing the speaking proficiency descriptors, qualitative data was also used. The qualitative data came through structured interview questions to provide a deeper insight into the quantitative findings, in particular, to support strengthening the development of the speaking proficiency descriptors. Therefore, the findings of the qualitative data regarding the respondents' responses to the speaking proficiency test and their speaking abilities derived from the interview transcripts.

The findings are classified into three major categories namely, clarity of sections A, B, and C of the test, difficulties encountered by the respondents during the test and preference of types of speaking questions. It should be noted that a total number of 9 respondents from different bands were selected to participate in collecting the qualitative data. It is worth mentioning that verbatim procedure was followed in transcribing the interviews.

The qualitative data were collected via structured interview questions. Three respondents from each band (Band Two, Band Three and Band Four) who participated in the PSPT were selected to examine precisely what the respondents are able to do and what they are not able to do. However, the qualitative data did not include Band One as none of the respondents of the study was categorized as Band One. Their responses were tabulated and later analyzed and interpreted.

Findings regarding the clarity of the three different sections of the speaking proficiency test have been presented in Table 5.7. The respondents' responses elicited from the structured interview excerpts.

Table 5.7

*Clarity of Sections A, B, and C*

| Item   | Interviewees' Responses   | Themes   |
|--|---|--|
| Do you understand the questions in Section A, B, and C?<br>Yes/no...Elaborate. |   |  |
| Band Two   | <ul style="list-style-type: none"> <li>• Yes, the questions in all sections are understandable.</li> <li>• Yes, the questions were easy to understand. This question is about the same as I ever answer questions during a test in school first.</li> <li>• Yes, I do understand the question in those three Section Because the words used was familiar to me and easy to understand.</li> </ul>   | <ul style="list-style-type: none"> <li>• Understandability of questions</li> </ul> |
| Band Three   | <ul style="list-style-type: none"> <li>• Yes, in Section A, we are given questions about and situation and have to give the most suitable answer according to the situation. For Section B, we did the job interview session by answering 5 common interview questions and for the last session we did group discussion where we are given a situation to discuss about it.</li> <li>• Yes, for Section A, I have to answer 10 question or situation given. While, in Section B, I have to answer a job interview based on flier where there will be 5 questions that need to answer. Then, Section C I have to attend a group discussion with my fellow friend discuss about certain subject.</li> <li>• Yes, I understood the questions. The questions in Section A is clear. Meanwhile for Section B, the questions asked are relevant and most appropriate to ask in job interview. For Section C, the question is clear and we could discuss and come to a conclusion at the end of the discussion.</li> </ul> | <ul style="list-style-type: none"> <li>• Clarity of sections A, B and C</li> </ul> |
| Band Four  | <ul style="list-style-type: none"> <li>• Yes, I do. The questions are clear, so I could imagine the situation and also give proper answer based on my view.</li> <li>• Yes, I am. The question is very straight forward and easy to understand. For question A, the question is related to daily activities, so it was not that difficult to answer. For question B, the interview is quite simple and I have done an interview before. For question C, I just need to act spontaneous and enjoy the play role as what it is in my mind.</li> <li>• Yes, the instructions were given clearly and I had fun answering them.</li> </ul>   | <ul style="list-style-type: none"> <li>• Understandability of questions</li> </ul> |

In general, it can be seen that the respondents agreed that the questions of all the three different sections of the speaking proficiency test were clear and understandable. Although the respondents were from different language proficiency levels, but as they stated, all the questions were clear for all the respondents of lower to higher levels of English language proficiency. For example, a respondent from the lower English language proficiency said:

... Yes, the questions were easy to understand. This question is about the same as I ever answer questions during a test in school first ...(B2-2-22/10/2016).

Likewise, other respondents from higher English language proficiency found that the questions were clear and they related to daily life activities. For example, one of the respondents from higher English language proficiency stated:

...Yes, I am. The question is very straight forward and easy to understand. For question A, the question is related to daily activities, so it was not that difficult to answer. For question B, the interview is quite simple and I have done an interview before. For question C, I just need to act spontaneous and enjoy the play role as what it is in my mind...(B4-2-29/10/2016).

Similarly, findings regarding the difficulties the respondents encountered during taking the speaking test have been summarized in Table 5.8. The findings regarding the difficulties encountered by the respondents obtained through the structured interviews and are followed by examples from the interview excerpts.

Table 5.8

*Difficulties Encountered During the Test*

| Item   | Interviewees' Responses  | Themes   |
|--|--|--|
| What are the difficulties you encountered when answering the questions in Section A, B, and C? |  |  |
| Band Two   | <ul style="list-style-type: none"> <li>• There were difficulties when answering questions in Section A. However, in Section B, it is hard to answer the questions without further knowledge of the background of the job. For me very difficult discuss in group discussion.</li> <li>• In Section A and Section B, I have no problem to answer but it is difficult for Section C because Section C requires to communicate with other friends. I have a little problem to say some words of English and told another friend. Besides, I also lack a vocabulary during the discussion took place. This led me to stutter and reduce the efficiency of the session. It was also difficult as to think of a better idea because the time allotted for thinking is limited.</li> <li>• It not much difficult in Section A, but in Section B, I have slight problem when I have to arrange my word when answering the interview question and in Section C I got problem to translate words from Malay to English especially in using the right grammar.</li> </ul> | <ul style="list-style-type: none"> <li>• Difficulty in Section A</li> <li>• Difficulty in Section B</li> <li>• Great difficulty in Section C</li> <li>• Lack of vocabulary in group discussion</li> <li>• Difficulty in grammar in Section C</li> <li>• Difficulty in constructing answer</li> </ul> |
| Band Three   | <ul style="list-style-type: none"> <li>• For Section A, I think there is no problem. For Section B and C, I feel that I cannot answer best because I had no confidence to have a conversation in English with people. Moreover, we are given a short time to prepare the answer.</li> <li>• For Section A, I have only little difficulties on answer it correctly and for Section B I have difficulties where I have no experience in job interview. Next, Section C, I'm not good at communication with people on giving an opinion.</li> <li>• I had no problem in answering Section A. However, I was quite nervous in answering Section B and C but still I managed to answer.</li> </ul>  | <ul style="list-style-type: none"> <li>• No difficulty in Section A</li> <li>• Little difficulty in Section B</li> <li>• Little difficulty in Section C</li> <li>• Little difficulty in communicative ability in group discussion</li> </ul>   |
| Band Four  | <ul style="list-style-type: none"> <li>• In Section A, there is no significant difficulties, but maybe I don't know whether I should answer very short or long. In Section B, everything is clear, since we required to speak and sometimes I got nervous. In Section C, we required to communicate with other members and little difficulties came when your friends do not agree. We have to find the best way to support arguments and also respect to our friends' arguments too.</li> <li>• For Section A, there is no difficulties at all. Section B, I think general knowledge and common sense to bring up yourself in front of interviewer is the most important thing. Section C, being spontaneous is somehow hard for me, but since I had already known the other friends there so I'm freely to express myself.</li> <li>• No problem for Section A. In B, nervous and lack of experience and C no problem at all.</li> </ul>   | <ul style="list-style-type: none"> <li>• No difficulty in Section A</li> <li>• No difficulty in Section B</li> <li>• No difficulty in Section C</li> </ul>   |

As it is evident from Table 5.8, respondents of different speaking proficiency levels encountered different difficulties in sections A, B and C during the speaking test. In general, respondents of lower band speaking proficiency (Band Two) encountered difficulties in sections A, B, and C of the ESL speaking proficiency test. As they claimed, they encountered difficulties, especially great difficulties in Section C. They also experienced problems in grammar. The respondents also faced difficulties in constructing ideas in Section C. Likewise, the respondents highlighted that constructing answers was a difficult task, especially in sections B and C as they were higher levels of questions in terms of difficulty. For example, one of the respondents said:

... There were difficulties when answering questions in Section A. However, in Section B, it is hard to answer the questions without further knowledge of the background of the job. For me very difficult discuss in group discussion...(B2-1-22/10/2016).

As for Band Three respondents, they faced less difficulties. As it is clear from Table 5.8, the respondents encountered less difficulties in Section A. However, they found Section B of the test a bit difficult. Like lower band respondents, Band Three respondents stressed that they had some difficulties in performing Section C of the test. Moreover, the respondents affirmed that they had difficulties in communicative tasks especially in Section C. For instance, one of the respondents stated:

... For Section A, I have only little difficulties on answer it correctly and for Section B I have difficulties where I have no experience in job interview. Next, Section C, I'm not good at communication with people on giving an opinion...(B3-2-25/10/2016 E).

By looking at Table 5.8, we can see that Band Four respondents did not experience major difficulties in performing the three different sections of the speaking proficiency test. However, some respondents experienced little nervousness when answering the question of Section B. Similarly, being spontaneous for some students was rather hard, especially in conducting a group discussion. For example, one of the Band Four respondents said:

...In Section A, there is no significant difficulties, but maybe I don't know whether I should answer very short or long. In Section B, everything is clear, since we required to speak and sometimes I got nervous. In Section C, we required to communicate with other members and little difficulties came when your friends do not agree. We have to find the best way to support our arguments and also respect to our friends' arguments too...(B4-1-29/10/2016).

Findings regarding the preference of the respondents regarding types of speaking tasks in speaking test have been summarized in Table 5.9. The findings regarding the types of speaking tasks obtained through the structured interviews and are followed by examples from the interview excerpts.

Table 5.9

*Preference of Types of Speaking Questions*

| Item  | Interviewees' Responses  | Themes   |
|---|--|--|
| Which type of questions do you prefer for speaking? Individual or group discussion? Give reasons. | <ul style="list-style-type: none"> <li>• I prefer individual test because I have no confidence to speaking English in front of many people I become nervous easily when speak if there are peoples around.</li> <li>• Well, for me individual test is easier to do because I able to talk when alone, but group discussion difficult to handle.</li> <li>• Type of question for speaking I prefer is alone test not group discussion because for me difficult to discuss topic.</li> </ul> | <ul style="list-style-type: none"> <li>• Preference of individual test</li> <li>• Nervousness in group discussion</li> </ul> |
| Band Three  | <ul style="list-style-type: none"> <li>• I prefer to engage in group discussions because I would have the chance to discuss and listen to the idea and opinion of other members.</li> <li>• I prefer having discussion group over individual one.</li> <li>• I would prefer group discussion because we don't get easily scared and bored and have lot of new ideas from other member. I feel comfortable with a group discussion with other member.</li> </ul>                            | <ul style="list-style-type: none"> <li>• Preference of group discussion</li> </ul>   |
| Band Four   | <ul style="list-style-type: none"> <li>• I prefer group discussion because it is more related to daily activities. And we speak with our friends, so I'm not that nervous.</li> <li>• I prefer B, because interview is what we always face in working life. We need to be good at it. For part C, you can learn it every day in a daily conversation.</li> <li>• Group discussion, it's fun to have interaction with other members rather than just facing the examiner alone.</li> </ul>  | <ul style="list-style-type: none"> <li>• Preference of group discussion</li> <li>• Preference of interview task</li> </ul>   |



In terms of preference of speaking tasks, respondents from different speaking proficiency levels prefer different types of speaking tasks based on their speaking ability. Similarly, in this study, respondents of lower band language proficiency (Band Two) favored different speaking tasks from those of higher language proficiency (Bands Three and Four). As such, respondents from lower band language proficiency preferred individual tasks as they reckoned that taking an individual speaking task is easier. They also highlighted that they feel nervous engaging in a group discussion task. For example, one of the respondents from lower band speaking proficiency stated:

...I prefer individual test because I have no confidence to speaking English in front of many people I become nervous easily when speak if there are peoples around...(B2-1-22/10/2016).

In contrast, respondents from higher band language proficiency preferred speaking tasks that include group discussion. Likewise, in this study, higher band language proficiency respondents preferred group discussion when taking a speaking task. Although interview tasks can be challenging for language learners, but as it can be seen from the responses of the respondents from Band Four language proficiency, they favored group discussion as well as interview tasks as they believed that they have the confidence in engaging similar speaking tasks. In this regard, one of the respondents from Band Four stated:

...I prefer B, because interview is what we always face in working life. We need to be good at it. For part C, you can learn it every day in a daily conversation...(B4-2-29/10/2016).

Finally, the data gathered were combined to enrich and help develop the speaking proficiency descriptors based on the responses of the respondents from different bands as well as the respondents' responses to the structured interview questions. As such, the combination of the quantitative and qualitative data led to the development of specific speaking proficiency descriptors (See Appendix L).

It is worth mentioning that the language lecturers can refer to the speaking proficiency descriptors to get precise and enough information pertaining to their students' speaking proficiency performance. With consequent to that, the language lecturers will be able to plan their teaching materials and instruction prepared for their students' needs in improving speaking proficiency.

### **5.3 Identifying Undergraduates' ESL Speaking Proficiency**

This section presents the respondents' speaking proficiency and it intends to answer research question 3.

Research Question 3:

How do the undergraduates perform on the ESL speaking proficiency test?

- a. What is the respondents' general ESL speaking proficiency?
- b. What is the overall ESL speaking proficiency performance according to gender, ethnicity, university, and school?

Identifying is considered as recognizing the respondents' speaking proficiency in the speaking proficiency test. The analysis of the respondents' speaking was based on the findings gained from the speaking proficiency test which included a total of 140

first-year undergraduates of different schools (4 students from each school) at two public universities in the northern part of Malaysia. Table 5.10 shows the percentages of university A respondents according to performance bands for each school.

Table 5.10

*Percentages of University A Respondents According to Performance Bands for Each School*

| School                                    | Band One (%) | Band Two (%) | Band Three (%) | Band Four (%) |
|---|--------------|--------------|----------------|---------------|
| 1. Management                             | 0            | 75           | 25             | 0             |
| 2. Physics                                | 0            | 75           | 25             | 0             |
| 3. Humanities                             | 0            | 0            | 0              | 100           |
| 4. Biology                                | 0            | 50           | 50             | 0             |
| 5. Housing, Building, and Planning        | 0            | 100          | 0              | 0             |
| 6. Educational Studies                    | 0            | 25           | 75             | 0             |
| 7. Aerospace Engineering                  | 0            | 25           | 75             | 0             |
| 8. Chemical Engineering                   | 0            | 75           | 25             | 0             |
| 9. Chemistry                              | 0            | 75           | 25             | 0             |
| 10. Civil Engineering                     | 0            | 75           | 25             | 0             |
| 11. Computer Science                      | 0            | 100          | 0              | 0             |
| 12. Industrial Technology                 | 0            | 100          | 0              | 0             |
| 13. Languages, Literacies and Translation | 0            | 0            | 25             | 75            |
| 14. Mathematics                           | 0            | 75           | 25             | 0             |
| 15. Pharmacy                              | 0            | 50           | 50             | 0             |
| 16. Social Sciences                       | 0            | 0            | 100            | 0             |
| 17. Communication                         | 0            | 50           | 25             | 25            |
| 18. Arts                                  | 0            | 25           | 75             | 0             |
| 19. Electronics Engineering               | 0            | 25           | 75             | 0             |
| 20. Materials Engineering                 | 0            | 50           | 50             | 0             |

The results of the respondents' speaking proficiency have been obtained and analyzed. Table 5.11 presents the performance bands, frequencies, and percentages of university A respondents who participated in the study.

Table 5.11

*University A Respondents' Speaking Proficiency*

| Performance Bands | Respondents (f) | Percentage |
|-------------------|-----------------|------------|
| Band Four         | 8               | 10%        |
| Band Three        | 30              | 37.5%      |
| Band Two          | 42              | 52.5%      |
| Band One          | 0               | 0.00%      |
| Total             | 80              |            |

The information in Table 5.10 shows that all the respondents from the School of Humanities were categorized in Band Four. Likewise, 3 respondents out of 4 respondents from the School of Languages, Literacies and Translation were also categorized in Band Four and only 1 respondent out 4 respondents from the School of Communication was categorized in Band Four. In contrast, all the 12 respondents from the Schools of Housing, Building and Planning, Computer Science and Industrial Technology were categorized in Band Two. Similarly, the percentages of university B respondents according to performance bands for each school have been shown in Table 5.12 below.

Table 5.12

*Percentages of University B Respondents According to Performance Bands for Each School*

| School                                      | Band One (%) | Band Two (%) | Band Three (%) | Band Four (%) |
|---|--------------|--------------|----------------|---------------|
| 1. Accountancy                              | 0            | 50           | 50             | 0             |
| 2. Applied Psychology                       | 0            | 0            | 75             | 25            |
| 3. Business Management                      | 0            | 75           | 25             | 0             |
| 4. Computing                                | 0            | 100          | 0              | 0             |
| 5. Economic and Finance                     | 0            | 100          | 0              | 0             |
| 6. Education                                | 0            | 75           | 25             | 0             |
| 7. Government                               | 0            | 100          | 0              | 0             |
| 8. International Studies                    | 0            | 50           | 25             | 25            |
| 9. Islamic Business                         | 0            | 25           | 75             | 0             |
| 10. Law                                     | 0            | 50           | 50             | 0             |
| 11. Multimedia Technology and Communication | 0            | 50           | 50             | 0             |
| 12. Quantitative Studies                    | 0            | 75           | 25             | 0             |
| 13. Social Development                      | 0            | 75           | 25             | 0             |
| 14. Technology Management                   | 0            | 50           | 50             | 0             |
| 15. Tourism Management                      | 0            | 50           | 50             | 0             |

The results of the respondents' speaking proficiency have been obtained and analyzed. Table 5.13 presents the performance bands, frequencies, and percentages of university B respondents who participated in the present study.

Table 5.13

*University B Respondents' Speaking Proficiency*

| Performance Bands | Respondents (f) | Percentage |
|-------------------|-----------------|------------|
| Band Four         | 2               | 3.33%      |
| Band Three        | 21              | 35%        |
| Band Two          | 37              | 61.66%     |
| Band One          | 0               | 0.00%      |
| Total             | 60              |            |

By looking at Table 5.13, it is evident that no student was categorized in band One (Novice performers) among university B respondents who participated in the present study. 37 respondents which are 61.66% of the respondents were categorized in Band Two (Intermediate performers) and 21 respondents which are 35% of the respondents were categorized in Band Three (Advanced Performers). Last, 2 respondents which are 3.33% of the respondents were categorized in Band Four (Superior performers) which is considered the highest band based on the bands developed specifically for the current study. The findings in Table 5.12 reveal that only two students from the Schools of Applied Psychology and International Studies were categorized in Band Four. However, all the 12 respondents from the three Schools of Computing, Economic, and Finance as well as the School of Government were in Band Two. The main research question 3, as well as the sub-research questions have been presented below. What follows is the detailed elaboration on research question 3 as well as the sub-research questions.

### **5.3.1 Respondents' General ESL Speaking Proficiency**

Based on the cut scores calculated (refer to section 5.3.1), the results of the respondents' general ESL speaking proficiency of respondents have been obtained and

analyzed. Table 5.14 presents the performance bands, frequencies, and percentages of respondents of both universities who participated in the study.

Table 5.14

*University A and University B Respondents' Speaking Proficiency*

| Performance Bands | Respondents (f) | Percentage |
|-------------------|-----------------|------------|
| Band Four         | 10              | 7.14%      |
| Band Three        | 51              | 36.43%     |
| Band Two          | 79              | 56.43%     |
| Band One          | 0               | 0.00       |
| Total             | 140             |            |

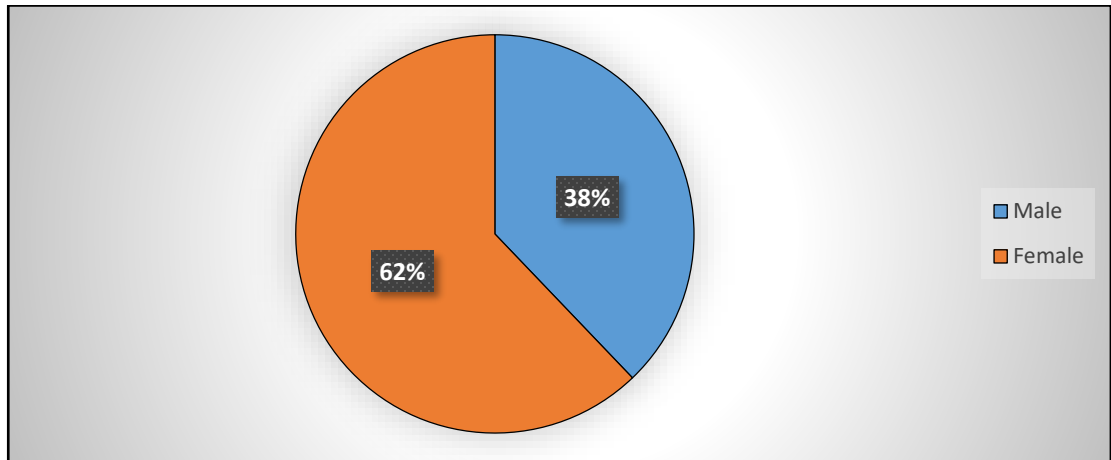
By looking at Table 5.14, it is evident that no student is categorized in Band One (Novice performers) among the respondents who participated in the present study. Seventy-nine respondents which are 56.43% of the respondents were categorized in Band Two (Intermediate performers) and 51 respondents which are 36.43% of the respondents were categorized in Band Three (Advanced Performers). Last, 10 respondents which are 7.14% of the respondents were categorized in Band Four (Superior performers) which is considered the highest band based on the bands developed specifically for the current study.

**5.3.2 Respondents' Overall ESL Speaking Proficiency Performance According to Gender, Ethnicity, University, and School**

This section presents the respondents' overall ESL speaking proficiency performance according to gender, ethnicity, university, and school. What follows is the detailed explanation of each section.

**a. Respondents' Overall ESL Speaking Proficiency According to Gender**

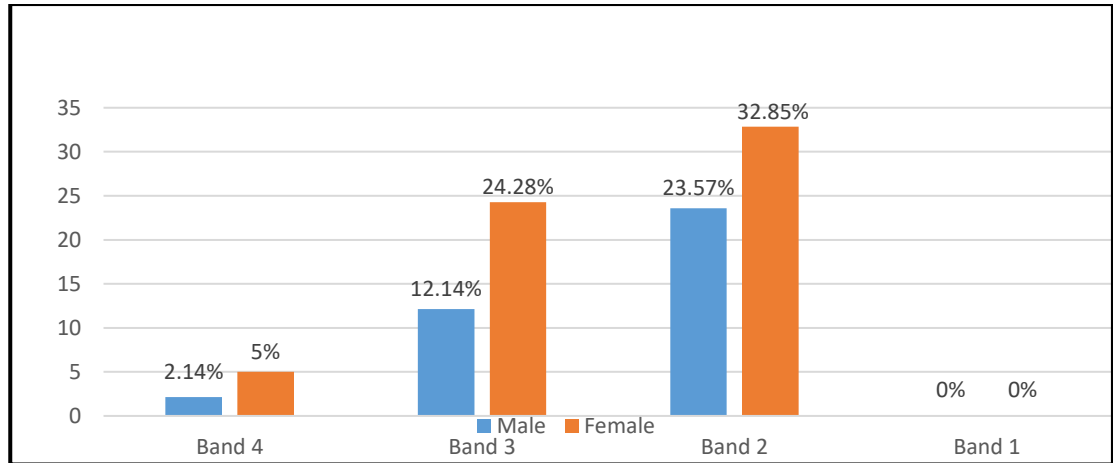
This subsection shows the respondents' speaking proficiency based on gender. Figure 5.4 illustrates the percentages of male and female respondents participated in the present study.



*Figure 5.4: Percentages of male and female respondents*

Key: N = 140 Respondents

The sample of male respondents is 53 respondents (38%) of the total number of respondents and female respondents is 87 respondents (62%), making it a total number of 140 respondents. In this study, based on the results achieved from the ESL speaking proficiency test, the respondents were categorized into 4 performance bands (Band One to Band Four). Figure 5.5 shows the respondents' speaking proficiency performance based on gender.



*Figure 5.5: Percentages of male and female respondents at the various bands*

Key: N = 140 Respondents

The percentages of the male and female respondents from Band Two to Band Four were calculated based on the total number of respondents (140). The findings of the study reveal much difference in gender speaking proficiency performance in Bands Two, Three and Four. It was found that the performance of the males who were classified in Band Four amounted to 2.14% while the performance of the females who were categorized in the same band amounted to 5%. Upon investigating further, it was revealed that the males who were classified in Band Three amounted to 12.14% whereas female respondents amounted to 24.28%. This leaves 56.42% of the total number of respondents (140) in Band Two. Of this, 23.57% were males and 32.85% were females.

In this study, 10 respondents (7.14%) of the total number of respondents (140) were categorized as superior performers. From those who were categorized as superior performers, the male respondents comprise of 3 respondents (30%) out of 10 respondents whereas the female respondents comprise of 7 respondents (70%).



Likewise, 51 respondents (36.43%) of the total number of respondents (140) were categorized as advanced performers. From those who were categorized as advanced performers, the male respondents comprise of 17 respondents (33.33%) out of 51 respondents whereas the female respondents comprise of 34 respondents (66.67%). 79 respondents (56.43%) of the total number of respondents (140) were categorized as intermediate performers. From those who were categorized as intermediate performers, the male respondents comprise of 33 respondents (41.77%) out of 79 respondents whereas the female respondents comprise of 46 respondents (58.23%). In this study, however, no student was categorized as novice performers. The present study is about assessing undergraduates' speaking proficiency. As a matter of fact, undergraduates' speaking proficiency is assessed before embarking on their academic studies at the tertiary level. Therefore, undergraduates are expected to speak and converse in English to some extent as the medium of instruction is English in most of the schools at the tertiary level education. Hence, it was expected that none of the respondents who were undergraduate students to be categorized as novice performers. As such, in this study, the results reveal that none of the respondents out of the total number of respondents was categorized as novice performers. Figure 5.6 shows the performers' performance based on gender.

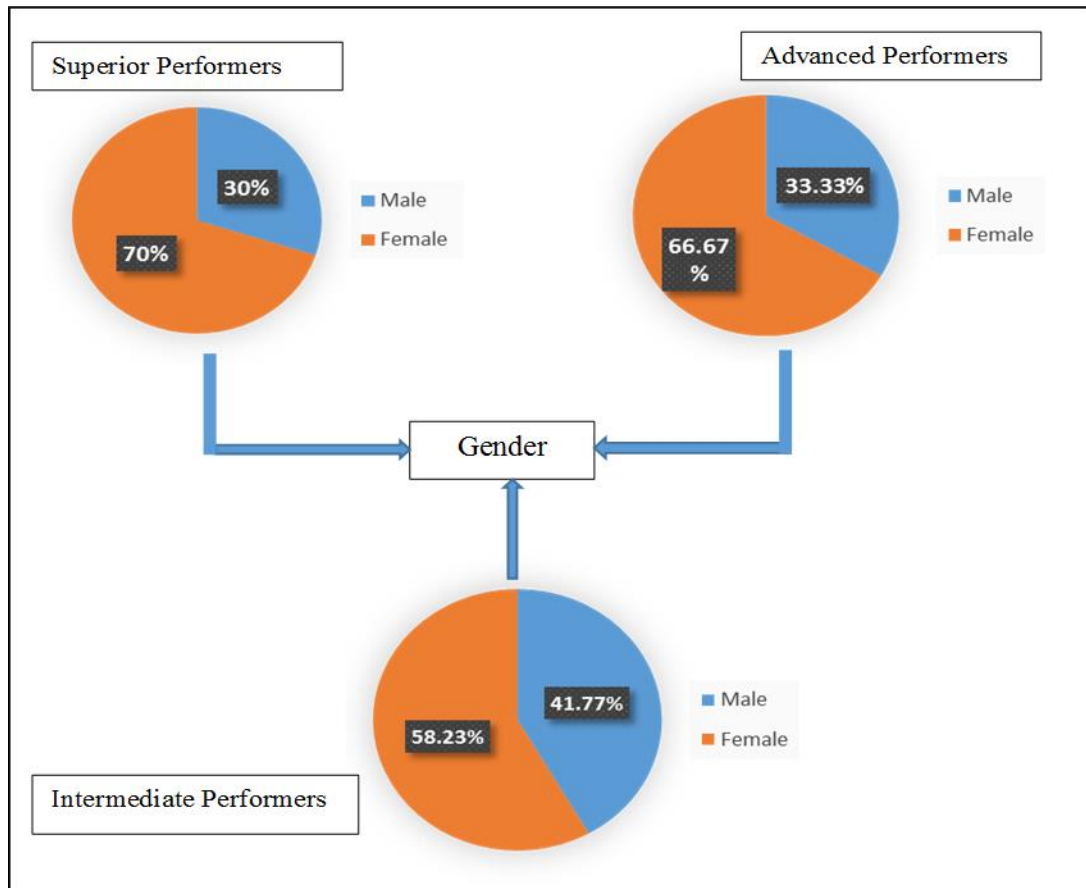
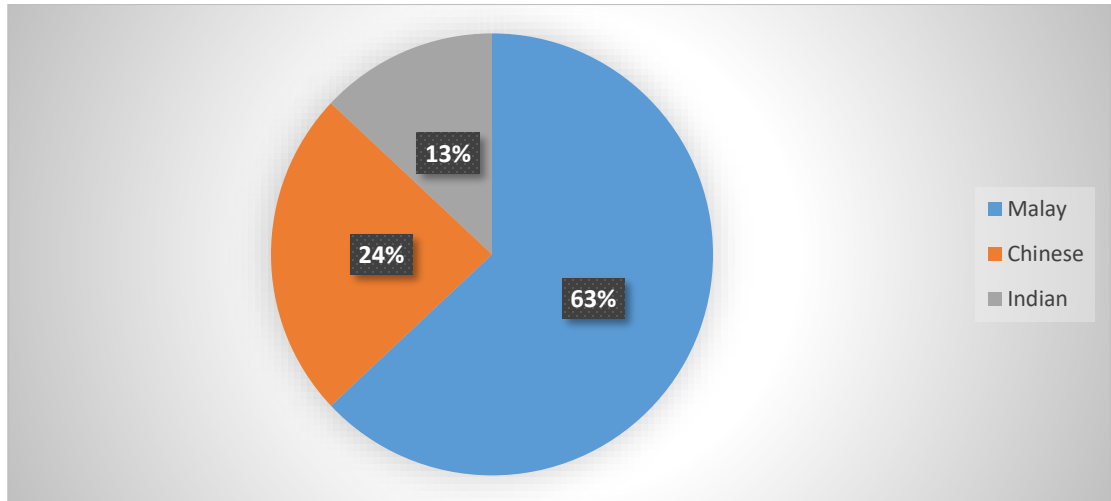


Figure 5.6: Gender performance

Key: Superior performers= 10 respondents, advanced performers= 51, intermediate performers= 79 respondents

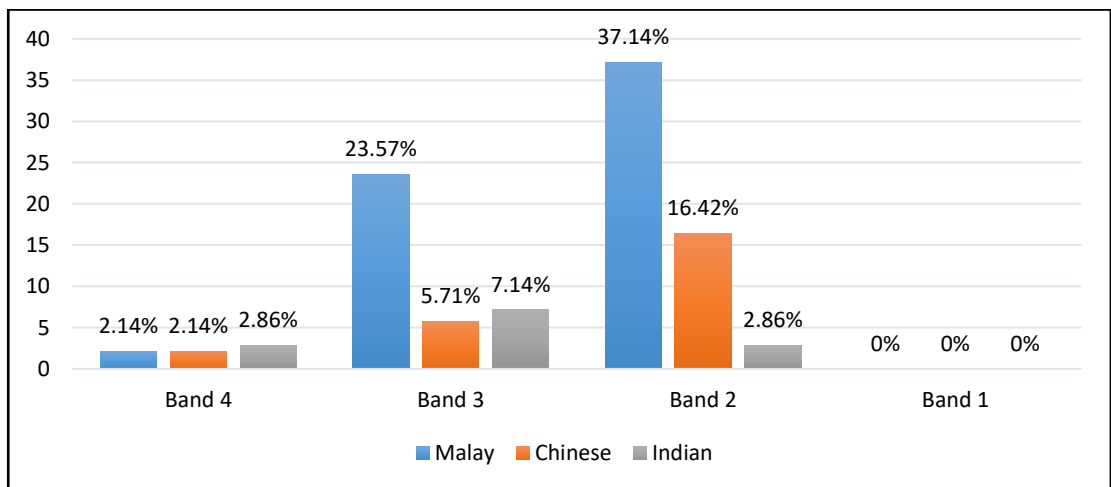
**b. Respondents' Overall ESL Speaking Proficiency According to Ethnicity**

This sub-section presents the respondents' speaking proficiency based on ethnicity. Figure 5.7 illustrates the percentages of the different ethnic groups.



*Figure 5.7: Percentages of respondents of different ethnic groups*  
Key: N = 140 Respondents

Out of 140 respondents, 88 respondents (63%) were Malays, the Chinese consisted of 34 (24%) whereas the Indians consisted of 18 respondents (13%) of the total number of respondents. Based on the findings obtained from the ESL speaking proficiency test, the respondents were categorized into 4 performance bands (Band One to Band Four). Figure 5.8 illustrates the respondents' speaking proficiency based on ethnicity.



*Figure 5.8: Percentages of different ethnic groups at the various bands*  
Key: N = 140 Respondents

The percentages of the Malay, Chinese and Indian respondents from Band Two to Band Four were calculated based on the total number of respondents (140). The results of the study reveal that there are differences in the speaking proficiency performance in Bands Two, Three and Four. It was found that the performance of the Malay respondents (three) who were in Band Four amounted to 2.14% and the performance of the Chinese respondents (three) who were categorised in the same band amounted to 2.14% whereas the performance of the Indian respondents (4) who were in Band Four amounted to 2.86%. Upon studying further, it was revealed that the Malays (33) who were categorized in Band Three amounted to 23.57%, the Chinese (eight) amounted to 5.71% and the Indian respondents (10) amounted to 7.14%. Likewise, Malay respondents (52) who were categorized in band 2 amounted to 37.14% as compared to the performance of Chinese respondents (23) which amounted to 16.42% and Indians (four) which amounted to 2.86%.

In this study, 10 respondents (7.14%) of the total number of respondents (140) were categorized as superior performers. From those who were categorized as superior performers, the Malay respondents comprise of 3 respondents (30%) out of 10 respondents, the Chinese respondents comprise of 3 respondents (30%) whereas the Indian respondents comprise of 4 respondents (40%). Similarly, 51 respondents (36.42%) of the total number of respondents (140) were categorized as advanced performers. From those who were categorized as advanced performers, the Malay respondents comprise of 33 respondents (64.7%) out of 51 respondents, the Chinese respondents comprise of 8 respondents (15.7%) whereas the Indian respondents comprise of 10 respondents (19.6%). 79 respondents (56.42%) of the total number of

respondents (140) were categorized as intermediate performers. From those who were categorized as intermediate performers, the Malay respondents comprise of 52 respondents (65.8%) out of 79 respondents, the Chinese respondents comprise of 23 respondents (29.1%) whereas the Indian respondents comprise of 4 respondents (5.1%). Figure 5.9 shows the performers' performance based on ethnicity.

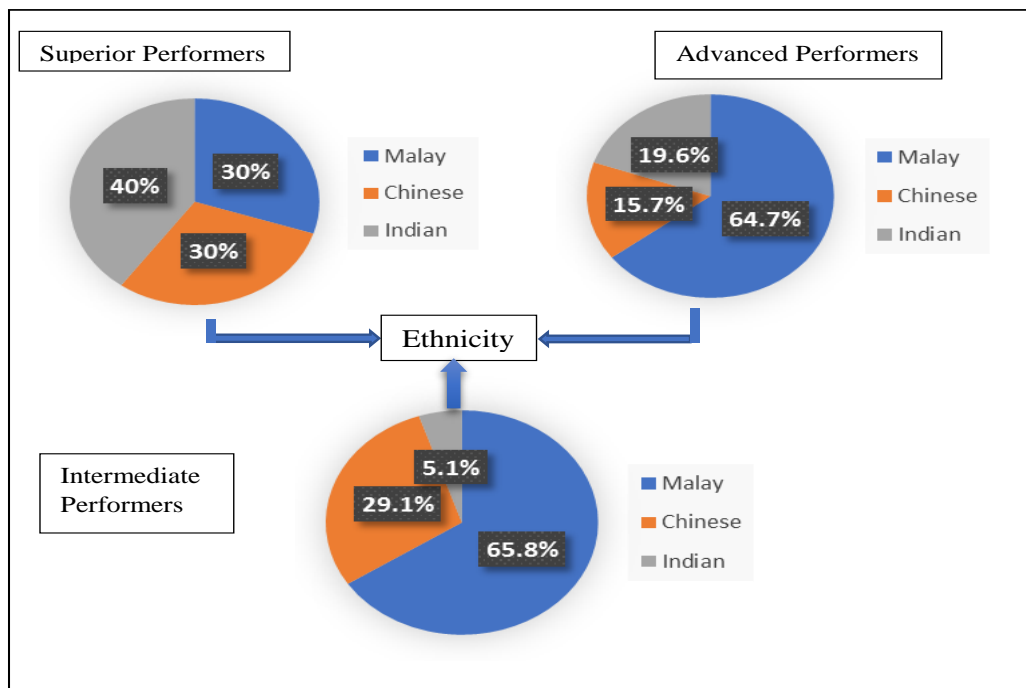
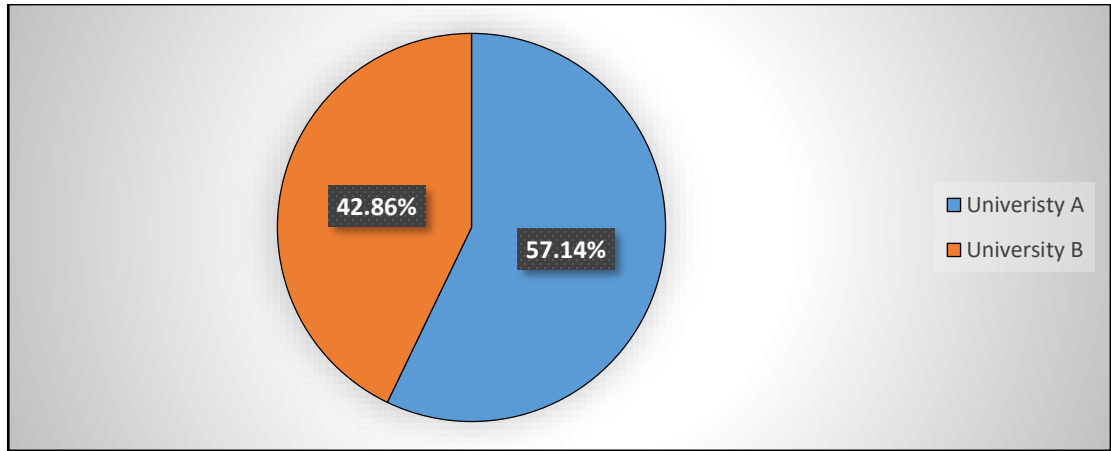


Figure 5.9: Ethnicity performance

Key: Superior performers= 10 respondents, advanced performers= 51, intermediate performers= 79 respondents

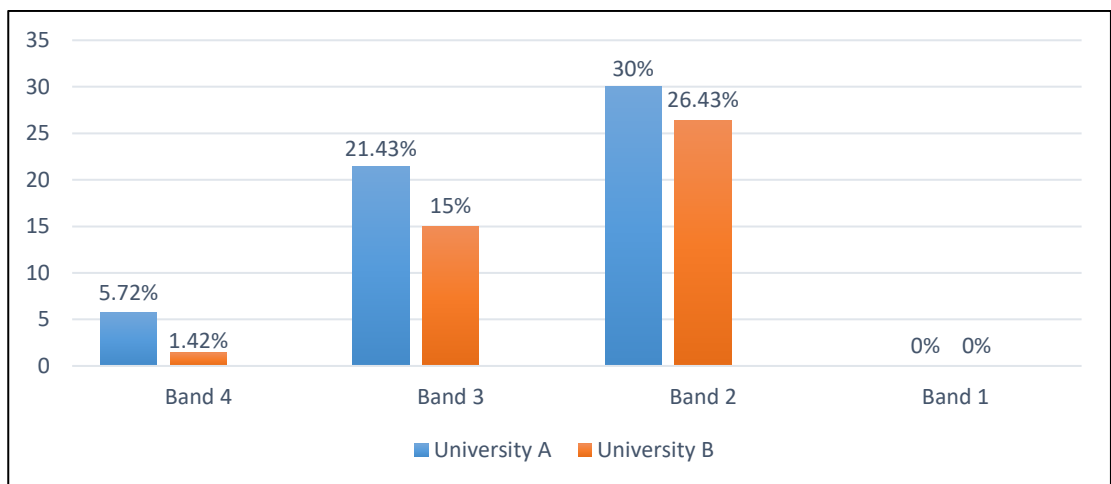
### c Respondents' Overall ESL Speaking Proficiency According to University

This sub-section presents the respondents' speaking proficiency based on University. Figure 5.10 illustrates the percentages of respondents from university A and university B.



*Figure 5.10: Percentages of respondents of university A and university B*  
 Key: N = 140 respondents

Out of 140 respondents, 80 respondents (57.14%) were respondents from university A whereas 60 respondents (42.86%) from university B of the total number of respondents (140). Based on the findings obtained from the ESL speaking proficiency test, the respondents were categorized into 4 performance bands (Band One to band Four). Figure 5.11 illustrates the respondents' speaking proficiency based on university.

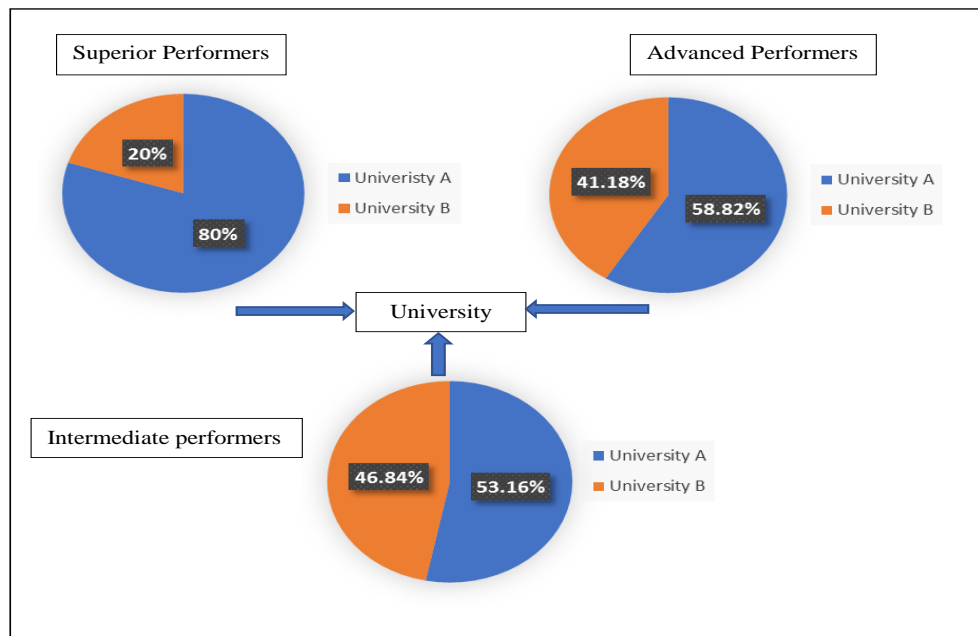


*Figure 5.11: Percentages of respondents in university A and university B at the various bands*  
 Key: N = 140 Respondents

The percentages of university A and university B respondents from Band Two to Band Four were calculated based on the total number of respondents (140). The results of the study reveal that there are differences in the speaking proficiency performance in Bands Two, Three and Four. It was found that the performance of university A respondents who were in Band Four amounted to 5.72% while the performance of university B respondents who were categorized in the same band amounted to 1.42%. Upon studying further, it was revealed that university A respondents who were categorized in Band Three amounted to 21.43% whereas university B respondents amounted to 15%. University A respondents who were categorized in Band Two amounted to 30% as compared to the performance of university B respondents which amounted to 26.43%.

Findings of the study show that 10 respondents (7.14%) of the total number of respondents (140) were categorized as superior performers. From those who were categorized as superior performers, university A respondents comprise of 8 respondents (80%) out of 10 respondents whereas university B respondents comprise of 2 respondents (20%). 51 respondents (36.43%) of the total number of respondents (140) were categorized as advanced performers. From those who were categorized as advanced performers, university A respondents comprise of 30 respondents (58.82%) out of 51 respondents whereas university B respondents comprise of 21 respondents (41.18%). However, 79 respondents (56.43%) of the total number of respondents (140) were categorized as intermediate performers. From those who were categorized as intermediate performers, university A respondents comprise of 42 respondents (53.16%) out of 79 respondents whereas university B respondents comprise of 37

respondents (46.84%). Figure 5.12 shows the performers' performance based on university.



5:12: University performance

Key: Superior performers= 10 respondents, advanced performers= 51, intermediate performers= 79 respondents

#### d. Respondents' Overall ESL Speaking Proficiency According to School

This subsection shows the respondents' speaking proficiency based on school in both universities.

##### i. Schools in University A

Table 5.15 shows the schools in university A, categorised by superior, advanced and intermediate performance with majority of students (i.e., when three or four respondents out of four respondents were categorized in each school). All the 4 respondents in the School of Humanities and the School of Languages, Literacies, and



Translation were categorized as superior performers. Respondents in the schools of Educational Studies, Aerospace Engineering, Social Sciences, Arts and Electronics Engineering were categorized as advanced performers. Whereas students from the Schools of Management, Physics, Housing, Building and Planning, Chemical Engineering, Chemistry, Civil Engineering, Computer Science, Industrial Technology, and Mathematics were categorized as intermediate performers. However, no schools were categorized as novice performance school.

Table 5.15

*School Performance*

| Performance              | Schools   |
|--------------------------|---|
| Superior Performance     | 1. Humanities<br>2. Languages, Literacies and Translation   |
| Advanced Performance     | 1. Educational Studies<br>2. Aerospace Engineering<br>3. Social Sciences<br>4. Arts<br>5. Electronics Engineering   |
| Intermediate Performance | 1. Management<br>2. Physics<br>3. Housing, Building, and Planning<br>4. Chemical Engineering<br>5. Chemistry<br>6. Civil Engineering<br>7. Computer Science<br>8. Industrial Technology<br>9. Mathematics |
| Novice Performers        | -   |
| Total                    | 16  |

**ii. Schools in University B**

Table 5.16 shows the schools grouped by superior, advanced and intermediate performance with majority of students (i.e., when three or four respondents out of four respondents were categorized in each specific school). Based on the performance of

the respondents no schools were categorized as superior and advanced performance, whereas 7 schools were categorized as intermediate performance. However, no schools were categorized as novice performance.

Table 5.16

*School Performance*

| Performance              | Schools   |
|--------------------------|---|
| Superior Performance     | -   |
| Advanced Performance     | -   |
| Intermediate Performance | <ol style="list-style-type: none"> <li>1. Business Management</li> <li>2. Computing</li> <li>3. Economics and Finance</li> <li>4. Education</li> <li>5. Government</li> <li>6. Quantitative Studies</li> <li>7. Social Development</li> </ol> |
| Novice Performers        | -   |
| Total                    | 7   |

#### 5.4 Profiling ESL Undergraduates' ESL Speaking Proficiency

The final research question in this study is; what are the profiles of the undergraduates' ESL speaking proficiency?

- a. Who are the superior performers?
- b. Who are the advanced performers?
- c. Who are the intermediate performers?
- d. Who are the novice performers?

In general, profiles do not show ability, but only performance. In other words, profiling is to identify the characteristics of the speakers in the different categories used specifically in this study. Further, a profile is an outline, especially one representing the shape of something. As such, in the profiling process, it is necessary to revert to the

performance bands and speaking proficiency descriptors developed specifically for the purpose of this study.

The speaking proficiency performance bands simply refer to the performance bands specifically developed in this study. The speaking proficiency descriptors refer to the indicators which show the detailed descriptions and information of the specific speaking proficiency at each performance band (Band One to Band Four) which is shown in four performance levels of speaking proficiency namely; superior performers, advanced performers, intermediate performers and novice performers. This is perhaps best illustrated in Figure 5.13.

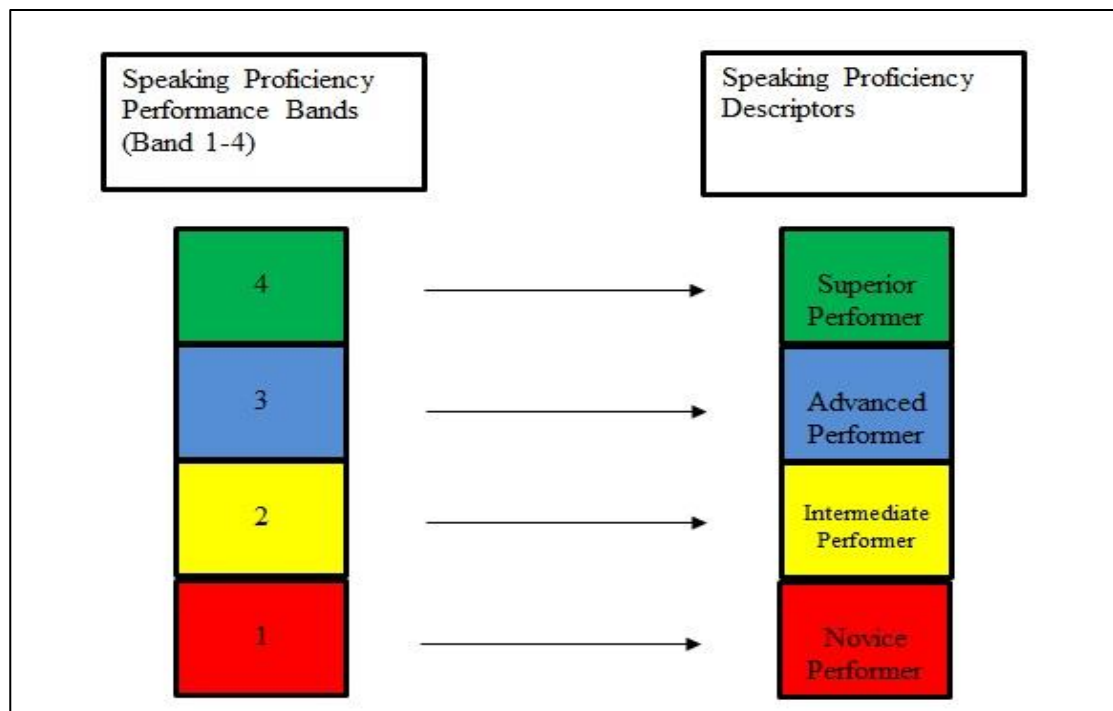


Figure 5.13: Speaking proficiency performance bands/speaking proficiency descriptors

Figure 5.13 presents the interaction between these two descriptors in the profiling process. Based on the performance of the participants of the study on the ESL speaking

proficiency test, these undergraduate students were then classified into the different performance bands (Band One to Band Four) and each band is followed by a descriptor.

As such, the undergraduates in:

- a. Band Four (those whose scores are within 78-100) are categorized as superior performers.
- b. Band Three (those whose scores are within 54-77) are categorized as advanced performers.
- c. Band Two (those whose scores are within 30-53) are categorized as intermediate performers.
- d. Band One (those whose scores are within 0-29) are categorized as novice performers.

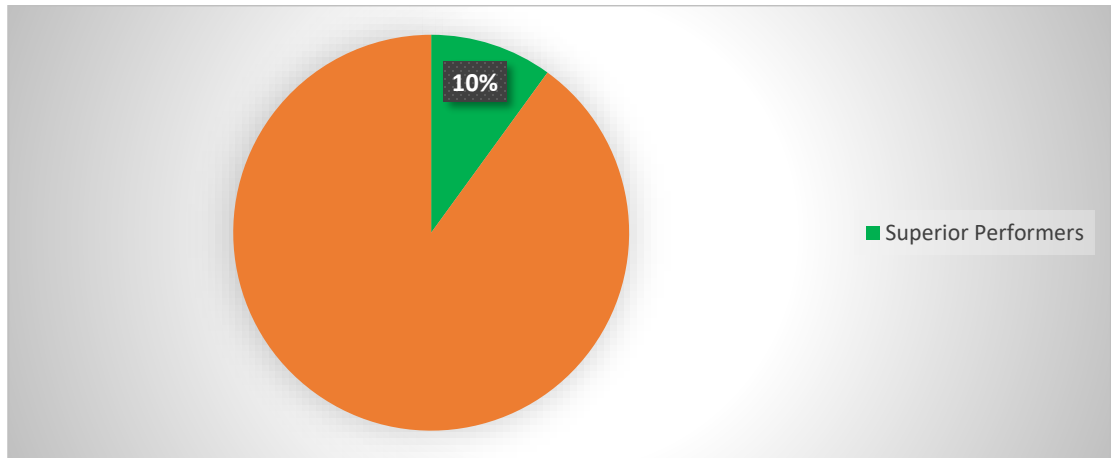
After the analysis of the respondents' performance by using cut scores (refer to section 5.2.1), the respondents were then classified into a number of performance bands, which are specifically exclusive only to this study. Therefore, it is crucial to identify what the undergraduate students are able to do and not able to do at each of the performance bands. In doing so, it was decided to reveal the profiles of the undergraduate students of both universities separately. As such, in the following section, no new data are introduced, it is rather to synthesize the data to chart out the profiles of the superior performers, advanced performers, intermediate performers, novice performers of undergraduates. What follows is a detailed elaboration of profiles of university A and university B undergraduates' speaking proficiency performance.

#### **5.4.1 Profiles of University A Undergraduates' Performance**

In this section, the profiles of university A undergraduate speakers' speaking proficiency have been presented.

**a) Profiles of Superior Performers**

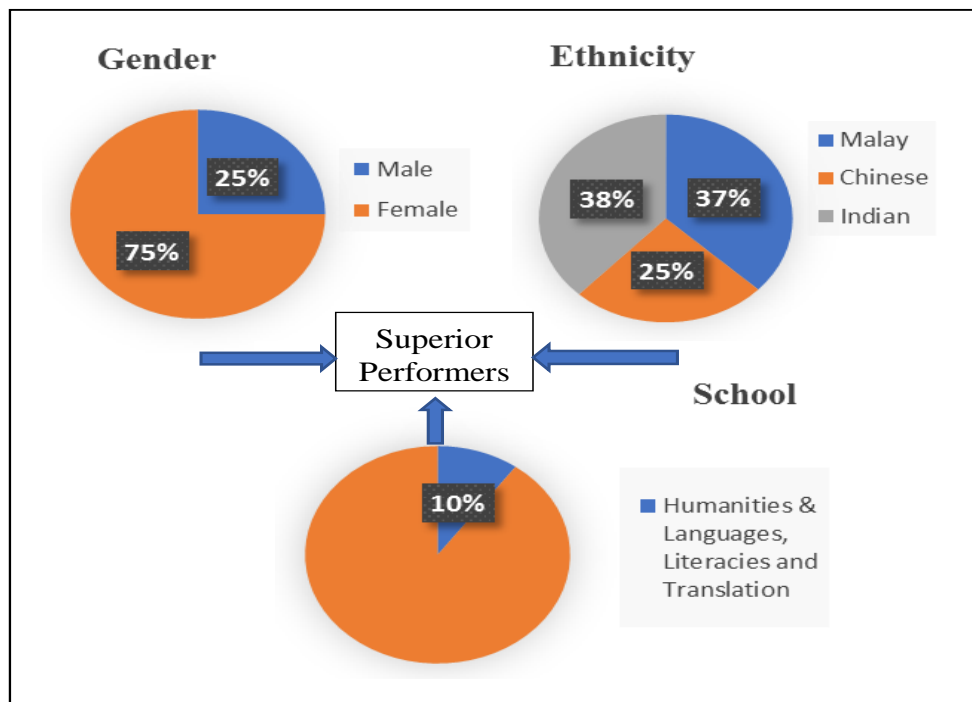
Out of the total number of university A respondents (80 respondents), only 8 respondents were categorized as superior performers. This comprises only 10% of the total number of university A respondents as presented in Figure 5.14.



*Figure 5.14:* University A superior performers

Key: N = 80 respondents

In profiling university A superior performers, data from the different sections of gender, ethnicity, and school were combined as shown in Figure 5.15. It should be noted that the calculation in Figure 5.15 is based on the total number of the respondents as superior performers (8 respondents). However, the calculation for school category is based on the total number of schools.



5.15: Percentages of respondents as superior performers

Key: n= 8 respondents

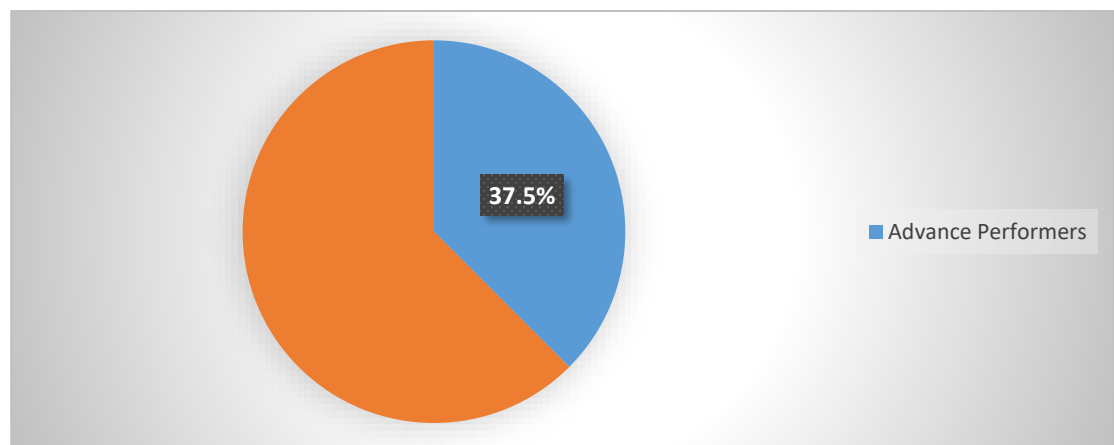
- Gender: In terms of gender, 75% of the respondents as superior performers are females (6) while 25% are males (2).
- Ethnicity: In terms of ethnicity, 37% of the respondents as superior performers are Malays (3) and the Chinese (2) comprise of 25%, while the Indians (3) comprise of 38%.
- School: In terms of school, 10% of the schools as superior performance schools are the School of Humanities and the School of Languages, Literacies, and Translation.

It should be noted that in the first stage of the profiling process, the gathered data were deconstructed to combine the profile of individual groups. After the details of all the individual groups were identified, all the groups were reconstructed to show how the groups are compared to one another. As such, it was possible to see the differences among the individual groups. By looking at Table 5.17 [refer to section 5.4.3(a)], we can see that the superior performers are prone to female respondents. Likewise, it can

be seen that the superior performers are balanced between Malays and Indians in terms of ethnicity. It is also important to note that those respondents who were categorized as superior performers were from the two schools of School of Humanities and School of Languages, Literacies and Translation. It is important to note that those students who are categorized as superior performers perform very well in the sections A, B, and C of the ESL speaking proficiency test.

**b. Profiles of Advanced performers**

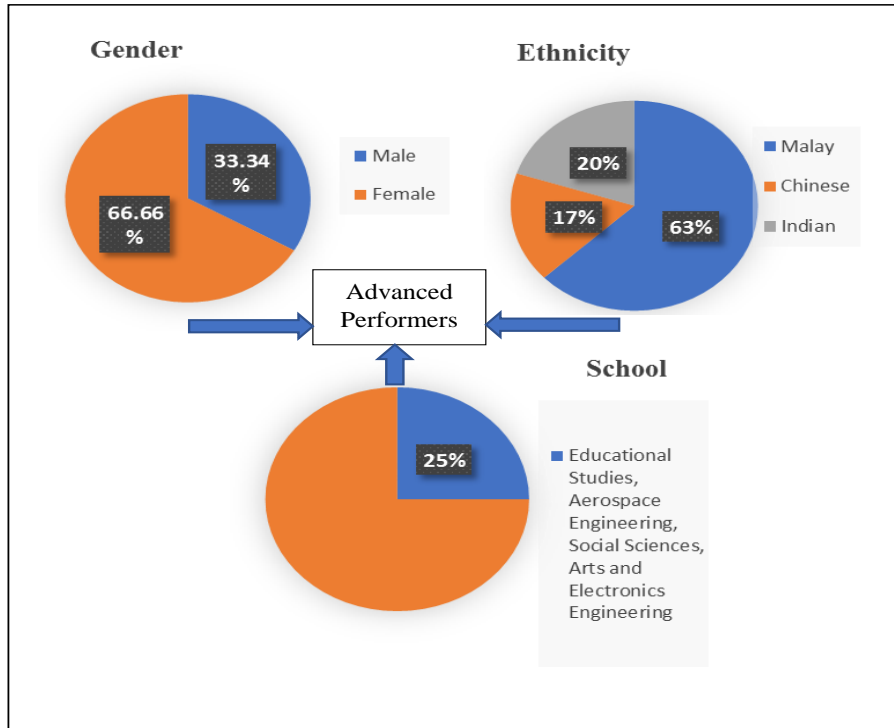
Out of the total number of university A respondents (80 respondents), 30 respondents were categorized as advanced performers. This comprises 37.5% of the total number of university A respondents as presented in Figure 5.16.



*Figure 5:16:* University A advanced performers  
Key: N= 80 respondents

In profiling university A advanced performers, data from the different sections of gender, ethnicity, and school were combined as shown in Figure 5.17 It should be noted that the calculation in Figure 5.17 is based on the total number of university A

respondents as advanced performers (30 respondents). However, the calculation for school category is based on the total number of schools.



*Figure 5.17: Percentages of respondents as advanced performers*  
Key: n=30 respondents

- a. Gender: In terms of gender, 66.66% of the respondents as advanced performers are females (20) while 33.34% are males (10).
- b. Ethnicity: In terms of ethnicity, 63% of the respondents as advanced performers are Malays (19), the Chinese (5) comprise of only 17% whereas Indians (6) comprise of 20%.
- c. School: In terms of school, 25% of the schools as advanced performance schools are the Schools of Educational Studies, Aerospace Engineering, Social Sciences, Arts and Electronics Engineering.

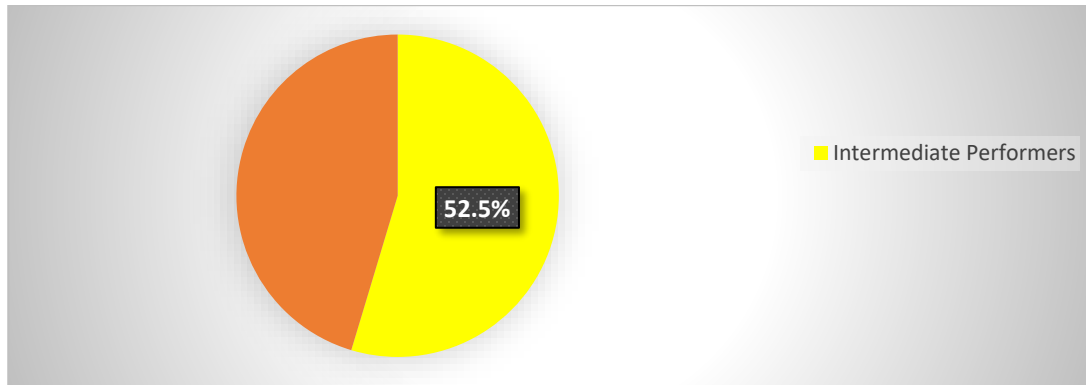
By referring to Table 5.17 [refer to section 5.4.3(a)], we can see that the advanced performers are rather prone to females. The data also shows that the ethnicity made-up



of the advanced performers are rather prone to Malays. Those performing at the advanced level are speakers whose performance are satisfactory in the sections A, B and C of the ESL speaking proficiency test.

**c. Profiles of Intermediate Performers**

Out of the total number of university A respondents (80 respondents), 42 respondents were categorized as intermediate performers. This comprises 52.5% of the total number of university A respondents as presented in Figure 5.18.



*Figure 5.18:* University A intermediate performers

Key: N=80 respondents

In profiling university A intermediate performers, data from the different sections of gender, ethnicity, and school were combined as shown in Figure 5.19. It should be noted that the calculation in Figure 5.19 is based on the total number of university A respondents as intermediate performers (42 respondents). However, the calculation for school category is based on the total number of schools.

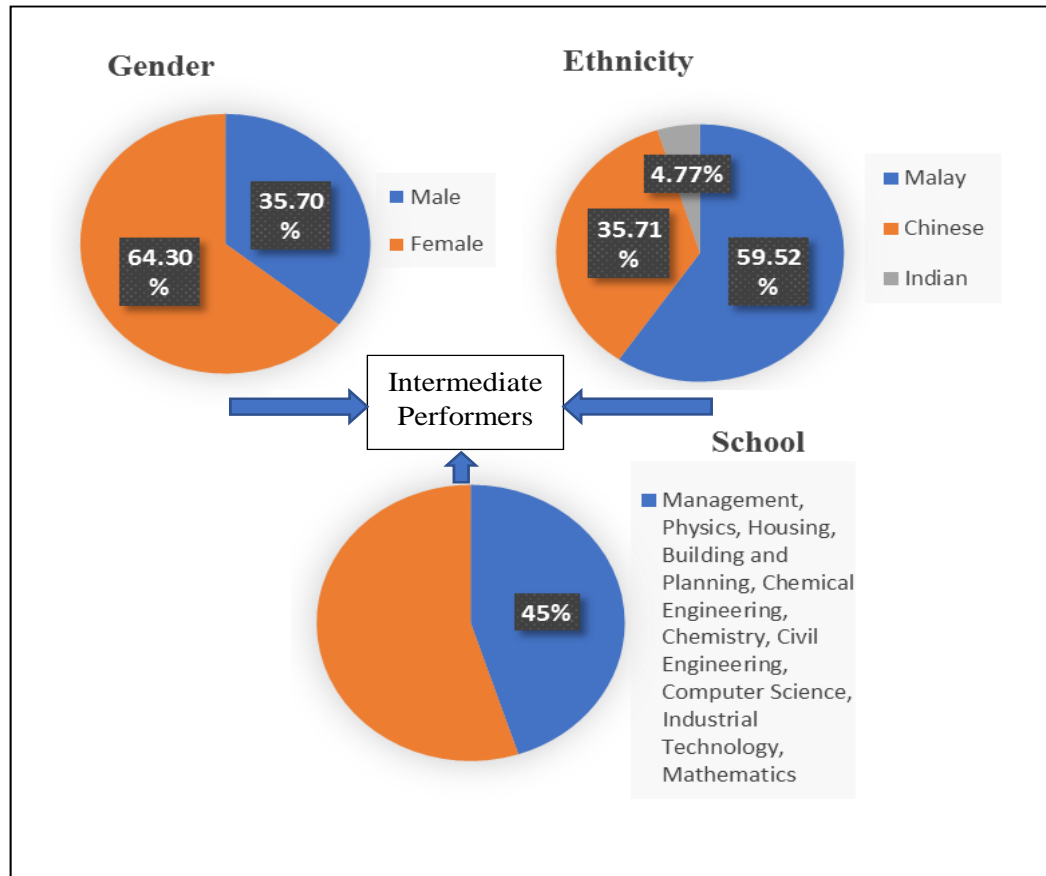


Figure 5.19: Percentages of respondents as intermediate performers  
Key: n=42 respondents

- Gender: In terms of gender, 64.30% of the respondents as intermediate performers are females (27) while 35.70% are males (15).
- Ethnicity: In terms of ethnicity, 59.52% of the respondents as intermediate performers are Malays (25), the Chinese (15) comprise of 35.71%, while the Indians (2) only comprise of 4.77%.
- School: In terms of school, 45% of the schools as intermediate performance schools are the Schools of Management, Physics, Housing, Building and Planning, Chemical Engineering, Chemistry, Civil Engineering, Computer Science, Industrial Technology, and Mathematics.

Findings show that the intermediate performers are rather prone to females. The data gathered from Table 5.17 [refer to section 5.4.3(a)] also shows that in terms of

ethnicity made-up of the intermediate performers are also prone to Malays. It should be noted that those performing at the intermediate level are speakers who have difficulty in performing sections A, B, and C of the ESL speaking proficiency test.

**d. Profiles of Novice performers**

The present study is about assessing undergraduates' speaking proficiency. As a matter of fact, undergraduates' speaking proficiency is assessed before embarking on their academic studies at the tertiary level. Therefore, undergraduates are expected to speak and converse in English to some extent as the medium of instruction is English in most of the schools at the tertiary level education. Hence, it was expected that none of the respondents who were undergraduate students to be categorized as novice performers. As such, in this study, the results reveal that none of the university A respondents out of the total number of respondents was categorized as novice performers. The next section presents the profiles of university B undergraduates' speaking proficiency performance.

**5.4.2 Profiles of University B Undergraduates' Performance**

In this section, the profiles of university B undergraduate speakers' speaking proficiency have been presented.

**a. Profiles of Superior Performers**

Out of the total number of university B respondents (60 respondents), only 2 respondents were categorized as superior performers. This comprises only 3.33% of the total number of university B respondents as presented in Figure 5.20.

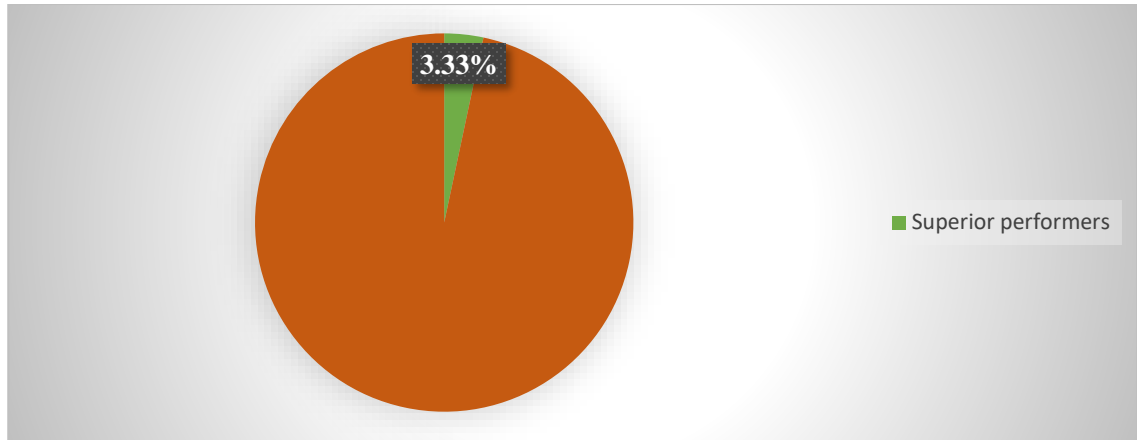


Figure 5.20: University B superior performers  
Key: N=60 respondents

In profiling university B superior performers, data from the different sections of gender, ethnicity, and school were combined as shown in Figure 5.21. It should be noted that the calculation in Figure 5.21 is based on the total number of the respondents as superior performers (2 respondents). However, the calculation for school category is based on the total number of schools.

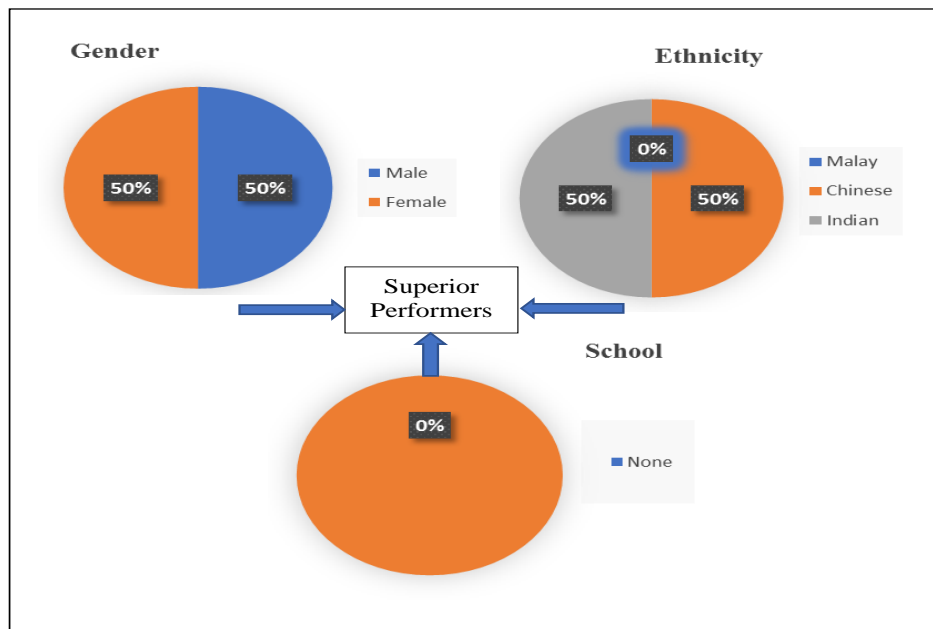


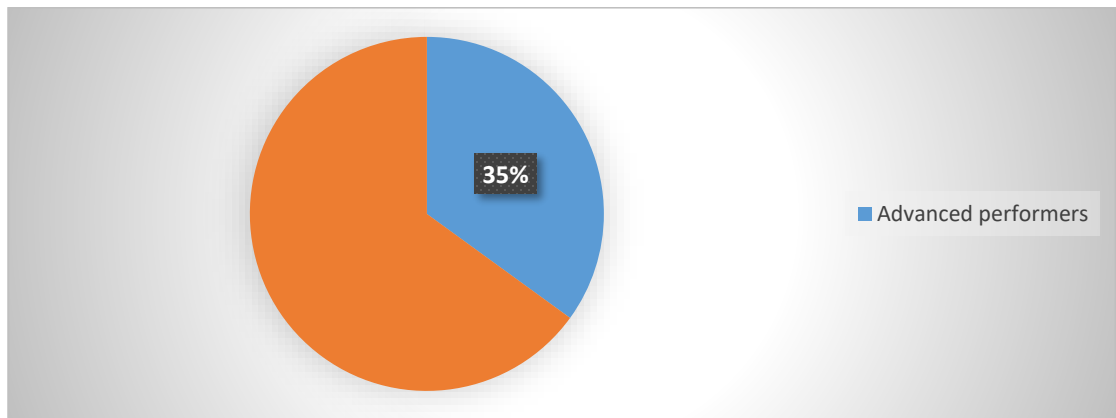
Figure 5.21: Percentages of respondents as superior performers  
Key: n= 2 respondents

- a. Gender: In terms of gender, 50% of the respondents as superior performers are females (1) and 50% are males (1).
- b. Ethnicity: In terms of ethnicity, 50% of the respondents as superior performers are Chinese (1), the Indians (1) comprise of 50%, while the Malays (0) comprise of 0.0%.
- c. School: In terms of school, none of the schools in university B was categorized as superior performance school.

By looking at Table 5.18 [refer to section 5.4.3(b)], we can see that the superior performers are balanced in terms of gender. From the findings, it can be seen that the superior performers are also balanced between Chinese and Indian respondents. It is important to note that those students who are categorized as superior performers perform very well in the sections A, B, and C of the ESL speaking proficiency test.

**b. Profiles of Advanced performers**

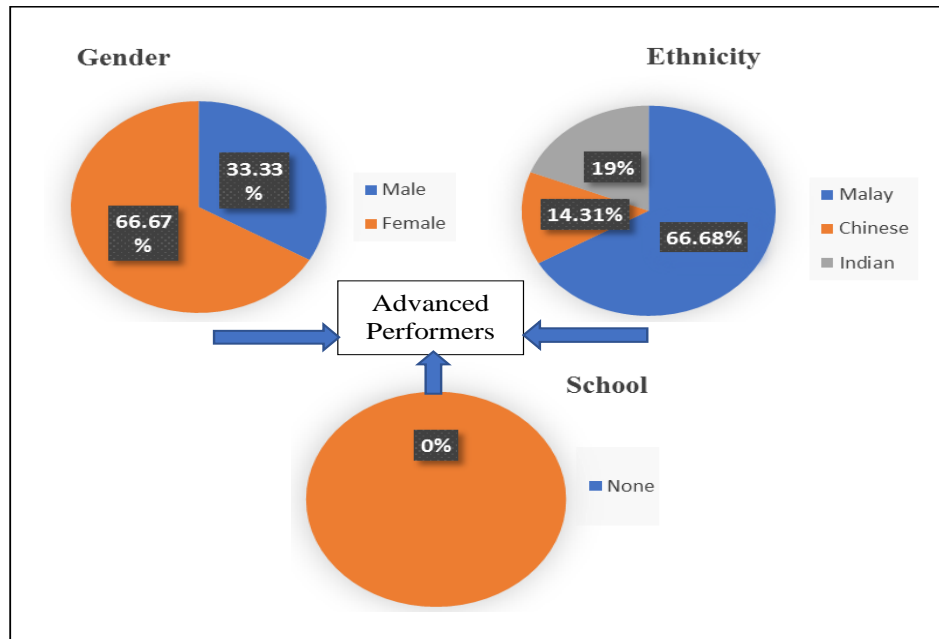
Out of the total number of university B respondents (60 respondents), 21 respondents were categorized as advanced performers. This comprises 35% of the total number of university B respondents as presented in Figure 5.22.



*Figure 5.22: University B advanced performers*

Key: N=60 respondents

In profiling university B advanced performers, data from the different sections of gender, ethnicity, and school were combined as shown in Figure 5.23. It should be noted that the calculation in Figure 5.23 is based on the total number of university B respondents as advanced performers (21 respondents). However, the calculation for school category is based on the total number of schools.



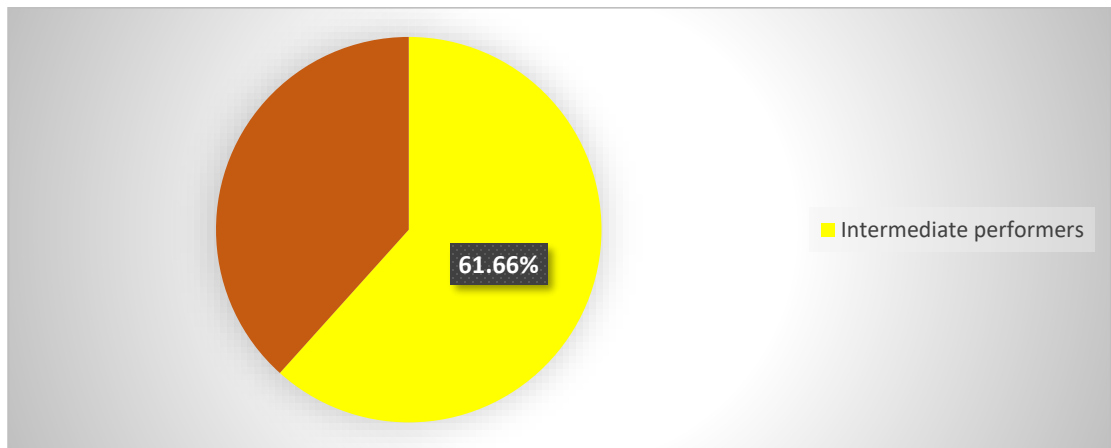
*Figure 5.23: Percentages of respondents as advanced performers*  
Key: n=21 respondents

- a. Gender: In terms of gender, 66.67% of the respondents as advanced performers are females (14) while 33.33% are males (7).
- b. Ethnicity: In terms of ethnicity, 66.68% of the respondents as advanced performers are Malays (14), the Chinese (3) comprise of 14.31%, while the Indians (4) comprise of 19%.
- c. School: In terms of school, none of the university B schools was categorized as advanced performance school.

By referring to Table 5.18 [refer to section 5.4.3(b)], we can see that the advanced performers are rather prone to females. The data also shows that the ethnicity made-up of the advanced performers are rather prone to Malays. Those performing at the advanced level are speakers whose performance are satisfactory in the sections A, B and C of the ESL speaking proficiency test.

**c. Profiles of Intermediate Performers**

Out of the total number of university B respondents (60 respondents), 37 respondents were categorized as intermediate performers. This comprises 61.66% of the total number of university B respondents as presented in Figure 5.24.



*Figure 5.24:* University B intermediate performers

Key: N=60 respondents

In profiling university B intermediate performers, data from the different sections of gender, ethnicity, and school were combined as shown in Figure 5.25. It should be noted that the calculation in Figure 5.25 is based on the total number of university B respondents as intermediate performers (37 respondents). However, the calculation for school category is based on the total number of schools.

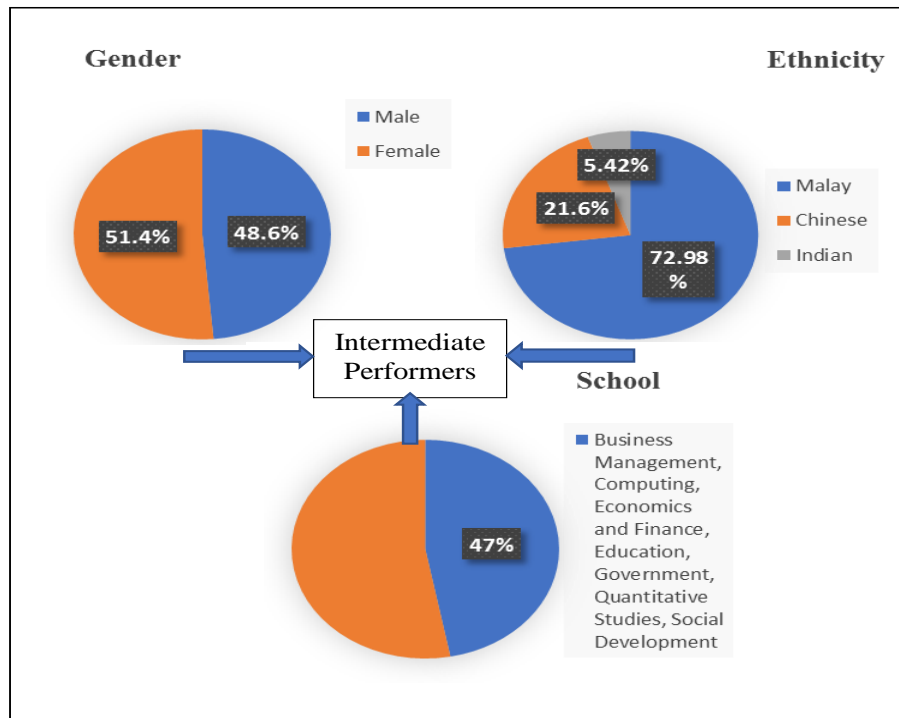


Figure 5.25: Percentages of respondents as intermediate performers

Key: n=37 respondents

- a. Gender: In terms of gender, 51.4% of the respondents as intermediate performers are females (19) while 48.6% are males (18).
- b. Ethnicity: In terms of ethnicity, 72.98% of the respondents as intermediate performers are Malays (27), the Chinese (8) comprise of 21.60%, while the Indians (2) comprise of only 5.42%.
- c. School: In terms of school, 47% of the schools as intermediate performance schools are the Schools of Business Management, Computing, Economics and Finance, Education, Government, Quantitative Studies and Social Development.

Findings show that the intermediate performers are balanced in terms of gender. The data gathered from Table 5.18 [refer to section 5.4.3(b)], however, shows that in terms of ethnicity made-up of the intermediate performers are prone to Malays. It



should be noted that those performing at the intermediate level are speakers who have difficulty in performing sections A, B, and C of the ESL speaking proficiency test.

**d. Profiles of Novice performers**

The present study is about assessing undergraduates' speaking proficiency. As a matter of fact, undergraduates' speaking proficiency is assessed before embarking on their academic studies at the tertiary level. Therefore, undergraduates are expected to speak and converse in English to some extent as the medium of instruction is English in most of the schools at the tertiary level education. Hence, it was expected that none of the respondents who were undergraduate students to be categorized as novice performers. As such, in this study, the results reveal that none of the university B respondents out of the total number of respondents was categorized as novice performers. The next section presents the speakers' speaking proficiency profile.

**5.4.3 The Speakers' Speaking Proficiency Profile**

Once the profiling process was done, a clear profile of the respondents of both universities began to show. What follows is a summary of the profiles of university A and university B speakers' profiles.

**a. University A Speakers' Speaking Proficiency Profile**

The summary of university A speakers' profiles has been presented in Table 5.17 below.

Table 5.17

*University A Undergraduate Speakers' Profile (N=80 Respondents)*

|                        | Gender                                   | Ethnicity   | School   |
|------------------------|--|---|--|
| Proficiency Level      |  |   |  |
| Superior Performer     | Male= 2(25%)<br>Female= 6(75%)           | Malay= 3(37.5%)<br>Chinese= 2(25%)<br>Indian= 3(38%)            | Humanities Languages, Literacies and Translation<br>2(10%)   |
| Advanced Performer     | Male= 10(33.34%)<br>Female = 20 (66.66%) | Malay= 19(63.33%)<br>Chinese= 5(17%)<br>Indian= 6(20%)          | Educational Studies, Aerospace Engineering, Social Sciences, Arts, Electronics Engineering<br>5(25%)   |
| Intermediate Performer | Male= 15(35.70%)<br>Female= 27(64.30%)   | Malay= 25(59.52%)<br>Chinese= 15 (35.71% %)<br>Indian= 2(4.77%) | Management, Physics, Housing, Building, and Planning, Chemical Engineering, Chemistry, Civil Engineering, Computer Science, Industrial Technology, Mathematics<br>9(45%) |
| Novice Performer       | Male= 0(0%)<br>Female= 0(0%)             | Malay= 0(0%)<br>Chinese= 0(0%)<br>Indian= 0(0%)                 | None (0%)  |

**b. University B Speakers' Speaking Proficiency Profile**

Like university A, the summary of university B speakers' profiles has been presented in Table 5.18 below.

Table 5.18

*University B Undergraduate Speakers' Profile (N=60 Respondents)*

| Proficiency Level      | Gender                                 | Ethnicity   | School  |
|------------------------|--|---|---|
| Superior Performer     | Male= 1(50%)<br>Female= 1(50%)         | Malay= 0(0%)<br>Chinese= 1(50%)<br>Indian= 1(50%)           | None (0%)   |
| Advanced Performer     | Male= 7(33.33%)<br>Female = 14(66.67%) | Malay= 14(66.68%)<br>Chinese= 3(14.31%)<br>Indian= 4(19%)   | None (0%)   |
| Intermediate Performer | Male= 18(48.6%)<br>Female= 19(51.4%)   | Malay= 27(72.98%)<br>Chinese= 8 (21.6%)<br>Indian= 2(5/42%) | Business Management,<br>Computing,<br>Economics and Finance,<br>Education, Government,<br>Quantitative Studies,<br>Social Development<br>7(47%) |
| Novice Performer       | Male= 0(0%)<br>Female= 0(0%)           | Malay= 0(0%)<br>Chinese= 0(0%)<br>Indian= 0(0%)             | None (0%)   |

## 5.5 Conclusion

In general, major findings have been achieved from conducting the mixed methods current study. The results of the present study could identify the undergraduates' speaking proficiency. Further, the type of speaking assessment is not utilized as part of classroom grades, but the emphasis is rather on providing informative feedback in pursuit of students' speaking proficiency improvement. In addition, the gathered data can be viewed as an interpretation of information collected about students' speaking proficiency. More important, with the gathered data, language lecturers and instructors will be able to identify the students' speaking proficiency and discover their students' learning problems and challenges and therefore permit them to design better speaking instruction plans for their students. As such, they would be able to teach their students based on their continuous formative assessment and data analysis and interpretation. Further, the profiles of the undergraduates' ESL speaking proficiency will also be able to help lecturers, English departments, as well as policy

makers, plan speaking proficiency courses for undergraduates with low ESL speaking proficiency performance. Therefore, it can be comfortably concluded that this way of assessment will massively assist language instructors to understand what students need and assist their students to understand what they should learn. The discussion and conclusion of the study findings are revealed in the next chapter.

## CHAPTER 6

### DISCUSSION OF FINDINGS AND CONCLUSION

#### 6.1 Introduction

This chapter presents the discussion of findings and conclusion of the current mixed methods study. The major topics to be discussed in the current chapter include discussion of the key findings, the overview of the study, restatement of the objectives, pedagogical implications of the findings, limitations of the study, recommendations for the stakeholders, recommendations for future research and summary of the chapter. What follows is a detailed elaboration on each topic.

#### 6.2 Discussion of Findings

This study has attempted to develop a valid and reliable ESL speaking proficiency test to accurately assess undergraduates' speaking proficiency. The study has also devised new specific speaking proficiency descriptors in order for language lecturers to identify their students' speaking proficiency, their strengths, and weaknesses in pursuit of betterment. Besides, the current study identified the students' speaking proficiency based on gender, ethnicity, university as well as school. Finally, the speakers' profiles were established. As such, this section provides a detailed discussion about the key findings of the study obtained from the analysis of the gathered data. That is, the key findings of the study are summarised and discussed below. References to previous studies are presented where relevant. This study set out

to answer four research questions. Therefore, what follows are four subsections and each subsection relates to one research question raised in chapter one.

### **6.2.1 Development of ESL Speaking Proficiency Test for Undergraduates**

Speaking proficiency is a significant portion of the curriculum in language teaching and this makes speaking a crucial object of assessment as well. Assessing speaking proficiency of language learners, however, is a difficult task due to different factors (Luoma, 2004). In this study, a new ESL speaking proficiency test was devised to assess the undergraduates' speaking proficiency. The developed speaking test consists of three sections ranging from elementary to advanced levels. That is, the developed speaking test is a more informative test that can be used to show the strengths and weaknesses of the undergraduate students in speaking proficiency. This helps language lecturers to accurately assess their students' speaking proficiency. Besides, the developed test in this study is a generic test that can be utilized to measure undergraduates' speaking proficiency. In doing so, a few steps were taken to set the test as well as assure its validity and reliability (Bachman & Palmer, 1996). They posit that in order to design a test, several points should be specified, namely: the sources of data; the tasks and the testees; ensuring the qualities of usefulness; and its management. As such, the first research question of the study is:

How is an ESL speaking proficiency test for undergraduates developed?

To answer the first research question, the following sub-research questions needed to be investigated:

- a. What are the relevant sources for the construction of the ESL prototype speaking

- proficiency test (PSPT)?
- b. What would be the best possible combination of ESL speaking proficiency questions selected for the ESL speaking proficiency test?
  - c. What is the validity and reliability of the ESL prototype speaking proficiency test?
  - d. What is the time allocated for the ESL speaking proficiency test?

One of the most crucial aspects of test development is the sources used in designing the test. In this regard, the syllabus of the English language courses offered at the School of Languages, Literacies and Translation, Littlewood's Methodological Framework (1981), Kathleen Bardovi-Harlig's operationalizing conversation speech acts (2015), IELTS part B speaking test and MUET past year examinations, Long's Interaction Hypothesis (1981) taken and analysed carefully in developing the ESL prototype speaking proficiency test. In this study, the relevant sources were taken and studied carefully in order to develop the ESL speaking proficiency test specifically for the purpose of this study. This procedure is in line with what Bachman and Palmer (1996) assert about test development. They state that in the process of test development, sources of data should be considered. As such, the researcher in the current study studied the necessary sources for constructing the test items.

Based on the sources of data reviewed and studied in this study, the researcher was able to determine and construct the items of the ESL speaking proficiency test which is constructed from elementary to advanced levels. A similar procedure has been taken as a framework for development and validation of the Diagnostic College English

Speaking Test (DCEST) by Zhao (2011). The framework shows that in developing a speaking test, needs analysis, test design, test piloting and administration should be considered. A similar study conducted by Luoma (2004) who found that test items are important in a test's construct. Therefore, test items should be selected based on the purpose of the test specifications.

In determining the validity of the developed ESL prototype speaking proficiency test, content validity of the test was first investigated by inviting two experts in the related field. The findings of the judgments obtained from the two experts showed the appropriateness of items of the prototype ESL test. When these findings are considered, the content validity of the test is fairly defined as satisfactory. Hence, it can be claimed that the ESL speaking proficiency test possesses an adequate degree of appropriateness for the targeted purpose. This procedure is consistent with Sak (2008) who obtained content validity through experts for validating an English-speaking exam. As it is known, to judge the content validity of a test, the test's content should be appropriate for the intended purpose of assessment (Bachman 1991; Brown, 1996). A similar procedure was followed by Sak (2008). In her study, she conducted content validity to assess a speaking proficiency exam for undergraduates. That was conducted to ascertain the content validity of a speaking exam. In developing a new test, it is a necessary procedure to consider the content validity of the new test. The same procedure was taken by Zhao (2011) in developing CDEST. Zhao obtained that in developing the test, it was necessary to validate the test as it assisted in improving the test items.



This procedure confirms what Zhao (2013) obtained in a study on designing speaking tests about test validation. He claims that test validation is a necessary phase in test developing as it helps the researcher to improve the test design and the quality of the test items and to examine the practicality of the developed test. In this study, however, only two experts were invited to validate the developed test. Therefore, it is recommended that future researchers invite more experts to guarantee the content validity of tests they develop.

Construct validity of the ESL speaking proficiency test was established in which the findings indicate that the test has construct validity. Findings of the construct validity indicate that the developed test was able to distinguish between high-performing students and low-performing students. This is an interesting fact that the test was able to distinguish between different groups of students from different language proficiency levels. This finding shows parallelism with what Fulcher (2003) found who claimed that task difficulty is related to construct validity of a language test.

In developing any new instrument, reliability is usually an aspect it should be considered. Parallel-forms reliability and Cronbach's Alpha are common measurements in judging the reliability of a test. Hence, in this study, Parallel-forms reliability and Cronbach's Alpha were generated to determine the reliability of the developed ESL prototype speaking proficiency test. To assure the reliability of the test and the scores of the respondents' speaking proficiency on the ESL speaking proficiency test, interrater reliability was also generated. As such, in this study, three

measurements were considered in determining the reliability of the developed ESL speaking proficiency test.

The findings of the analysis indicate that both sets of the developed test are identical and therefore can be used interchangeably for measuring the same construct which is speaking proficiency of undergraduate students (Refer to Appendix E). Likewise, the findings of Cronbach's Alpha reveal that both sets of the ESL speaking proficiency test are reliable and hence can be used to measure the speaking proficiency of undergraduate students (Refer to Appendix F and Appendix G).

As the result of the correlation computed using interrater reliability, statistically significant results were achieved since the p-values were significantly high between the two raters. This finding is consistent with the findings obtained by Halleck (1996) and Sak (2008) investigating the inter-rater reliability of raters on speaking proficiency exams. It should be noted that in this study only two raters scored the students' performance on the ESL speaking proficiency test. However, previous research suggests that high correlation is generally obtained when several raters are involved to score performance (Fulcher, 2003). This points to the important role of raters, in particular, trained raters in a speaking test. Therefore, the primary concern of test designers attempting to validate their tests should be to increase the quality of their tests as much as possible by taking the crucial aspects of validity and reliability into their account.

In administering any test, time is also considered as an important factor to determine. During administering the test, the time taken by students of different groups was observed and recorded. Based on the observation, the average time taken by students of different language proficiency levels was considered as the time allocated for the ESL speaking proficiency test. Finally, considering these issues and the fact that the test used in this study is an ESL speaking proficiency test, it can be said that it possesses the quality of content and construct validity to a moderately high degree. As such, this study provides a strong evidence supporting the validity of the ESL speaking proficiency test used in the study to measure undergraduates' speaking proficiency.

### **6.2.2 Development of a Set of Descriptors to Identify the Undergraduates' ESL Speaking Proficiency**

A word or phrase that can be used to describe an item in an information retrieval system can be referred to as descriptor. In this study, speaking descriptors consisted of characteristics of student performance at specific band; and the speaking proficiency descriptors were developed based on the students' performance at each band. As such, the second research question of the present study is:

How are the speaking proficiency descriptors developed to identify the undergraduates' speaking proficiency performance?

To answer the second research question, the following sub-research questions needed to be investigated:

- a. What are the most suitable cut scores for the performance bands?
- b. What are the speaking skills undergraduates have acquired?

In this study, z-scores were used to determine the cut scores. After the prototype

ESL speaking proficiency test was administered, the scores were gathered and broken up into groups in order for the researcher to determine the performance bands based on the cut scores. As such, the cut scores helped in identifying the current speaking proficiency of the respondents at each specific band (Band One to Band Four) determined specifically in the present study. The cut scores for Band One to Band Four were established based on the z-scores which is the most standard type of scores. This procedure can be explicated with the same procedure that Rafieyan (2014) followed. In his study, he utilized z-scores in determining four categories of performers based on cut scores. However, the current study had the advantage of determining the performance bands of the undergraduate students as well. Once the performance bands were determined, the speaking proficiency descriptors were established to identify the students' specific speaking proficiency, their strengths, and weaknesses in terms of speaking proficiency.

The speaking proficiency descriptors in this study provide specific speaking proficiency of the students based on formative assessment. In addition, this provides a starting point for the language learners' speaking proficiency instruction. Each speaking proficiency descriptor is a statement of the specific knowledge and speaking proficiency and each speaking proficiency descriptor shows the ability at each proficiency level. Hence, if a student performs at the superior level, it shows that the student has already learned all the speaking performance descriptors of that specific proficiency level to a great extent. Consequently, the speaking proficiency descriptors could be used as an "all-inclusive" document which provides enough information to

assist language instructors in designing, developing and implementation of the speaking proficiency courses at the undergraduate studies level.

Developing specific speaking proficiency descriptors in this study can be explained through the previous literature providing band descriptors in language teaching and learning. Previous research, reporting systems as well as several English language tests recognized internationally and locally have provided various band descriptors in language learning in general and speaking proficiency specifically such as MUET speaking descriptors, IELTS Speaking band descriptors, ACTFL proficiency guidelines-speaking (Breiner-Sanders et al., 2000), Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEF) (Council of Europe, 2001), just to name a few. In general, language lecturers and instructors can utilize them for the purpose of identifying students' language learning. However, these descriptors are not specific speaking proficiency descriptors and some of them have merely provided general language band descriptors which is too general. In contrast, the current study had the advantage of developing specific speaking proficiency descriptors in such a way that students' specific speaking proficiency can be identified. As such, the students' speaking strengths and weaknesses with regards to their speaking proficiency can also be recognized.

### **6.2.3 Identifying Undergraduates' ESL Speaking Proficiency**

In this study, student assessment is utilized to identify the undergraduate students' speaking proficiency as a continuous formative assessment in pursuit of

betterment. Therefore, the current study identifies the students' speaking proficiency.

As such, the third research question of the present study is:

How do the undergraduates perform on the ESL speaking proficiency test?

To answer the third research question, the following sub-research questions need to be investigated:

- a. What is the respondents' general ESL speaking proficiency?
- b. What is the overall speaking proficiency according to gender, ethnicity, university, and school?

In the discussion of the third research question which explored the respondents' general ESL speaking proficiency, the findings of the study reveal that most of the students are at a lower level of English speaking proficiency at both universities. That is, the findings show that students lack English speaking proficiency to some extent. Nonetheless, none of the students was categorized as novice performers as expected. The fact that none of the students was categorized as novice performers can be explained based on the fact that undergraduate students are expected to possess some English proficiency level before entering the university.

The findings can be explained in view of conducting more on-going formative assessments to monitor undergraduates' speaking proficiency improvement. In other words, the findings of the study support the view that a closer attention should be paid towards speaking proficiency assessment of undergraduates in this context. Ultimately, language lecturers would be able to identify strengths and weaknesses of their students. Having said that, they would be able to adapt their teaching materials based on the

needs of their students. This is based on the view that when this new diagnostic assessment tool is adopted, the researchers stress that it should have a positive impact on the teaching and learning of speaking proficiency in the classroom. In supporting the aforementioned view, previous research has shown that assessment in language teaching should be aligned with its goals and aims so that improvement can take place. This finding can be discussed based on the findings of Tsai and Tsou (2009) who found that adopting a new assessment tool should help to have a positive influence on the process of teaching and learning.

Moreover, this finding confirms the findings of Mewald et al. (2009) who set the specifications of speaking assessment test, E8 Speaking Test Specifications; and the aim of E8 was to recognize the strengths and weaknesses of language learners' speaking proficiency. More important, the outcome of this finding is also important for the students to identify their strengths as well as weaknesses. It also aids the lecturers as well as students' parents in order to be applied as an on-going formative assessment to monitor and enhance students' speaking proficiency.

In fact, there is not much research conducted about gender differences and speaking proficiency on its own, but in a general sense, female students outperform male students in English language ability in all language performance skills (Xiufang, 2013). In this study, this assumption has been tested. Findings of this study indicate that female students outperformed male students at both superior and advanced proficiency levels. This finding is similar to the findings obtained by Gorjian et al. (2011) who found that females outperformed males in their speaking proficiency. This

result also shows parallelism with the findings of Hunter et al. (2005). In their study, they found that female students performed better than male students in their speaking proficiency performance. However, this finding in this present study does not confirm the results found by Koosha et al. (2011). In their study, they found that there is no statistically significant difference between male and female students in their speaking proficiency performance. They believed that both male and female students were exposed to the same materials, the same methods and were usually taught by the same lecturer. Hence, they performed the same in speaking proficiency. This points to the important point that this study calls for further investigation in terms of the relationship between gender and speaking proficiency of students in this context.

The findings of the study indicate ethnicity difference in students' performance. As for superior performers, Indian students performed better compared to Malay and Chinese students. However, Malay students outperformed at the advanced proficiency level. This finding can be explained through the findings of previous research. In this regard, this finding is consistent with the findings obtained by Mahyuddin et al. (2006). They found that Indian students tend to outperform in speaking English, followed by the Chinese and Malay students. This finding is also suggested by Renganathan and Chong (2007) who concluded that Indian students outperformed Chinese and Malay students. The current study falls short of explaining why such difference was found in terms of ethnicity. However, the previous study suggested that it may be due to that fact that Indian students see English language as having a higher status than their own mother tongue. Likewise, being proficient in English, as put forward by Renganathan and Chong (2007) is seen as a good investment to the Indian students. Similarly, with



regards to Chinese students, this may be due to the fact that the Chinese community in Malaysia is more likely to use English more frequently in their daily conversation. Therefore, they are more confident in using the English language (Renganathan & Chong, 2007).

In the discussion of the students' performance in relation to both universities (university A and university B), results of the current study show that students of university A outperformed students of university B. In fact, different universities admit students based on different criteria. This difference can be attributed to acceptance criteria for being admitted to both universities. Outstanding examples of this criteria are interview and MUET results. In university A students are admitted to programmes through interview. In addition, the minimum MUET band requirement is Band Two. In contrast, students are not admitted to university B based on interview. Moreover, the minimum MUET band requirement is Band One at university B. In this study, 69% of university B students had MUET Band Three and below and 31% of the students had MUET Band Four and above. In contrast, only 52% of university A students had MUET Band Three and below while 48% of the students had MUET Band Four and above. This may imply that there is an association between MUET band results and their speaking proficiency, in particular in SPT.

There is generally consistent evidence to suggest that student's field of study can be a determining factor influencing their speaking proficiency performance. For example, students who are majoring the English language are expected to perform better compared to students who are majoring other fields other than English (Bachman

& Palmer, 1996). Therefore, this assumption in this study has been tested whether or not this affects students' speaking proficiency performance in the context of Malaysia. Hence, a descriptive analysis was taken in this study to examine this assumption. The study's results show that students from the School of Languages, Literacies, and Translation as well as students from the school of Humanities who are majoring English language studies performed successfully on the ESL speaking proficiency test compared to students from other schools of the two universities. The reason behind this result is that students from those two schools possess prior knowledge of the English language. Consequently, they outperformed students of other schools at both universities. Therefore, this finding in this study shows that this assumption has been appeared to be true in the context of Malaysia. This finding supports the previous results e.g., the results by Khabbazbashi (2017) who concluded that students majoring English language with prior knowledge tend to perform better in speaking proficiency.

#### **6.2.4 Profiling the Undergraduates' ESL Speaking Proficiency**

In this study, based on the notion that identifying and classifying students' performance proves to be crucial, the students' profiling has been established. Students profiling which consists of gender, ethnicity, and school are used to recognize characteristics of students who are performing well and students who are not performing well at both universities. As such, the last research question in this study is:

What are the profiles of the undergraduates' ESL speaking proficiency?

- a. Who are the superior performers?
- b. Who are the advanced performers?

c. Who are the intermediate performers?

d. Who are the novice performers?

In this study, particular attention was paid to classifying students into four profiles of superior performers, advanced performers, intermediate performers and novice performers as a measure to identify students who are performing well and those who are not performing well. Results of the current study show that superior performers are mostly females and they are from the School of Humanities and the School of Languages, Literacies, and Translation. The reason behind such finding is that students at those schools are basically majoring English language studies. Superior performers are students who are able to perform very well in the three sections of the ESL speaking proficiency test.

Similarly, findings of the present study indicate that most of the advanced performers were female students. Advanced performers are students who are able to perform satisfactorily in the ESL speaking proficiency test. Likewise, intermediate performers are students who have difficulties in performing on the three different sections of the ESL speaking proficiency test. Lastly, novice performers are students who can hardly speak English and can hardly perform on the ESL speaking proficiency test. This finding concurred with the results of various studies which also found that female students performed higher than male students in performing speaking proficiency (Gorjian et al., 2011; Hunter et al., 2005).

A similar study has been conducted by Littlewood and Liu (1996) who profiled students entering universities. The study was a two-year study profiling entrants

comprehensibly so that prior information can be gathered for the purpose of future intervention. Similarly, this method tends to support the previous work by Evans and Green (2007) who profiled tertiary students identifying their needs of English language for academic purposes.

The results of profiling the undergraduate students' ESL speaking proficiency provide crucial information for pedagogical improvement of speaking proficiency of the language learners and a clear image of their English language speaking proficiency. This method is particularly important so that strengths and weaknesses of students can be measured for the purpose of future improvement. The findings of the study show much difference in gender and ethnicity. However, the study falls short of explaining why such difference happened in gender and ethnicity. The results of the study would be a useful basis in the future to identify groups of students with different speaking proficiency and thus enable lecturers to tackle them for better performance in speaking. Other factors should also be explored in order to identify the reasons behind the students' low speaking proficiency.

### **6.3 Overview of the Study**

The current study has investigated the undergraduates' speaking proficiency in pursuit of improvement. Oosterhof (2001) states that students are usually assessed after a course of study which indicates solely what they have achieved at the end of the course. Nonetheless, speaking proficiency assessment is dynamic. That is, speaking proficiency assessment of language learners should be based on continuous formative assessment so that improvement can be monitored and guaranteed (Masters, 2015). To

this end, a valid and reliable ESL speaking proficiency test was developed. The present study has also devised new specific speaking proficiency descriptors in order for language instructors to use them to identify the students' specific speaking proficiency, their strengths as well as weaknesses in pursuit of betterment. In this respect, providing the ESL language lecturers with a set of speaking proficiency descriptors that consists of specific information about what students are able to do and are not able to do would be crucially significant. Besides, identifying undergraduates' speaking proficiency based on gender, ethnicity, university, and school was also conducted. By identifying students at the group level, I was able to profile the groups into Superior Performers, Advanced Performers, Intermediate Performers, and Novice Performers.

The main research instrument of the study was an ESL speaking proficiency test (SPT). The SPT was complimented with a section on the students' biodata. The administration of the SPT consisted of 140 first-year undergraduates from different schools at two public universities in Malaysia. The gathered data was used to identify and profile the students' speaking proficiency. Eventually, by conducting the current mixed-methods study, the researcher was able to:

- a. Develop an ESL speaking proficiency test;
- b. Establish the speaking proficiency descriptors;
- c. Identify the students' speaking proficiency and;
- d. Profile the speakers of different speaking proficiency

#### **6.4 Restatement of the Objectives**

As mentioned earlier in the introductory chapter, this study followed four major objectives. The first objective sought to develop an ESL speaking proficiency test for undergraduates by selecting the relevant sources for the construction of the ESL prototype speaking proficiency test, determining the best combination of questions for the test, conducting a pilot study to test the validity and reliability of the ESL speaking test, as well as determining the time allocated for the ESL speaking test. The second objective sought to develop a set of descriptors to identify the undergraduates' speaking proficiency by determining the best cut scores for the performance bands and identifying speaking skills the undergraduates have acquired. The third objective of the study sought to identify the undergraduates' speaking proficiency based on gender, ethnicity, university, and school. The final objective of the study sought to profile the undergraduates' speaking proficiency into Superior Performers, Advanced Performers, Intermediate Performer, and Novice Performers.

#### **6.5 Pedagogical Implications of the Findings**

The issue is whether the current assessment is able to pinpoint language learners' strengths and weaknesses pertaining to their speaking proficiency. Therefore, the main thrust in this study was assessing undergraduates' speaking proficiency based on formative assessment in pursuit of betterment and improvement. As such, it is reasonable to state that there is a need to refine the speaking proficiency syllabus based on the students' needs and to link assessment more closely to learning objectives so that necessary actions can be taken into consideration to achieve quality learning in terms of speaking proficiency of undergraduates.

The results of the study indicate that undergraduate students can generally speak English. However, their speaking proficiency needs to be enhanced based on ongoing and meaningful formative assessment where the assessment will be able to inform future decision and provision for teaching and learning. To the best knowledge of the researcher, speaking assessment should be dynamic (on-going formative assessment) to recognize the current level of students in pursuit of improvement. As such, the findings of the current study suggest that universities introduce new speaking proficiency courses at the undergraduate studies level so that they will be able to improve their students' speaking proficiency accordingly.

Language lecturers and instructors can utilize the speaking rubric in speaking courses to assign expectations in the beginning of their speaking instruction to provide feedback on student betterment. This appears consistent with previous research, demonstrating that providing learners with rubric helps to enhance language learning, because rubric recognizes areas for improvement in teaching (Fleming, 2001; Gschwend, 2000; Song, 2006). Although language lecturers may see rubrics merely as tools for marking and assessment, students believe that rubrics assist them in improvement (Reddy & Andrade, 2010).

A further implication of this research is that lecturers can tailor their instructional materials needed based on the students' weaknesses in areas of speaking in universities because the speaking rubric and descriptors serve as a powerful analytical instrument to assess the effectiveness of their instructional strategies for what their students have or have not achieved. Last, language lecturers might use the method in the current study

to the continuous explication on developing assessment rubrics to increase the marking reliability. The method utilized in this research has provided a practicable model for devising of new assessment scales, which can serve as a prototype model for future scale designers.

## **6.6 Limitations of the Study**

Although the current study was conducted carefully and provided useful information on undergraduates' speaking proficiency, it holds a number of limitations imposed by the study design as well as the availability of the data that are necessary to be addressed. These limitations must be kept in mind when considering how these findings might apply to larger contexts.

First, it is important to note that the scope of this study was limited to assessing, identifying as well as profiling the ESL undergraduates' speaking proficiency of two universities only, involving male and female students from different ethnic groups. Having said that, it would have been broad enough if data were collected from other universities. Therefore, the researcher is prevented from drawing any broad conclusions or generalization of the findings of the study.

Second, the present study only focused on undergraduate students. Therefore, future research would target and involve postgraduate students to explore fully to the speaking proficiency of language learners based on the on-going formative assessment to meet its intended goals and objectives. Hence, the findings of the study are considered to be interpreted with caution.



Third, the data collected from the respective schools included only four students from each school. This concern should be kept in mind as it makes the data insufficient to establish a well representative sample of the population. As such, the findings of this study can only be specifically utilized within the context of this study because it does not represent the larger context of the undergraduates' speaking proficiency performance. Therefore, collecting a larger sample of students from each school would be a better representative of the schools in terms of their performance.

The last limitation concerns the difficulties of using video recording during the data collection process. This limitation was due to concerns related to showing students' face during recording, and it was a recognized limitation in research of this nature. Therefore, the researcher faced difficulties in this regard during the data collection process.

## **6.7 Recommendations for Stakeholders**

The findings of the current study achieved through the quantitative and qualitative data suggest some recommendations for the following stakeholders: Malaysian Ministry of Higher Education, universities, lecturers of English language, and parents of English language learners.

Malaysian Ministry of Higher Education is recommended to necessitate introducing new speaking courses at the tertiary level so undergraduates can improve their speaking proficiency more effectively and powerfully. Likewise, universities are recommended to necessitate the use of the speaking proficiency descriptors developed

in the present study in order for them to identify the strengths, weaknesses as well as needs of the undergraduates at the beginning of their studies in pursuit of betterment.

Lecturers of English language are then recommended to identify the language learners' speaking proficiency by administering the ESL speaking proficiency test prior to instruction. This could be considered as a diagnostic test where language lecturers are able to find out what their students can do and what they cannot do. The results should be utilized to provide clear and specific feedback to the lecturers so they will be able to assist their students by adapting their instruction and teaching materials to enhance the language learners' speaking proficiency based on their current needs. That is achieved based on effective assessment. As such, by providing effective assessment, instruction could be more effective and would meet the objectives of optimal learning. Proper assessment can provide rich information and feedback about students' achievement and inform students how to improve (Stiggins, 2006). This is obtained merely through the concept of *assessment for learning* as it helps in informing instructional decisions.

The specific speaking proficiency descriptors could further offer the language lecturers with a clearer image of the specific abilities of the students in the different performance bands (Band One to Band Four) in terms of the sub-skills of speaking proficiency.

The identifying results can be used to categorize the respondents according to the different categories based on their performance. As such, by identifying the profiles of

Superior Performers, Advanced Performers, Intermediate Performers, and Novice Performers, universities and relevant authorities might want to plan what needs to be done to improve the ESL speaking standards in Malaysia.

Similarly, providing parents with such detailed information through an optimal reporting system regarding the students' speaking proficiency might provide detailed information about the progress of the students' English language speaking proficiency. Consequently, parents will be fully aware of the students' speaking proficiency.

## **6.8 Recommendations for Further Research**

Based on the findings and discussion of this study, the researcher proposes suggestions for further studies in speaking assessment. Perhaps, future research could be done in the other universities because the current study was merely restricted to undergraduates of two public universities. As such, future research in the field of speaking assessment could be considered within a state or even at the national level so that the generation of the results could be guaranteed. This would also provide the chance to test the adaptability of the ESL speaking proficiency test in different contexts.

Further, the current study only focused on undergraduate students. Thus, researchers in the relevant field might want to include postgraduate students to investigate the full understanding of speaking proficiency of language learners based on formative assessment to meet the intended goals and objectives of the study.

Due to time constraints, only four students were taken from each school of the two public universities. Hence, further investigation could also consider including a larger sample size of students from each respective school. As such, future studies can draw generalization of findings regarding undergraduates' speaking proficiency performance.

Given the significance of these issues, future studies will lead to a better understanding of proper speaking assessment and better practice of measuring the construct (speaking proficiency) in the context of English speaking assessment in Malaysia. Thus, it is hoped that this research can be used as one of the references for other researchers who conduct the similar research studies related to the improvement of the students' speaking proficiency based on conducting proper assessment.

## **6.9 Summary**

The main reason behind conducting the current study was the unhappiness in the way speaking assessment is conducted at the tertiary level. The assessment has been solely for the purpose of evaluation. Therefore, we are not able to identify the strengths and weaknesses of the students so we can assist them in enhancing their speaking proficiency.

Speaking proficiency assessment is not static, but it is rather dynamic (on-going formative assessment) because it can be improved by providing proper instruction. Additionally, language instructors can use the ESL speaking test to gather information on the current status of their students' speaking proficiency. As such, students can also

be aware of their own speaking proficiency, strengths, and weaknesses with the descriptors provided.

The gathered data collected for the current study allowed me to establish the performance bands to display the speaking proficiency of the students. Likewise, the performance bands permitted me to see the speaking proficiency of the respondents whereby I could assess the respondents' proficiency level at Superior, Advanced, Intermediate and Novice Performers. From the practical perspective, the current framework established in the research can assist as a basis for English language lectures to develop more diagnostic tests and systems based on the curricular goals and teaching objectives at the tertiary level in Malaysia.

Finally, it is important to acknowledge that the ESL speaking proficiency test, the performance bands, and speaking proficiency descriptors may not be faultless. Nevertheless, the test that has been developed showed that far more information can be attained and it can be accurate enough in providing useful information about students' current speaking proficiency. Ultimately, the researcher is content that this piece of work has been fruitful and a satisfying attempt.

## REFERENCES

- Abdullah, K. I., & Rahman, N. L. A. (2010). *A study on second language speaking anxiety among UTM students*. Fakulti Pendidikan, Universiti Teknologi Malaysia. Retrieved from [http://eprints.utm.my/10275/2/Nurul\\_Lina\\_Bt\\_Abdul\\_Rahman.pdf](http://eprints.utm.my/10275/2/Nurul_Lina_Bt_Abdul_Rahman.pdf)
- ACTFL (1999). *The ACTFL Proficiency Guidelines: Speaking (Revised, 1999)*. Yonkers, NY: ACTFL.
- Afflerbach, P., Cho, B.-Y., Kim, J.-Y., Crassas, M. E., & Doyle, B. (2013). Reading: What else matters besides strategies and skills? *The Reading Teacher*, 66(6), 440– 448. <http://dx.doi.org/10.1002/TRTR.1146>
- Alberola Colomar, M. P. (2014). A classroom-based assessment method to test speaking skills in English for Specific Purposes. *Language Learning in Higher Education*, 4(1), 9-26.
- Alderson, J. C. (2007). The CEFR and the need for more research. *Modern Language Journal*, 91, 659–663.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Allen, M. J. (2004). *Assessing academic programs in higher education*. Bolton: Anker.
- Ambrose, S. A., Bridges, M. W., Dipietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. San Fransisco, CA: Jossey-Bass.
- American Council on the Teaching of Foreign Languages (ACTFL). (2006). *Standards for foreign language learning in the 21st century*. Lawrence, KS: Allen Press.

- Azman, H., & Razak, N. A. (2007). Change in the language of instruction policy and representation of readiness among tertiary educators in Malaysia. In R. Azman, S. Lee & N. M. Noor (Eds.), *Transforming learning realities in the ELT world*, (pp.38-54). Petaling Jaya: Pearson Malaysia Sdn. Bhd.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Bailey, K. M., & Savage, L. (1994). *New Ways in Teaching Speaking. New Ways in TESOL Series: Innovative Classroom Techniques*. Alexandria, VA: TESOL.
- Baker, J., & Westrup, H. (2003). *Essential speaking skills: A handbook for English language teachers*. London, UK: Continuum.
- Banta, T. W. (2002). *Building a scholarship of assessment*. San Francisco, CA: Jossey-Boss.
- Barvdovi-Halrig, K. (2015). Operationalizing conversation in studies of instructional effect in L2 pragmatics. *System*, 48(1), 21-34.
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85(5), 536–553.
- Birjandi, P., Bagheridoust, E., & Mossalanejad, P. (2004). *Language testing, A concise collection for graduate applicants* (3<sup>rd</sup> ed.). Tehran, Iran: Shahid Mahdavi Publications.
- Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research and Evaluation*, 8(9). Retrieved from <http://pareonline.net/getvn.asp?v=8&n=9>

- Bransford, J. D., Brown, A., & Cocking, R. (1999). *How people learn: Brain, mind experience and school*. Washington, DC: National Academy Press.
- Breiner-Sanders, K., Lowe, P., Miles, J., & Swender, E. (2000). ACTFL Proficiency Guidelines-Speaking, (Revised, 1999). *Foreign Language Annals* 33, 13–18.
- Bresciani, M. J., Gardner, M. M., & Hickmott, J. (2009). *Demonstrating student success: A practical guide to outcomes-based assessment of learning and development in student affairs*. Sterling, VA: Stylus.
- Britton, T. (2011). Using formative and alternative assessments to support instruction and measure student learning. *Science Scope*, 34(5), 16-21.
- Brookhart, S. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12.
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ, USA: Prentice Hall Regents.
- Brown, D. (2001). *Teaching by principles: An interactive approach to language pedagogy* (2<sup>nd</sup> ed.). USA: Longman.
- Brown, J. D. (2005). *Testing in Language Programs: A Comprehensive Guide to English Language Assessment*: McGraw-Hill.
- Brown, S. (2004). Assessment for learning. *Learning and teaching in higher education*, 1(1), 81-89.
- Burke, K. (Ed.) (1992). *Authentic Assessment: A Collection*: Arlington Heights IRI/SkyLight Training and Publishing, Inc.
- Burns, A., & Joyce, H. (1997). *Focus on Speaking*. Sydney: National Centre for English Language Teaching and Research.



- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 2-27). New York: Longman Press.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Castejón, J. I., Gilar, R., Minano, P., & González, M. (2016). Latent class cluster analysis in exploring different profiles of gifted and talented students. *Learning and Individual Differences*, 50, 166-174.  
<http://doi.org/10.106/j.lindif.2016.08.003>
- Cavanagh S. (1997) Content analysis: concepts, methods and applications. *Nurse Researcher* 4(3), 5–16.
- Cenoz, J., & Jessner, U. (2000). *English in Europe: The acquisition of a third language*. Clevedon: Multilingual Matters.
- Chambers, F., & Richards, B. (1992). Criteria for oral assessment. *Language Learning Journal*, 6(1), 5-9.
- Clark, I. (2008). Assessment is for learning: Formative assessment and positive learning interactions. *Florida Journal of Educational Administration and Policy*, 2(1), 1-16.
- Cohen, A. D., & Shively, R. L. (2007). Acquisition of requests and apologies in Spanish and French: Impact of study abroad and strategy-building intervention. *The Modern Language Journal*, 91(2), 189-212.
- Costel, S., Simona, M., Ana, O., & Stefan, G. (2015). Measures of transforming summative assessment in formative assessment in evaluation students activities in Constanta Maritime University. Considerations about formative feedback. *Universitatii Maritime Constanta Annals*, 16(23), 181-188.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

- Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4<sup>th</sup> ed.). Boston: Pearson.
- Cross, D. (1991). *A Practical handbook of language teaching*. London: Longman.
- Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. E. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing*, 21(2), 107-145.
- Dahan, H. B. A. M. (2012). *Language Testing: The Construction and Validation* (2<sup>nd</sup> ed.). Kuala Lumpur: Malaysia: University of Malaya Press.
- Damro, C. (2015). Market power Europe: exploring a dynamic conceptual framework', *Journal of European Public Policy*, 22(9), 1336-1354. <http://doi.org/10.1080/13501763.2015.1046903>
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.
- Descriptor. (n.d.) In *Merriam-Webster's collegiate dictionary*. Retrieved from <http://www.merriam-webster.com/dictionary/descriptor>
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessments: The limited scientific evidence of the impact of formative assessments in education. *Practical Assessment, Research and Evaluation*, 14(7), 1-11. Retrieved from <http://pareonline.net/getvn.asp?v014&n07>
- Earl L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks: Corwin Press.
- Egan, K.B. (1999). Speaking: A critical skill and a challenge. *Calico Journal*, 16(3), 277-293.
- Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics*, 16(4), 401-449.

- Ellis, R. (1999). *Learning a second language through interaction*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Elo, S., & Kyngas, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. <http://dx.doi.org/10.1111/j.1365-2648.2007.04569.x>
- Eslami, Z., & Liu, C. N. (2013). Learning pragmatics through computer-mediated communication in Taiwan. *International Journal of Society, Culture and Language*, 1(1), 52-73.
- Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, 6(1), 3-17. <http://dx.doi.org/10.1016/j.jeap.2006.11.005>
- Ewell, P. T. (2002). An emerging scholarship: A brief history of assessment. In Trudy, W. Banta (Ed.), *Building a scholarship of assessment*, 3-25. San Francisco: Jossey-Bass.
- Fleming, V. M. (2001). Helping students learn to learn by using a checklist, modified rubrics, and e-mail. *Journal on Excellence in College Teaching*, 12, 5-22.
- Florez, M. A. C. (1999). Improving adult English language learners' speaking skills. *National Clearinghouse for ESL Literacy Education*. Retrieved from <http://www.ericdigests.org/2000-3/adult.htm>
- Fraenkel, J. R., & Wallen, N. E. (2009). *How to design and evaluate research in education* (7<sup>th</sup> ed.). New York: McGraw-Hill.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8<sup>th</sup> ed.). New York: McGraw-Hill.
- Fulcher, G. (2003). *Testing second language speaking*. London: Pearson Longman.
- Fulcher, G. (2012). Assessment Literacy for the Language Classroom. *Journal of Language Assessment Quarterly*, 9(2), 113-132. <http://dx.doi.org/10.1080/15434303.2011.642041>

- Galaczi, E. (2013). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics*, 35(5), 553-574.
- Gavilán, F. (2008). A study of the dominant type of technique (controlled, semi-controlled and free) of two English teachers from a languages teaching program. *Profile Issues in Teachers' Professional Development*, 10(1), 163-180.
- Gay, L. R., & Airasian, P. (2003). *Educational research: Competencies for analysis and application* (7th ed.). Upper Saddle River, New Jersey: Merrill Prentice-Hall.
- Gorjian, B., Moosavinia, R., & Shahramiri, P. (2011). Effects of Oral Summary of Short Stories on Male/Female Learners' Speaking Proficiency: High vs. Low Achievers. *Iranian EFL Journal*, 7, 34-50.
- Gravetter, F. J., & Wallnau, L. B. (2013). *Statistics for the behavioral sciences*. Belmont, CA: Wadsworth Publishing.
- Gschwend, L. (2000). Every student deserves an assessment tool that teaches. *Communication Teacher*, 14(3), 1-5.
- Hadley, G., & Mort, J. (1999). An Investigation of Interrater Reliability in Oral Testing. *Nagoya National College of Technology Journal*, 35(2), 45-51.
- Halleck, G. B. (1996). Interrater reliability of the OPI: Using academic trainee raters. *Foreign Language Annals*, 29(2), 223-238.
- Hazzi, O., & Maldaon, I. (2015). A pilot study: Vital methodological issues. *Business: Theory and Practice*, 16(1), 53-62.
- Henly D. (2003). Use of Web-based formative assessment to support student learning in a metabolism/nutrition unit. *European Journal of Dental Education* 7(3), 116-122.

- Hiew, W. (2012). English language teaching and learning issues in Malaysia: Learners' perceptions via Facebook dialogue journal. *Journal of Arts, Science and Commerce*, 3(1), 11-19.
- Hinkel, E. (2005). *Handbook of research in second language teaching and learning*. Mahwah, NJ: Lawrence Erlbaum.
- House, J. (2002). Communicating in English as a lingua franca. *Eurosla Yearbook*, 2(1), 243-261.
- Howarth, P. (2001). Process speaking 1: Preparing to repeat yourself. *Modern English Teacher Journal*, 10(1), 39-44.
- Hsieh H.F., & Shannon S.E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277-1288.  
<http://dx.doi.org/10.1177/1049732305276687>
- Huba, M.E., & Freed, J.E. (2000). *Learner-centered assessment on college campuses: Shifting the focus from teaching to learning*. Boston: Allyn and Bacon.
- Hulstijn, J. H. (2010). Linking L2 proficiency to L2 acquisition: Opportunities and challenges of profiling research. In I. Bartning, M. Martin & I. Vedder (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp.233–238). EUROSLA Monograph series 1. Retrieved from <http://eurosla.org/monographs/EM01/EM01home.html>
- Hunter, D., Gambell, T., & Randhawa, B. (2005). Gender gaps in group listening and speaking: Issues in social constructivist approaches to teaching and learning. *Educational Review*, 57, 329 – 355.
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29(1), 24-49.
- Jacoby, J., Heugh, S., Bax, C., & Branford-White, C. (2014). Enhancing learning through formative assessment. *Innovations in Education and Teaching International*, 51(1), 72-83. <http://dx.doi.org/10.1080/14703297.2013.771970>

Jaya, P. P. (2003). *An introduction to testing and evaluation*. Kuala Lumpur: English Language Teaching Centre, Kementerian Pelajaran Malaysia.

Khabbazbashi, N. (2017). Topic and background knowledge effects on performance in speaking assessment. *Language Testing*, 34(1), 23-48.

Koosha, B.; Ketabi, S., & Kassaian, Z. (2011). The effects of self-esteem, age and gender on the speaking skills of intermediate university EFL learners. *Theory and Practice in Language Studies*, 1(10), 1328, 1337.  
<http://dx.doi.org/10.4304/tpls.1.10.1328-1337>

Kubiszyn, T., & Borich, G. (2000). *Educational testing and measurement: Classroom application and management* (6<sup>th</sup> ed.). New York, USA: John Wiley and Sons, Inc.

Kubiszyn, T., & Borich, G. (2010). *Educational testing and measurement: classroom application and practice* (9<sup>th</sup> ed.). New Jersey, USA: John Wiley and Sons, INC.

Lan, C. O. T., Khaun, A. L. C., & Singh, P. K. S. (2011). Employer expectations of language at the workplace. *Malaysian Journal Of ELT Research*, 7(2), 82-103.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom Assessment: Minute by Minute, Day by Day. *Educational Leadership*, 63(3), 18-24.

Lemus, C. (2014). Controlled and Free Practice activities. Retrieved from <https://prezi.com/wmhrgrztcv3/controlled-and-free-practice-activities/>

Liao, H., & Hsu, L. (2014). Using an Analytical Rubric to Improve the Writing of EFL College Students. *The Asian TEFL Journal*, 16(1), 296-333.

Linn, R. L., & Miller M. D. (2005). *Measurement and Assessment in Teaching* (9<sup>th</sup> ed.). Upper Saddle River, New Jersey: Merrill Prentice Hall.

Little, D. (2007). The Common European framework of reference for languages: Perspectives on the making of supranational language education policy. *Modern Language Journal*, 91, 645–655.

- Littlewood, W. (1981). *Communicative language teaching: An introduction*: Cambridge University Press.
- Littlewood, W., & Liu, N.-F. (1996). *Hong Kong students and their English: LEAP*. Hong Kong: Macmillan.
- Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *The Modern Language Journal*, 92(1), 71-86.
- Long, M. (1981). Input, interaction and second-language acquisition. *Annals of the New York Academy of Sciences*, 379(1), 259-278.
- Long, M. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics*, 4, 126-141.
- Long, M. (1996). The role of the linguistic environment in second language acquisition. In: Ritchie, C., Bhatia, T. K. eds., *Handbook of Language Acquisition*. Second Language Acquisition. Academic Press, New York, 413-468.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Madsen, H. S. (1983). *Techniques in Testing*. New York: Oxford University Press.
- Mahyuddin, R., Elias, H., Loh S. C., Muhamad, M. F, Noordin, N., & Abdullah, M. C. (2006). The relationship between students' self-efficacy and their English language achievement. *Jurnal Pendidik dan Pendidikan*, 21, 61-71.
- Malaysian Examinations Council (2015). *Malaysian University English Test (MUET): Regulations, test specifications, test format and sample questions*. Selangor, Malaysia. Retrieved from [http://www.mpm.edu.my/download\\_MUET/MUET\\_Test\\_Specification\\_2015VersiPortal.pdf](http://www.mpm.edu.my/download_MUET/MUET_Test_Specification_2015VersiPortal.pdf)

- Masters, G. N. (2013). Reforming educational assessment: Imperatives, principles and challenges. *Australian Council for Educational Research*, 57. Retrieved from <http://research.acer.edu.au/aer/12/>
- Masters, G. N. (2015). Rethinking formative and summative assessment. *Teacher Magazine*. Retrieved from <https://www.teachermagazine.com.au/geoff-masters/article/rethinking-formative-and-summative-assessment>
- Mewald, C., Gassner, O., & Sigott, G. (2009). *E8 Speaking Test Specifications Version 01*. Retrieved from LTC, Klagenfurt: [http://www.uni-klu.ac.at/ltc/downloads/4\\_Speaking\\_Test\\_Specifications\\_October\\_2009.pdf](http://www.uni-klu.ac.at/ltc/downloads/4_Speaking_Test_Specifications_October_2009.pdf)
- Miller, K. S. (1998). Teaching speaking. In K. Johnson & H. Johnson (Eds.), *Encyclopedic dictionary of applied linguistics* (pp. 335-341). Oxford: Blackwell.
- Mok, S.S. (2000). *Ilmu pendidikan untuk KPLI (Kursus Perguruan Lepas Ijazah)*. Subang Jaya: Kumpulan Budiman Sdn. Bhd.
- Morovat, E. (2014). Effects of reflectivity/impulsivity on IELTS candidates' band scores in the speaking module of the test. *Procedia-Social and Behavioral Sciences*, 98, 1232-1239.
- Nakamura, Y. (1993). *Measurement of Japanese college students' English speaking ability in a classroom setting*. Unpublished doctoral dissertation, International Christian University, Tokyo.
- New Zealand Qualifications Authority (NZQA). (2013). New Zealand qualifications framework. Wellington, New Zealand: NZQA. Retrieved from <http://www.nzqa.govt.nz/assets/Studying-in-NZ/New-Zealand-Qualification-Framework/requirements-nzqf.pdf>
- Nopiah, Z. M., Ismail, N. A., Othman, H., Asshaari, I., Razali, N., Othman, M. H., & Jamalludin, M. H. (2011). Identification of student achievement and academic profile in the Linear Algebra course: An analysis using the Rasch model. *20113rd International Congress on Engineering Education (ICEED)*, 197-202. <http://dx.doi.org/10.1109/ICEED.2011.6235389>
- Norris, C.B. (1991). Evaluating English Oral Skills through the Technique of Writing as if Speaking. *System*, 19(3), 203-216.



- Nunan, D. (1991). *Language teaching methodology*. A textbook for teachers. London: Prentice Hall International.
- Nunan, D. (2003). *Practical English language teaching*. New York: McGraw-Hill.
- Oosterhof, A. (2001). *Classroom applications of educational measurement* (3<sup>rd</sup> ed.). Upper Saddle River, New Jersey, USA: Merrill Prentice-Hall.
- O'Sullivan, B. (2008). Notes on assessing speaking. *Cornell University–Language Resource Center*. Retrieved from <http://lrc.cornell.edu/events/past/2008-2009/papers08/osull1.pdf>
- Pead, M. J. (2008). Assessment: Cinderella or sleeping beauty? Evolution of final examinations at the Royal Veterinary College. *Journal of Veterinary Medical Education*, 35(4), 607–611.
- Peat, J. (2001). *Health science research: A handbook of quantitative methods*. Sydney: Allen and Unwin.
- Penuel, W. R., & Shepard, L. A. (2016). Social models of learning and assessment. In A. A. Rupp & J. P. Leighton (Eds.), *Handbook of cognition and assessment: Frameworks, methodologies, and applications*. Hoboken, NJ: John Wiley & Sons.
- Peterson, M. W., Einarson, M. K., Augustine, C. H., & Vaughan, D. S. (1999). *Institutional support for student assessment: Methodology and results of a national survey*. Stanford, CA: National Center for Postsecondary Improvement.
- Prototype. (n.d.) In *Merriam-Webster's collegiate dictionary*. Retrieved from <http://www.merriam-webster.com/dictionary/prototype>
- Rabinni, R. (2002). An Introduction to syllabus design and evaluation. *The Internet TESL Journal*, 8(5). Retrieved from <http://iteslj.org/Articles/Rabinni-Syllabus.html>

- Radzi, M., Hanadi, A., Azmin, A. M., Zolhani, N. H., & Abdul Latif, S. (2007). *Adopting Communicative Language Teaching (CLT) Approach to Enhance Oral Competencies Among Students: Teachers' Attitudes and Beliefs*. In: The Second Biennial International Conference on Teaching and Learning of English in Asia: Exploring New Frontiers (TELiA2). 1–23.
- Rafieyan, V. (2014). *The effect of 'focus on form' versus 'focus on forms' pragmatic instruction on the development of pragmatic comprehension* (Doctoral thesis). Retrieved from Repository USM <http://eprints.usm.my/29924/>.
- Rashid, R.A., Mohamed, S.B., Rahman, M.F.A., & Wan Shamsuddin, S.N. (2017). Developing speaking skills using virtual speaking buddy. *International Journal of Emerging Technologies in Learning*, 12(5), 195-201.
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment and Evaluation in Higher Education*, 35(4), 435–448. <http://dx.doi.org/10.1080/02602930902862859>
- Renganathan, S., & Chong, S. L. (2007). The use of English as a social practice in a multilingual and multi-ethnic community. Conference Proceedings. Paper presented at *International Conference on Social Sciences and Humanities, Bangi, 2007*. Malaysia.
- Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching* (2<sup>nd</sup> ed.). Cambridge, UK: Cambridge University Press.
- Risager, K. (2007). *Language and culture pedagogy: From a national to a transnational paradigm*. Clevedon: Multilingual Matters.
- Sak, G. (2008). *An investigation of the validity and reliability of the speaking exam at a Turkish university*. CiteSeer.
- Sawaki, Y. (2007). Construct validation of analytic rating scales in a speaking assessment: Reporting a score profile and a composite. *Language Testing*, 24(3), 355 – 390.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.

- Shumin, K. (2002). Factors to consider: Developing adult EFL students speaking abilities. J. C. Richards, & W. A. Renandya (Eds.), *Methodology in language teaching* (pp. 204-211). Cambridge: Cambridge University Press.
- Sindermann, G., & Horsella, M. (1989). Strategy markers in writing. *Applied Linguistics*, 10(4), 438–446.
- Song, K. H. (2006). A conceptual model of assessing teaching performance and intellectual development of teacher candidates: A pilot study in the US. *Teaching in Higher Education*, 11(2), 175-190.
- Srikaew, D., Tangdhanakanond, K., & Kanjanawasee, S. (2015). Development of an English speaking skill assessment model for grade 6 students by using portfolio. *Procedia-Social and Behavioral Sciences*, 191, 764-768.
- Stiggins, R. (2006). Assessment for learning: A key to motivation and achievement. *Phi Delta Kappan*, 2(2), 3-19. Retrieved from [http://www.michigan.gov/documents/mde/Kappan\\_Edge\\_Article\\_188578\\_7.pdf](http://www.michigan.gov/documents/mde/Kappan_Edge_Article_188578_7.pdf)
- Subramaniam, P. (2014, December 9). Muhyiddin ‘baffled’ by Malaysians’ poor grasp of English. *Malay Mail Online*. Retrieved from <http://www.themalaymailonline.com/malaysia/article/muhyiddin-baffled-by-malaysians-poor-grasp-of-english>
- Taylor, L. (2011). *Examining speaking: Research and practice in assessing second language speaking*. Cambridge: UCLES/Cambridge University Press.
- Thompson, S. K. (2012). *Sampling* (3<sup>rd</sup> ed.). New Jersey, USA: A John Wiley and Sons, INC.
- Thornbury, S. (2005). *How to teach speaking*: London: Pearson Longman.
- Tsai, Y., & Tsou, C. (2009). A standardized English Language Proficiency test as the graduation benchmark: student perspectives on its application in higher education. *Assessment in Education: Principles, Policy and Practice*, 16(3), 319–330.

- Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques*: Cambridge, UK: Cambridge University Press.
- Ur, P. (1996). *A course in language teaching: Practice and theory*. Cambridge, UK: Cambridge University Press.
- Van Teijlingen E. R., & Hundley V. (2001). The Importance of Pilot Studies. *Social Research Update*, (35). Retrieved from <http://sru.soc.surrey.ac.uk/SRU35.html>
- Velan G., Kumar R., Dziegielewski M., & Wakefield D. (2002) Web-based assessments in pathology with Questionmark Perception. *Pathology* 34(3), 282–284.
- Walvoord, B. E. (2004). *Assessment clear and simple: A practical guide for institutions, departments and general education*. San Francisco: Jossey-Bass.
- Wang, K. H., Wang, T. H., Wang, W. L., & Huang, S. C. (2006). Learning styles and formative assessment strategy: Enhancing student achievement in web-based learning. *Journal of Computer Assisted Learning*, 22(3), 207–217.
- Wang, Q., & Castro, C. D. (2010). Classroom interaction and language output. *English Language Teaching*, 3(2), 175-186.
- Wellington, J. (2015). *Educational Research: Contemporary Issues and Practical Approaches* (2<sup>nd</sup> ed.). London, UK: Bloomsbury Academic.
- White, E. (2009). Are you assessment literate? Some Fundamental Questions Regarding Effective Classroom-based Assessment. *OnCUE Journal*, 3(1), 3-25.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>.
- William, D. (2013). Assessment: The bridge between teaching and learning. *Voices from the Middle*, 21(2), 15 – 20.

- Wong, M. W. (2014, December 12). Education Minister's blind spot on English baffling. *The Heat Malaysia*. Retrieved from <http://www.theheatmalaysia.com/VIEWS/Education-Minister-s-blind-spot-on-English-baffling/>
- Xiufang, X. (2013). Gender Differences in Using Language. *Theory and Practice in Language Studies*, 3(8), 1485-1489.
- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., ...Furtak, E. M. (2008). On the impact of formative assessment on student motivation, achievement and conceptual change. *Applied Measurement in Education*, 21(4), 335–359.
- Zhao, Z. B. (2011). *Development and Validation of the Diagnostic College English Speaking Test* (Unpublished doctoral dissertation, Shanghai Jiao Tong University).
- Zhao, Z. (2013). An Overview of Models of Speaking Performance and Its Implications for the Development of Procedural Framework for Diagnostic Speaking Tests. *International Education Studies*, 6(3), 66-75.

## APPENDICES

### Appendix A

#### Structured Interview Questions

1. Do you understand the questions in Section A, B, and C? Yes/no... Elaborate.
2. What are the difficulties you encountered when answering the questions in Section A, B, and C?
3. Which type of questions do you prefer for speaking tests? Individual or group discussion? Give reasons.

## Appendix B

### Respondents' Biographical Information

Dear respondent,

I am conducting a Ph.D. study on speaking assessment. I would be very much grateful if you could help by filling out this form and answering the questions of section (A) on the next page regarding my research. You can put a tick mark (√) to indicate your chosen option in the appropriate columns below.

Thank you for your kind cooperation.

Karwan Mustafa Saeed  
PhD Candidate  
School of Educational Studies, USM

---

I have read the above instructions and I understand all of the conditions. I freely give consent and voluntarily agree to participate in the above aspects of this study. I understand that my identity will be protected and that all records will be coded to guarantee anonymity; all data including recordings will be used only for research purposes.

Respondent's signature -----

|                       |                                |                                  |                                 |                                |
|-----------------------|--------------------------------|----------------------------------|---------------------------------|--------------------------------|
| Identification Number |                                |                                  |                                 |                                |
| Gender                | Male <input type="checkbox"/>  | Female <input type="checkbox"/>  |                                 |                                |
| Ethnicity             | Malay <input type="checkbox"/> | Chinese <input type="checkbox"/> | Indian <input type="checkbox"/> | Other <input type="checkbox"/> |
| Name of School        | <input type="text"/>           |                                  |                                 |                                |
| MUET result           | <input type="text"/>           |                                  |                                 |                                |

## Appendix C

### Permission to Conduct Pilot Study



**Pusat Pengajian Ilmu Pendidikan**  
School of Educational Studies

Universiti Sains Malaysia  
11800 USM Pulau Pinang  
Tel.:(6)04-653 3888 ext.3235/2572,  
(6)04-653 3255 (Direct);  
Fax.:(6)04-657 2907;  
Email : dean\_edu@usm.my  
Website : education.usm.my

**Date : 1 February, 2016**

**Dr. Thomas Chow Voon Foo**  
Deputy Dean (Academic)  
School of Languages, Literacies and Translation  
Greetings,

**Permission to conduct a speaking proficiency test for a thesis, PhD in TESOL programme at the School of Educational Studies, USM**

**Karwan Mustafa Saeed, matric number P-PD0056/14(R)**

With regard to the above, I would like to kindly request your assistance and approval in allowing the above named PhD candidate to conduct a speaking proficiency test involving undergraduates of USM who are currently enrolled in the following English courses organised by the School of Languages, Literacies and Translation, USM, 24 students in each course.

LMT 100: Preparatory English  
LSP 300: Academic English  
LSP 403: Business and Communication English  
LHP 456: Spoken English

The above mentioned student is doing a PhD study and is required to collect data for a pilot study. It is hoped that the outcome of the study would be beneficial and useful for both the said student and also the school of Languages, Literacies and Translation.

Your cooperation and understanding in this matter are most appreciated.

Thank you.

Yours sincerely,

**[PROFESOR DR. MUHAMMAD KAMARUL KABILAN ABDULLAH]**  
Deputy Dean (Research)

**PROF. MADYA DR. SHAIK ABDUL MALIK MOHAMED ISMAIL**  
Timbalan De'kan (Akademik)  
Pusat Pengajian Ilmu Pendidikan  
Universiti Sains Malaysia



## **Appendix D**

Content Validity for ESL Prototype Speaking Proficiency Test

Dear.....,

Greetings,

In this study, the researcher attempts to develop an ESL prototype speaking proficiency test to assess the ESL language learners' speaking proficiency. Expected respondents will be the first-year undergraduate students.

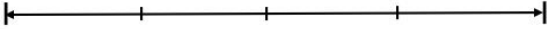
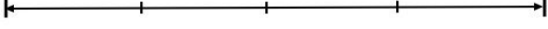
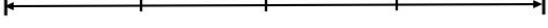
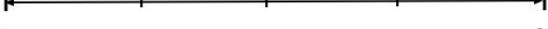

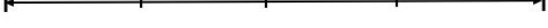






Enclosed is an ESL prototype speaking proficiency test which consists of three sections namely; A, B and C, ranging from elementary to advanced levels. In Section A, respondents are given 10 situations where they need to write down the answers accordingly. While, in Section B, respondents are given a job application brochure (a flyer) and then they are asked 5 questions that they need to answer verbally. Lastly, in Section C, in groups of 4, respondents need to conduct a group discussion where a question is given and therefore, each respondent will play a role.

This speaking proficiency test will be used to assess the speaking proficiency of the first-year undergraduates in order for the researcher to collect the necessary data for the study. Please examine the items of the test based on the scale of 1 (least appropriate) to 5 (most appropriate) to show that the items can be used to assess the said skill. Then please express your comments to improve the items.

Sincerely Yours,

Karwan Mustafa Saeed  
Ph.D. candidate, TESOL

Evaluation of Speaking Proficiency Test Items (Continued)

| Section A | Evaluation   | Comment |
|-----------|--|---------|
| 01        | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p>   |         |
| 02        | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p>   |         |
| 03        | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p>   |         |
| 04        | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p>   |         |
| 05        | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p>   |         |
| 06        | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p> |         |
| 07        | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p> |         |
| 08        | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p> |         |
| 09        | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p> |         |
| 10        | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p> |         |
| Section B | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p> |         |
| Section C | <p style="text-align: center;">1                      2                      3                      4                      5</p>  <p>Least appropriate <span style="float: right;">Most appropriate</span></p> |         |

## Appendix E

### Parallel-Forms Reliability

[DataSet1] C:\Users\Karwan\Desktop\Pilot Study\Parallel-forms reliability.sav

#### *Descriptive statistics*

|       | Mean  | Std. Deviation | N  |
|-------|-------|----------------|----|
| Set 1 | 79.00 | 10.660         | 12 |
| Set 2 | 77.08 | 11.712         | 12 |

#### *Correlations*

|                           | Set 1 | Set 2 |
|---------------------------|-------|-------|
| Set 1 Pearson Correlation | 1     | .807* |
| Sig. (2-tailed)           |       | .002  |
| N                         | 12    | 12    |
| Set 2 Pearson Correlation | .807* | 1     |
| Sig. (2-tailed)           | .002  |       |
| N                         | 12    | 12    |

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

## Appendix F

Reliability Coefficient for Set One of ESL Prototype Speaking Proficiency Test

[DataSet] C:\Users\Karwan\Desktop\Pilot Study\Reliability of ESL Prototype Speaking Proficiency Test Alpha.sav

Scale: Reliability Coefficient of ESL Prototype Speaking Proficiency Test

### *Case Processing Summary*

|       |                       | N  | %     |
|-------|-----------------------|----|-------|
| Cases | Valid                 | 48 | 100.0 |
|       | Excluded <sup>a</sup> | 0  | .0    |
|       | Total                 | 48 | 100.0 |

a. Listwise deletion based on all variables in the procedure

### *Reliability Statistics*

| Cronbach's Alpha | N of Items |
|------------------|------------|
| .83              | 12         |

### *Item-Total Statistics*

|     | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|-----|----------------------------|--------------------------------|----------------------------------|----------------------------------|
| A1  | 164.52                     | 886.340                        | .543                             | .814                             |
| A2  | 164.23                     | 926.436                        | .312                             | .830                             |
| A3  | 164.60                     | 863.266                        | .558                             | .812                             |
| A4  | 164.42                     | 886.716                        | .365                             | .828                             |
| A5  | 164.06                     | 821.336                        | .551                             | .812                             |
| A6  | 164.15                     | 831.191                        | .572                             | .810                             |
| A7  | 164.06                     | 824.230                        | .548                             | .812                             |
| A8  | 164.42                     | 853.397                        | .524                             | .814                             |
| A9  | 165.17                     | 822.652                        | .626                             | .805                             |
| A10 | 165.25                     | 866.574                        | .391                             | .827                             |
| B   | 167.02                     | 929.425                        | .464                             | .821                             |
| C   | 167.50                     | 927.702                        | .526                             | .819                             |

## Appendix G

Reliability Coefficient for Set Two of ESL Prototype Speaking Proficiency Test

[DataSet] C:\Users\Karwan\Desktop\Pilot Study\Reliability of ESL Prototype Speaking Proficiency Test Alpha.sav

Scale: Reliability Coefficient of ESL Prototype Speaking Proficiency Test

### *Case Processing Summary*

|       |                       | N  | %     |
|-------|-----------------------|----|-------|
| Cases | Valid                 | 48 | 100.0 |
|       | Excluded <sup>a</sup> | 0  | .0    |
|       | Total                 | 48 | 100.0 |

a. Listwise deletion based on all variables in the procedure

### *Reliability Statistics*

| Cronbach's Alpha | N of Items |
|------------------|------------|
| .82              | 12         |

### *Item-Total Statistics*

|     | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|-----|----------------------------|--------------------------------|----------------------------------|----------------------------------|
| A1  | 159.44                     | 1338.634                       | .375                             | .819                             |
| A2  | 159.17                     | 1298.567                       | .565                             | .804                             |
| A3  | 159.42                     | 1235.057                       | .668                             | .795                             |
| A4  | 159.71                     | 1264.509                       | .656                             | .797                             |
| A5  | 159.67                     | 1303.206                       | .487                             | .810                             |
| A6  | 160.02                     | 1329.383                       | .431                             | .814                             |
| A7  | 158.83                     | 1298.908                       | .359                             | .824                             |
| A8  | 158.15                     | 1254.766                       | .635                             | .798                             |
| A9  | 159.46                     | 1297.998                       | .567                             | .804                             |
| A10 | 159.10                     | 1299.627                       | .498                             | .809                             |
| B   | 160.37                     | 1294.963                       | .393                             | .820                             |
| C   | 161.58                     | 1401.482                       | .223                             | .831                             |

## Appendix H

### Inter-rater Reliability for Section A

[DataSet] C:\Users\karwan\Desktop\PhD\Inter-rater reliability\Data interrater A.sav

#### *Case Processing Summary*

|       |                       | N  | %     |
|-------|-----------------------|----|-------|
| Cases | Valid                 | 96 | 100.0 |
|       | Excluded <sup>a</sup> | 0  | .0    |
|       | Total                 | 96 | 100.0 |

a. Listwise deletion based on all variables in the procedure

#### *Reliability Statistics*

| Cronbach's Alpha | N of Items |
|------------------|------------|
| .81              | 2          |

#### *Intraclass Correlation Coefficient*

|                  | Intraclass Correlation <sup>b</sup> | 95% Confidence Interval |             | F Test with true Value 0 |     |     |      |
|------------------|-------------------------------------|-------------------------|-------------|--------------------------|-----|-----|------|
|                  |                                     | Lower Bound             | Upper Bound | Value                    | df1 | df2 | Sig  |
| Single Measures  | .671 <sup>a</sup>                   | .539                    | .770        | 5.348                    | 95  | 95  | .000 |
| Average Measures | .803 <sup>c</sup>                   | .700                    | .870        | 5.348                    | 95  | 95  | .000 |

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

## Appendix I

### Inter-rater Reliability for Section B

[DataSet] C:\Users\karwan\Desktop\PhD\Inter-rater reliability\Data interrater A.sav

#### *Case Processing Summary*

|       |                       | N  | %     |
|-------|-----------------------|----|-------|
| Cases | Valid                 | 96 | 100.0 |
|       | Excluded <sup>a</sup> | 0  | .0    |
|       | Total                 | 96 | 100.0 |

a. Listwise deletion based on all variables in the procedure

#### *Reliability Statistics*

| Cronbach's Alpha | N of Items |
|------------------|------------|
| .79              | 2          |

#### *Intraclass Correlation Coefficient*

|                  | Intraclass<br>Correlation <sup>b</sup> | 95% Confidence<br>Interval |                | F Test with true Value 0 |     |     |      |
|------------------|--|----------------------------|----------------|--------------------------|-----|-----|------|
|                  |  | Lower<br>Bound             | Upper<br>Bound | Value                    | df1 | df2 | Sig  |
| Single Measures  | .655 <sup>a</sup>                      | .525                       | .756           | 4.839                    | 95  | 95  | .000 |
| Average Measures | .792 <sup>c</sup>                      | .688                       | .861           | 4.839                    | 95  | 95  | .000 |

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

## Appendix J

### Inter-rater Reliability for Section C

#### Reliability

[DataSet] C:\Users\karwan\Desktop\PhD\Inter-rater reliability\Data interrater C.sav

#### *Case Processing Summary*

|       |                       | N  | %     |
|-------|-----------------------|----|-------|
| Cases | Valid                 | 96 | 100.0 |
|       | Excluded <sup>a</sup> | 0  | .0    |
|       | Total                 | 96 | 100.0 |

a. Listwise deletion based on all variables in the procedure

#### *Reliability Statistics*

| Cronbach's Alpha | N of Items |
|------------------|------------|
| .81              | 2          |

#### *Intraclass Correlation Coefficient*

|                  |                   | 95% Confidence Interval |             | F Test with true Value 0 |     |     |      |
|------------------|-------------------|-------------------------|-------------|--------------------------|-----|-----|------|
|                  |                   | Lower Bound             | Upper Bound | Value                    | df1 | df2 | Sig  |
| Single Measures  | .695 <sup>a</sup> | .575                    | .785        | 5.527                    | 95  | 95  | .000 |
| Average Measures | .820 <sup>c</sup> | .730                    | .880        | 5.527                    | 95  | 95  | .000 |

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.



## Appendix K

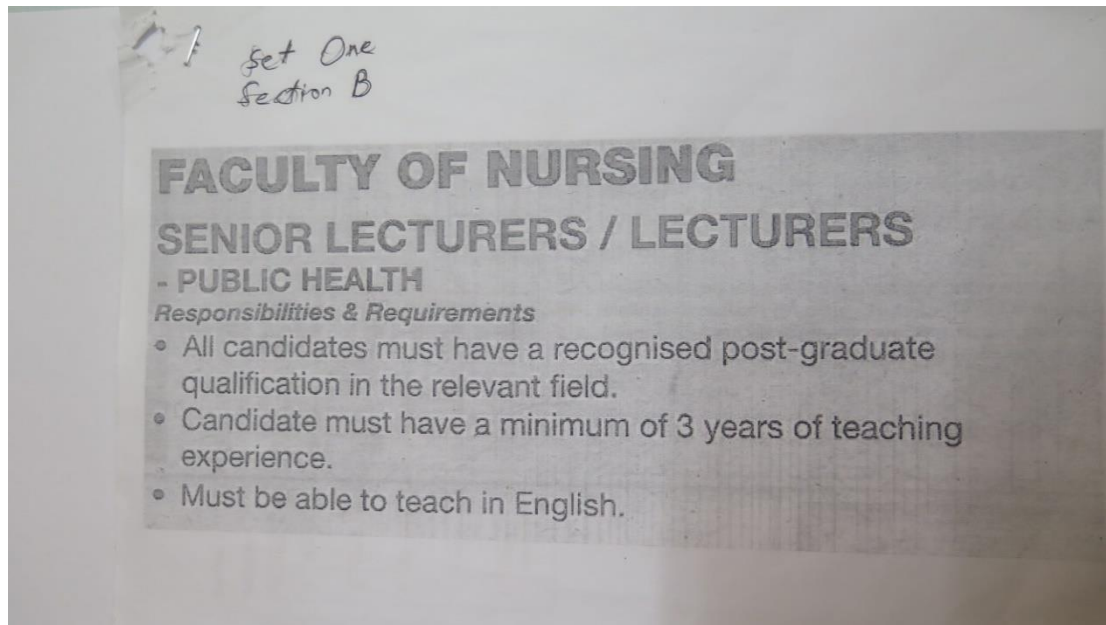
### ESL Speaking Proficiency Tests

#### Set One: Section A

There are ten situations described below. Please read the description of each situation and write down what you would say in that situation.

| No | Item   |
|----|--|
| 1  | You are at your father's office. One of his friends comes over and your father introduces his friend to you. What would you say to your father's friend?<br>.....  |
| 2  | You are a student. You forgot to do the assignment for your English course. Your teacher whom you have known for a while now asks for your assignment. You apologize to your teacher. What do you say to him/her?<br>.....                           |
| 3  | You have a difficult exam tomorrow. You don't understand some of the topics included in the exam. You want to ask one of your friends to help. What do you say to him/her?<br>.....  |
| 4  | It is raining hard and you are walking to school. A friend stops his car to offer you a ride. What would you say to him?<br>.....  |
| 5  | In a group discussion, your class is discussing spending time on Facebook. One of your classmates believes that nowadays people spend much time on Facebook and you have the same opinion as your classmate. What would you say to him/her?<br>..... |
| 6  | You went to see a movie at the cinema at the Queensbay Mall and you loved the movie so much. You believe it was an awesome movie, but your friend, Ali says: The movie was so boring. What would you say to Ali?<br>.....                            |
| 7  | You are trying to apply to do a master's degree in management in the USA. You are required to provide a recommendation letter from one of your professors. What would you say to your professor to write you a letter?<br>.....                      |
| 8  | You go to your school library with several books in your hands. Suddenly, you see a librarian. How do you ask him to help you to open the library door for you?<br>.....   |
| 9  | You need to talk to your lecturer. You go to his office to know if he has time to talk to you. His office door is open. How do you ask him if he has time to talk to you?<br>.....   |
| 10 | You and your friends have been invited by a new friend for dinner. You want to accept your friend's invitation. What would you say to him?<br>.....  |

Section B:



Section B Questions:

---

| No | Item  |
|----|---|
| 1  | How did you know about this teaching job vacancy?                                     |
| 2  | Can you tell me about your qualifications for this job?                               |
| 3  | What is your teaching experience in the relevant field?                               |
| 4  | Can you explain how qualified you are for this job?                                   |
| 5  | What are your salary expectations? What if we can't fulfill your salary expectations? |

---

Section C:

Read the following scenario and discuss the question among you. Each one plays a role in the discussion.

Scenario

---

It has been said that young people in Malaysia today are considered lucky.

---

Which of the following has helped young Malaysians today the most?

Candidate A: They grew in a time of peace and prosperity.

Candidate B: They have easy access to more information.

Candidate C: The government has provided better facilities for sports and recreation.

Candidate D: The education system has offered them more opportunities.

---

Set Two: Section A

There are ten situations described below. Please read the description of each situation and write down what you would say in that situation.

| No | Item   |
|----|--|
| 1  | You are invited by a new classmate for dinner. You want to accept your classmate's invitation. What would you say to him?<br>.....   |
| 2  | You want to talk to your boss. You go to his office to know if he has time to talk to you. His office door is open. What would you say to him?<br>.....                                |
| 3  | You go to your university library with books in your hands. Suddenly, you see a friend. How would you ask him to open the door for you?<br>.....                                       |
| 4  | You are trying to apply a teaching job. You are required to provide a recommendation letter from one of your lecturers. How do you ask your lecturer to provide you a letter?<br>..... |
| 5  | You want to go to Paris for the coming new year vacation, but your friend wants to go to London. You want to disagree with him. What do you say to him?<br>.....                       |
| 6  | In your family, you are discussing moving to a new city. One of your sisters prefers to move to Kuala Lumpur and you have the same opinion. What do you say to her?                    |

- 
- .....
- 7 It is raining heavily and you are walking to your hotel. A classmate stops her car to offer you a ride. What would you say to her?
- .....
- 8 You have an exam next week. You are not clear about some of the topics included in the exam. How do you ask your friend to assist you?
- .....
- 9 You forgot to do the assignment for your geography course. You want to apologize to your teacher. What do you say to him/her?
- .....
- 10 You are at your friend's house. One of his friends comes in and your friend introduces him to you. What do you say to him?
- .....
-

Continued.

Section B:

*Set to a Section B*

# SHARIQ PARTNERS

We are hiring while others are retrenching!!! Job seekers, Shariq Partners (M) Sdn Bhd is one of the most recognized and well respected debt recovery agency in Malaysia. Due to the rapid expansion and growth in our organization, we are seeking suitable candidates preferably with banking experience for the below position:

## CREDIT MANAGEMENT EXECUTIVES / RECOVERY OFFICERS

Currently recruiting at KL (Brem Mall, Kepong), Petaling Jaya (Leisure Commerce Square), Penang (Sunway Carnival Mall, Seberang Jaya) & Johor (Johor Bharu City Square) as:

### Remuneration & Benefits

- ☉ Basic Salary up to RM2,000
- ☉ Unlimited Commission and incentives
- ☉ Medical Insurance Coverage

### Requirements

- ❖ Diploma / Graduate or Minimum SPM with good credits
- ❖ Experience in banking and credit control would be an added advantage
- ❖ Training provided for those without experience
- ❖ Computer Knowledge
- ❖ Must be able to converse in English

Section B Questions:

---

| No | Item   |
|----|--|
| 1  | Why are you applying to work in this company?  |
| 2  | Can you tell me your working experience in management?                               |
| 3  | What are your qualifications?  |
| 4  | What can you do for our company within two years?                                    |
| 5  | What are your salary expectations? What if we can't fulfil your salary expectations? |

---

Continued.

Section C:

Read the following scenario and discuss the question among you. Each one plays a role in the discussion.

Scenario

---

Your brother has been given the opportunity to study in England on a government scholarship. You and your family are very proud of him. Discuss and decide on a special gift for him to prepare for his trip abroad. Discuss which of the following would be most useful for him and why.

---

Candidate A: Cash

Candidate B: A computer

Candidate C: Some suitable clothing

Candidate D: Some Malaysian food stuff

---

## Appendix L

### Speaking Proficiency Descriptors

| Difficulty Levels:<br>Sections | Criteria                  | Band One<br>(Novice Performers)<br>(0-29) raw score   | Band Two<br>(Intermediate Performers)<br>(30-53) raw score   | Band Three<br>(Advanced Performers)<br>(54-77) raw score  | Band Four<br>(Superior Performers)<br>(78-100) raw score  |
|--------------------------------|---------------------------|---|--|---|---|
| A                              | Appropriateness           | <ul style="list-style-type: none"> <li>• He/she can <b>hardly</b> answer in given context appropriately for the intended purpose.</li> <li>• He/she understands questions, but can <b>hardly</b> perform in good command of form and function.</li> </ul> | <ul style="list-style-type: none"> <li>• He/she <b>has difficulty</b> in answering Section A appropriately in given context for the intended purpose.</li> <li>• He/she understands questions but <b>has difficulty</b> in good command of form and function.</li> </ul> | <ul style="list-style-type: none"> <li>• He/she answers in given context for the intended purpose <b>satisfactorily</b>.</li> <li>• He/she understands questions and has a <b>satisfactory</b> command of form and function.</li> </ul>         | <ul style="list-style-type: none"> <li>• He/she answers <b>very well</b> and appropriately in given context for the intended purpose.</li> <li>• He/she understands questions and <b>Very well</b> command of form and function.</li> </ul> |
|                                | Grammar                   | <ul style="list-style-type: none"> <li>• He/she <b>hardly</b> uses accurate and correct grammar.</li> </ul>   | <ul style="list-style-type: none"> <li>• He/she <b>has difficulty</b> in using accurate and correct grammar.</li> </ul>  | <ul style="list-style-type: none"> <li>• He/she uses accurate and correct grammar <b>satisfactorily</b>.</li> </ul>   | <ul style="list-style-type: none"> <li>• He/she uses accurate and correct grammar <b>very well</b>.</li> </ul>  |
| B                              | Appropriateness           | <ul style="list-style-type: none"> <li>• He/she understands questions but <b>hardly</b> speaks appropriately in given context for the intended purpose.</li> <li>• He/she can <b>hardly</b> answer interview questions.</li> </ul>                        | <ul style="list-style-type: none"> <li>• He/she understands questions but <b>has difficulty</b> in speaking appropriately in given context for the intended purpose.</li> <li>• He/she <b>has difficulty</b> in answering interview questions.</li> </ul>                | <ul style="list-style-type: none"> <li>• He/she understands questions and speaks appropriately in given context for the intended purpose <b>satisfactorily</b>.</li> <li>• He/she answers interview questions <b>satisfactorily</b>.</li> </ul> | <ul style="list-style-type: none"> <li>• He/she understands questions and speaks appropriately in given context for the intended purpose <b>very well</b>.</li> <li>• He/she answers interview questions <b>very well</b>.</li> </ul>       |
|                                | Communicative ability     | <ul style="list-style-type: none"> <li>• He/she is <b>hardly</b> able to answer questions meaningfully.</li> <li>• He/she is <b>hardly</b> able to demonstrate well in conveying his/her message.</li> </ul>  | <ul style="list-style-type: none"> <li>• He/she <b>has difficulty</b> to answer questions meaningfully.</li> <li>• He/she <b>has difficulty</b> to demonstrate well in conveying his/her message.</li> </ul>   | <ul style="list-style-type: none"> <li>• He/she is <b>satisfactorily</b> able to answer questions meaningfully.</li> <li>• He/she is able to demonstrate well in conveying his/her message <b>satisfactorily</b>.</li> </ul>                    | <ul style="list-style-type: none"> <li>• He/she is able to answer questions <b>very well</b>.</li> <li>• He/she is able to demonstrate <b>very well</b> in conveying his/her message.</li> </ul>  |
|                                | Fluency and pronunciation | <ul style="list-style-type: none"> <li>• He/she can <b>hardly</b> speak fluently and smoothly.</li> <li>• He/she can <b>hardly</b> speak without any pausing for</li> </ul>   | <ul style="list-style-type: none"> <li>• He/she <b>has difficulty</b> in speaking fluently and smoothly.</li> <li>• He/she <b>has difficulty</b> in</li> </ul>   | <ul style="list-style-type: none"> <li>• He/she speaks fluently and smoothly <b>satisfactorily</b>.</li> <li>• He/she speaks without any pausing for too long</li> </ul>  | <ul style="list-style-type: none"> <li>• He/she speaks fluently and smoothly <b>very well</b>.</li> <li>• He/she speaks without any pausing for too long</li> </ul>   |

|   |                        |  |  |   |  |
|---|------------------------|--|--|---|--|
|   |                        | <ul style="list-style-type: none"> <li>too long.</li> <li>He/she <b>hardly</b> pronounces the individual words correctly.</li> <li>He/she is <b>hardly</b> able to express stress and intonation correctly.</li> </ul>   | <ul style="list-style-type: none"> <li>speaking without any pausing for too long.</li> <li>He/she <b>has difficulty</b> to pronounce the individual words correctly.</li> <li>He/she <b>has difficulty</b> to express stress and intonation correctly.</li> </ul>  | <ul style="list-style-type: none"> <li><b>satisfactorily.</b></li> <li>He/she pronounces the individual words <b>satisfactorily.</b></li> <li>He/she is <b>satisfactorily</b> able to express stress and intonation correctly.</li> </ul>   | <ul style="list-style-type: none"> <li><b>very well.</b></li> <li>He/she pronounces the individual words <b>very well.</b></li> <li>He/she is able to express stress and intonation <b>very well.</b></li> </ul>   |
|   | Grammar and vocabulary | <ul style="list-style-type: none"> <li>He/she <b>hardly</b> uses accurate and correct grammar.</li> <li>He/she <b>hardly</b> uses a range of correct grammatical sentences.</li> <li>He/she <b>hardly</b> uses a wide range of vocabulary effectively.</li> <li>He/she <b>hardly</b> uses appropriate vocabulary.</li> </ul> | <ul style="list-style-type: none"> <li>He/she <b>has difficulty</b> to use accurate and correct grammar.</li> <li>He/she <b>has difficulty</b> to use a range of correct grammatical sentences.</li> <li>He/she <b>has difficulty</b> to use a wide range of vocabulary effectively.</li> <li>He/she <b>has difficulty</b> to use appropriate vocabulary.</li> </ul> | <ul style="list-style-type: none"> <li>He/she uses accurate and correct grammar <b>satisfactorily.</b></li> <li>He/she uses a range of correct grammatical sentences <b>satisfactorily.</b></li> <li>He/she uses a wide range of vocabulary <b>satisfactorily.</b></li> <li>He/she uses appropriate vocabulary <b>satisfactorily</b></li> </ul> | <ul style="list-style-type: none"> <li>He/she uses accurate and correct grammar <b>very well.</b></li> <li>He/she uses a range of correct grammatical sentences <b>very well.</b></li> <li>He/she uses a wide range of vocabulary effectively <b>very well.</b></li> <li>He/she uses appropriate vocabulary <b>very well.</b></li> </ul>             |
| C | Appropriateness        | <ul style="list-style-type: none"> <li>He/she understands questions but <b>hardly</b> speaks appropriately in given context for the intended people.</li> <li>He/she is <b>hardly</b> able to construct ideas in a group discussion.</li> </ul>  | <ul style="list-style-type: none"> <li>He/she understands questions but <b>has difficulty</b> to speak appropriately in given context for the intended people.</li> <li>He/she <b>has difficulty</b> to construct ideas in a group discussion.</li> </ul>  | <ul style="list-style-type: none"> <li>He/she understands questions and speaks appropriately in given context for the intended people <b>satisfactorily.</b></li> <li>He/she is able to construct ideas in a group discussion <b>satisfactorily.</b></li> </ul>   | <ul style="list-style-type: none"> <li>He/she understands questions and speaks appropriately in given context for the intended people <b>very well.</b></li> <li>He/she is able to construct ideas in a group discussion <b>very well.</b></li> </ul>  |
|   | Communicative ability  | <ul style="list-style-type: none"> <li>He/she is <b>hardly</b> able to communicate effectively with the other candidates.</li> <li>He/she is <b>hardly</b> able to demonstrate good interactive ability in carrying out the discussion.</li> <li>He/she is <b>hardly</b> able to maintain eye contact with</li> </ul>        | <ul style="list-style-type: none"> <li>He/she <b>has difficulty</b> to communicate effectively with the other candidates.</li> <li>He/she <b>has difficulty</b> to demonstrate good interactive ability in carrying out the discussion.</li> <li>He/she <b>has difficulty</b> to maintain eye contact with the other candidates skillfully.</li> </ul>               | <ul style="list-style-type: none"> <li>He/she is able to communicate <b>satisfactorily</b> with the other candidates.</li> <li>He/she is able to demonstrate interactive ability in carrying out the discussion <b>satisfactorily.</b></li> <li>He/she is able to maintain eye contact with the other</li> </ul>                                | <ul style="list-style-type: none"> <li>He/she is able to communicate effectively with the other candidates <b>very well.</b></li> <li>He/she is able to demonstrate interactive ability in carrying out the discussion <b>very well.</b></li> <li>He/she is able to maintain eye contact with the other candidates skillfully <b>very</b></li> </ul> |



|                           |  |   |   |  |
|---------------------------|--|---|---|--|
| Managing a discussion     | <p>the other candidates skillfully.</p> <ul style="list-style-type: none"> <li>• He/she is <b>hardly</b> able to develop a discussion and manage it in terms of: <ul style="list-style-type: none"> <li>➤ Initiating</li> <li>➤ Turn-taking</li> <li>➤ Interrupting</li> <li>➤ Negotiating</li> <li>➤ Closing</li> </ul> </li> </ul>                 | <ul style="list-style-type: none"> <li>• He/she <b>has difficulty</b> to develop a discussion and manage it in terms of: <ul style="list-style-type: none"> <li>➤ Initiating</li> <li>➤ Turn-taking</li> <li>➤ Interrupting</li> <li>➤ Negotiating</li> <li>➤ Closing</li> </ul> </li> </ul>  | <p>candidates <b>satisfactorily</b>.</p> <ul style="list-style-type: none"> <li>• He/she is <b>satisfactorily</b> able to develop a discussion and manage it in terms of: <ul style="list-style-type: none"> <li>➤ Initiating</li> <li>➤ Turn-taking</li> <li>➤ Interrupting</li> <li>➤ Negotiating</li> <li>➤ Closing</li> </ul> </li> </ul>                               | <p><b>well</b>.</p> <ul style="list-style-type: none"> <li>• He/she is able to develop a discussion and manage it <b>very well</b> in terms of: <ul style="list-style-type: none"> <li>➤ Initiating</li> <li>➤ Turn-taking</li> <li>➤ Interrupting</li> <li>➤ Negotiating</li> <li>➤ Closing</li> </ul> </li> </ul>                    |
| Fluency and pronunciation | <ul style="list-style-type: none"> <li>• He/she can <b>hardly</b> speak fluently and smoothly.</li> <li>• He/she can <b>hardly</b> speak without any pausing for too long.</li> <li>• He/she <b>hardly</b> pronounces the individual words correctly.</li> <li>• He/she is <b>hardly</b> able to express stress and intonation correctly.</li> </ul> | <ul style="list-style-type: none"> <li>• He/she <b>has difficulty</b> to speak fluently and smoothly.</li> <li>• He/she <b>has difficulty</b> to speak without any pausing for too long.</li> <li>• He/she <b>has difficulty</b> to pronounce the individual words correctly.</li> <li>• He/she <b>has difficulty</b> to express stress and intonation correctly.</li> </ul>                      | <ul style="list-style-type: none"> <li>• He/she speaks fluently and smoothly in a <b>satisfactory</b> way.</li> <li>• He/she speaks without any pausing for too long <b>satisfactorily</b>.</li> <li>• He/she pronounces the individual words <b>satisfactorily</b>.</li> <li>• He/she is <b>satisfactorily</b> able to express stress and intonation correctly.</li> </ul> | <ul style="list-style-type: none"> <li>• He/she speaks fluently and smoothly <b>very well</b>.</li> <li>• He/she speaks without any pausing for too long <b>very well</b>.</li> <li>• He/she pronounces the individual words <b>very well</b>.</li> <li>• He/she is able to express stress and intonation <b>very well</b>.</li> </ul> |
| Grammar and vocabulary    | <ul style="list-style-type: none"> <li>• He/she <b>hardly</b> uses accurate and correct grammar.</li> <li>• He/she <b>hardly</b> uses a range of correct grammatical sentences.</li> <li>• He/she <b>hardly</b> uses a wide range of vocabulary effectively.</li> <li>• He/she <b>hardly</b> uses appropriate vocabulary.</li> </ul>                 | <ul style="list-style-type: none"> <li>• He/she <b>has difficulty</b> to use accurate and correct grammar.</li> <li>• He/she <b>has difficulty</b> to use a range of correct grammatical sentences.</li> <li>• He/she <b>has difficulty</b> to use a wide range of vocabulary effectively (lack of vocabulary).</li> <li>• He/she <b>has difficulty</b> to use appropriate vocabulary.</li> </ul> | <ul style="list-style-type: none"> <li>• He/she uses accurate and correct grammar <b>satisfactorily</b>.</li> <li>• He/she uses a range of correct grammatical sentences <b>satisfactorily</b>.</li> <li>• He/she uses a wide range of vocabulary <b>satisfactorily</b>.</li> <li>• He/she uses appropriate vocabulary <b>satisfactorily</b>.</li> </ul>                    | <ul style="list-style-type: none"> <li>• He/she uses accurate and correct grammar <b>very well</b>.</li> <li>• He/she uses a range of correct grammatical sentences <b>very well</b>.</li> <li>• He/she uses a wide range of vocabulary <b>very well</b>.</li> <li>• He/she uses appropriate vocabulary <b>very well</b>.</li> </ul>   |

---