Assessing EFL Speaking Skills in Vietnamese Tertiary Education

by
Thanh Nam Lam
B.A. (EFL Pedagogy), M.A. (TESOL)

A dissertation submitted in fulfilment of the requirements for
the Degree of Doctor of Philosophy (Education)

Supervisors:
Professor James Albright
Professor John Fischetti
Professor Greg Preston

School of Education
Faculty of Education and Arts

October 2018
STATEMENT OF ORIGINALITY

I hereby certify that the work embodied in the thesis is my own work, conducted under normal supervision. The thesis contains no material which has been accepted, or is being examined, for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made. I give consent to the final version of my thesis being made available worldwide when deposited in the University’s Digital Repository, subject to the provisions of the Copyright Act 1968 and any approved embargo.

Thanh Nam Lam

Signature: ………………

Date: ….. October, 2018
ACKNOWLEDGEMENTS

After an intensive period of more than four years, today is the day: writing this note of thanks is the finishing touch on my thesis. It has been a period of rewarding learning for me, not only in the scientific arena, but also on a personal level. I would like to reflect on the people who have supported and helped me so much throughout my academic journey.

First and foremost, I would like to express my sincere gratitude to my wonderful supervisors Professor James Albright, Professor John Fischetti, and Professor Greg Preston for their continuous support of my PhD study, for their patience, responsibility, enthusiasm, and immense knowledge. I am fortunate to have had them as my supervisors. Without their dedicated guidance, this research project would never have been possible.

I acknowledge my debt to the Government of Vietnam through the Ministry of Education and Training and the University of Newcastle, Australia, for awarding me a VIED-TUIT scholarship to support my doctoral studies.

I am very thankful to the University of Newcastle for providing such a rich research resource and excellent student services and, in particular, to the friendly and supportive academic staff of the School of Education. I must say a special thank you to Vietnam Aviation Academy, where I am working, for facilitating my study with the most favourable conditions they can.

My profound gratitude goes to Professor Max Smith, Professor Allyson Holbrook, and Associate Professor James Ladwig, who gave me interesting lectures and useful research skills in my coursework at the University of Newcastle.

I wish also to express my great appreciation to Associate Professor Kylie Shaw, Associate Professor Mitchell O’Toole, Dr. Maura Sellars, Dr. Rachel Burke, Ms. Helen Thursby, Ms. Helen Hopcroft, and Mr. Nicholas Collier for their suggestions, interest, language assistance and constructive feedback on my very first manuscript. My additional thanks goes to Ms. Ruth McHugh for helping me with the tedious and long running job of reading proof.
I would like to express my heartfelt thanks to Dr Ho Thanh My Phuong. As my former supervisor, she has taught me and motivated me more than I could ever give her credit for here.

I would also like to take this opportunity to thank my former instructor, Dr Vu Thi Phuong Anh. She inspired me with research ideas about language testing and assessment. She has shown me, by her example, what a good scientist (and person) should be.

My deepest respect and appreciation go to Heads of the EFL Departments at the tertiary institutions in Vietnam for their generosity giving consent to my data collection in their EFL classes. I would like to thank the EFL teacher and student participants in Vietnam. Their information and ideas constituted the core of my study. Many of the participants set aside considerable amounts of their time to provide me with a profound understanding about their experiences and perceptions of oral assessment. The EFL experts’ insights enriched my research results and contributed enormously to the eventual conclusions from the investigation. I am greatly appreciative of their enthusiastic cooperation in my study.

I wish to thank my family for their unconditional love and encouragement during my time of studying away from home. They have been and still are ever ready to assist me in my various endeavours.

Finally, there are my lovely schoolmates here in Callaghan campus. We were not only able to support each other by deliberating over our study problems and findings, but also happily by talking about things other than just our papers. I cannot forget my friends from Vietnam and the US. Their messages and emails gave me motivational strength to complete my study.

Newcastle, 30 October 2018

Nam Lam
# TABLE OF CONTENTS

STATEMENT OF ORIGINALITY .................................................................................. ii  
ACKNOWLEDGEMENTS ............................................................................................ iii  
TABLE OF CONTENTS ............................................................................................ v  
LIST OF ABBREVIATIONS ....................................................................................... x  
LIST OF TABLES ..................................................................................................... xi  
LIST OF FIGURES ................................................................................................... xiv  
LIST OF APPENDICES ............................................................................................ xv  
GLOSSARY ................................................................................................................ xvii  
ABSTRACT ................................................................................................................ xviii  

Chapter One: INTRODUCTION .............................................................................. 21  
1.1 Background of the study ................................................................................ 21  
   1.1.1 A brief history of testing L2 speaking .................................................. 22  
   1.1.2 Trends in language testing ................................................................. 26  
   1.1.3 Context of the research ..................................................................... 29  
1.2 Research questions ....................................................................................... 44  
1.3 Significance of the research ........................................................................ 47  
1.4 Organisation of the thesis .......................................................................... 49  

Chapter Two: LITERATURE REVIEW ................................................................. 51  
2.1 Introduction .................................................................................................... 51  
2.2 Key issues in designing spoken language tests .......................................... 51  
   2.1.1 Construct validity in oral language testing ....................................... 56  
   2.1.2 Content aspect of construct validity .............................................. 59  
   2.1.3 Reliability ...................................................................................... 60  
2.3 Conceptual framework for validating speaking tests ............................... 63  
2.4 Formats of speaking tests .......................................................................... 65  
2.5 Technological applications in oral assessment .......................................... 68
2.6 Factors affecting test validity and reliability ................................................................. 69
   2.6.1 Assessment criteria ..................................................................................................... 69
   2.6.2 Rating scales ............................................................................................................. 70
   2.6.3 Test tasks ................................................................................................................ 71
2.7 Washback of oral language assessment ........................................................................ 74
2.8 Summary ........................................................................................................................ 76

Chapter Three: METHODOLOGY ..................................................................................... 78
3.1 Introduction .................................................................................................................... 78
3.2 Rationale for the research design .................................................................................. 78
   3.2.1 Adoption of a mixed methods approach .................................................................. 78
   3.2.2 Using a convergent design ..................................................................................... 80
3.3 Research setting ............................................................................................................. 81
   3.3.1 Research sites ......................................................................................................... 82
   3.3.2 Research participants ............................................................................................ 83
3.4 Data sources .................................................................................................................. 87
   3.4.1 Test room observation ............................................................................................ 89
   3.4.2 Questionnaire surveys ........................................................................................... 90
   3.4.3 Interviews ............................................................................................................... 95
   3.4.4 Expert judgements ................................................................................................. 98
   3.4.5 Documents .............................................................................................................. 99
3.5 Data collection procedures ........................................................................................... 100
3.6 Methods of data analysis ............................................................................................. 105
   3.6.1 Quantitative data processing ................................................................................. 105
   3.6.2 Qualitative data processing ................................................................................... 106
3.7 Presenting data analysis ............................................................................................... 112
3.8 Research ethics and compliance ................................................................................... 113
3.9 Assuring research quality ............................................................................................. 115
3.10 Summary ...................................................................................................................... 117

Chapter Four: RESULTS: TEST TAKER CHARACTERISTICS AND TEST ADMINISTRATION ......................................................................................................................... 119
4.1 Introduction .................................................................................................................... 119
Chapter Nine: SUMMARY AND CONCLUSIONS .......................................................... 317

9.1 Introduction ........................................................................................................ 317

9.2 Summary of research results ............................................................................. 318

9.2.1 Issues in test administration affecting test fairness and candidates’ speaking
performance ........................................................................................................... 319

9.2.2 Test content relevance and inequality in test questions’ degree of difficulty ... 323

9.2.3 Diversity in test tasks elicited different speech patterns for assessment........ 324

9.2.4 Inconsistency in rating and scoring spoken language performance .......... 326

9.2.5 Impact of oral testing on EFL teaching and learning .................................... 327

9.3 Limitations of the study and recommendations for future research .................. 335

9.4 Implications ........................................................................................................ 340

9.4.1 Implications for speaking test administrators ............................................ 340

9.4.2 Implications for speaking test designers .................................................. 342

9.4.3. Implications for oral test raters and scorers ............................................ 343

9.4.4 Implications for oral test takers ................................................................. 344

9.4.5 Implications for educational policy makers .............................................. 345

9.4.6 Implications for language testing researchers ......................................... 347

9.5 Conclusions ....................................................................................................... 348

REFERENCES .......................................................................................................... 352

APPENDICES .......................................................................................................... 385
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>Bachelor of Arts</td>
</tr>
<tr>
<td>BEC</td>
<td>Business English Certificate</td>
</tr>
<tr>
<td>CA</td>
<td>Conversation analysis</td>
</tr>
<tr>
<td>CALT</td>
<td>Computer-assisted language testing</td>
</tr>
<tr>
<td>CBT</td>
<td>Computer-based testing</td>
</tr>
<tr>
<td>CEFR</td>
<td>Common European framework of reference for languages</td>
</tr>
<tr>
<td>CLT</td>
<td>Communicative language teaching</td>
</tr>
<tr>
<td>EFL</td>
<td>English as a foreign language</td>
</tr>
<tr>
<td>ELT</td>
<td>English language teaching</td>
</tr>
<tr>
<td>ESL</td>
<td>English as a second language</td>
</tr>
<tr>
<td>ESP</td>
<td>English for specific purposes</td>
</tr>
<tr>
<td>ETS</td>
<td>Educational Testing Service</td>
</tr>
<tr>
<td>IELTS</td>
<td>International English Language Testing System</td>
</tr>
<tr>
<td>L1</td>
<td>First language</td>
</tr>
<tr>
<td>L2</td>
<td>Second/Foreign Language</td>
</tr>
<tr>
<td>MOET</td>
<td>Ministry of Education and Training (Vietnam)</td>
</tr>
<tr>
<td>OPI</td>
<td>Oral Proficiency Interview</td>
</tr>
<tr>
<td>SBA</td>
<td>School-based assessment</td>
</tr>
<tr>
<td>SE</td>
<td>Spoken English</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package for Social Sciences</td>
</tr>
<tr>
<td>TESOL</td>
<td>Teaching English to Speakers of Other Languages</td>
</tr>
<tr>
<td>TOEFL</td>
<td>Test of English as a Foreign Language</td>
</tr>
<tr>
<td>TOEIC</td>
<td>Test of English for International Communication</td>
</tr>
<tr>
<td>UCLES</td>
<td>University of Cambridge Local Examinations Syndicate</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

*Table 2.1* Purposes of language assessment ................................................................. 53  
*Table 2.2* Contrasting categories of L2 assessment ......................................................... 54  
*Table 2.3* Kinds of language tests .................................................................................. 55  
*Table 2.4* Categories of speaking assessment tasks ......................................................... 73  
*Table 3.1* Sampling of participants and research instruments used in the study .............. 87  
*Table 3.2* Summary of key points covered in the test room observation scheme .......... 90  
*Table 3.3* Summary of contents covered in the questionnaire for EFL teachers .......... 92  
*Table 3.4* Summary of contents covered in the questionnaire for EFL students .......... 95  
*Table 3.5* Summary of the EFL teacher interview protocol ............................................ 96  
*Table 3.6* Summary of the EFL student interview protocol ............................................ 97  
*Table 3.7* Information about EFL experts ...................................................................... 99  
*Table 3.8* Data collection methods in a three-stage process .......................................... 101  
*Table 3.9* Data collection before the test ...................................................................... 102  
*Table 3.10* Data collection during the test ..................................................................... 103  
*Table 3.11* Data collection after the test ....................................................................... 104  
*Table 4.1* Test takers’ gender profile ........................................................................... 123  
*Table 4.2* Speaking test anxiety across institutions ......................................................... 127  
*Table 4.3* Candidates’ oral test anxiety in relation to general self-evaluation of English 129  
*Table 4.4* Test takers’ English education prior to the oral test ..................................... 130  
*Table 4.5* Test takers’ profile of English-speaking class attendance ........................... 133  
*Table 4.6* Candidates’ oral test anxiety in relation to class attendance ....................... 134  
*Table 4.7* Comparing test administration methods across institutions .......................... 136  
*Table 4.8* Advantages and disadvantages of direct assessment ..................................... 139  
*Table 4.9* Raters’ opinions for and against audio-recording oral tests ........................... 141  
*Table 4.10* Examiners’ performance during an oral test session ................................... 143  
*Table 4.11* Means and standard deviations (SD) for candidates’ perceptions of the test administration ........................................................................................................ 146
Table 4.12 Candidates’ comments for and against of computer-assisted oral testing .... 150
Table 4.13 Means and standard deviations (SD) for raters’ perceptions of the test administration................................................................. 152
Table 4.14 Teacher raters’ comments for and against computer-assisted oral testing.... 155
Table 5.1 Comparison of the English-speaking courses across the institutions .......... 159
Table 5.2 Summary of test questions used for oral assessment across the institutions .. 163
Table 5.3 Information collected from EFL experts’ content judgement protocol .......... 166
Table 5.4 Content validity index of test items (I-CVI) across institutions .................... 167
Table 5.5 Examples of test items with I-CVIs of 1.0 and .83 .................................. 168
Table 5.6 The average of the I-CVIs for all items on the scale ................................. 169
Table 5.7 Examples of items rated ‘Relevant/Highly relevant’ and positive comments by all the experts ................................................................. 171
Table 5.8 Examples of problematic items and experts’ suggestions for revision ......... 172
Table 5.9 Examples of confusing test questions with experts’ comments and suggestions ......................................................................................... 173
Table 5.10 Question lengths in numbers of total words (TW) and content words (CW) across different task types adopted at the institutions .......................................... 176
Table 5.11 Examples of picture-cued questions with experts’ comments ............... 179
Table 6.1 Response formats used in oral assessment across institutions .................. 194
Table 6.2 Description of underlying purposes of oral test tasks ............................... 207
Table 6.3 Summary of characteristics of the test tasks across institutions ............ 208
Table 6.4 Rubric words and phrases indicating task purposes ............................... 209
Table 6.5 Elicitation question types used for interview tasks .................................. 220
Table 6.6 Time constraints in minutes for speaking tasks across institutions .......... 226
Table 6.7 Channels of communication in EFL oral assessment across institutions ...... 231
Table 6.8 Means and standard deviations (SD) for EFL teachers’ perceptions of test tasks ...................................................................................... 236
Table 6.9 Means and standard deviations (SD) for EFL students’ perceptions of test tasks ...................................................................................... 238
Table 7.1 Age profile of EFL teacher participants in the survey .............................. 244
Table 7.2 Assessment weighting of the Listening and Speaking components .......... 246
Table 7.3 Assessment criteria across institutions ................................................................. 247
Table 7.4 Features to be assessed and scored in a rating scale ............................................ 249
Table 7.5 Example of a holistic rating scale for the interlocutor ........................................ 250
Table 7.6 Purposes of the oral test for EFL majors ............................................................. 253
Table 7.7 Raters’ tendency of giving bonus points ............................................................... 262
Table 7.8 Oral rating affected by candidates’ performance in class ....................................... 265
Table 7.9 Means and standard deviations (SD) for raters’ perceptions of the oral rating
    and scoring ....................................................................................................................... 268
Table 7.10 Means and standard deviations (SD) for candidates’ perceptions of the oral
    rating and scoring ............................................................................................................. 270
Table 7.11 Candidates’ opinions for and against the number of raters in each test room ....... 272
Table 7.12 Descriptive statistics of oral test scores across the institutions ......................... 273
Table 7.13 Comparing categories of test scores across institutions ..................................... 274
Table 7.14 Correlation of test scores by University A’s raters ............................................. 279
Table 7.15 Descriptive statistics of University A’s raters’ scores in the second time of
    scoring ............................................................................................................................... 280
Table 7.16 Correlation of test scores by University B’s raters ............................................. 281
Table 7.17 Descriptive statistics of University B’s raters’ scores in the second time of
    scoring ............................................................................................................................... 281
Table 7.18 Correlation of test scores by University C’s raters ............................................. 282
Table 7.19 Descriptive statistics of University C’s raters’ scores in the second time of
    scoring ............................................................................................................................... 283
Table 8.1 Means and standard deviations (SD) for EFL students’ perceptions of the test’s
    impact on learning .......................................................................................................... 292
Table 8.2 Candidates’ desired strategies to improve English speaking skills ..................... 295
Table 8.3 Means and standard deviations (SD) for EFL teachers’ perceptions of the test’s
    impact on teaching and learning speaking skills ......................................................... 304
Table 8.4 Teachers’ changes and adjustments as a consequence of the oral test ............... 311
LIST OF FIGURES

Figure 2.1 Model of spoken language ability in the Cambridge Assessment ............ 57
Figure 2.2 Construct validity of test score interpretation ...................................... 58
Figure 2.3 Socio-cognitive framework for validating speaking tests (Weir, 2005, p. 46) 64
Figure 2.4 Interrelationship between assessment, teaching and learning ................ 74
Figure 3.1 Example of a paragraph in the First Coding Cycle ............................... 110
Figure 3.2 Framework for validating speaking tests (adapted from Weir, 2005) .... 112
Figure 4.1 Age groups of test takers taking the oral test .................................... 122
Figure 4.2 Candidate self-evaluation of English proficiency in terms of general
performance, accuracy, and fluency ................................................................. 124
Figure 4.3. Candidates’ experience with oral testing ............................................ 132
Figure 4.4 Seat arrangements for oral assessment at different institutions .......... 138
Figure 5.1 Content validity index (CVI) ............................................................... 165
Figure 7.1 EFL teachers using spoken English in Speaking classes .................... 244
Figure 7.2 Challenges for raters in oral rating .................................................... 254
Figure 7.3 Aspects in candidates' oral performance the rater paid attention to .... 259
Figure 7.4 Inter-rater agreement in oral test scoring across the institutions ......... 277
Figure 8.1 Impact of test scores on candidates .................................................... 287
Figure 8.2 Useful learning activities for candidates before the test ...................... 291
Figure 8.3 Major desired changes in teaching speaking skills ............................. 299
Figure 8.4 Factors affecting teaching oral skills ................................................. 301
Figure 8.5 Useful activities to prepare students for the oral test ......................... 303
Figure 9.1. Example of a wrap-up activity for a speaking skills lesson ............... 331
LIST OF APPENDICES

Appendix A: Ethics approval documents ............................................................... 385
A.1 HREC Approval (17/11/2015)
A.2 Information Statement for (a) Head of EFL Faculty
(b) EFL teachers
(c) EFL students (*)
A.3 Consent Form for (a) Head of EFL Faculty
(b) EFL teachers
(c) EFL students (*)
A.4 HREC Expedited Approval of a protocol variation (22/02/2017)
A.5 Information Statement for (a) EFL teacher raters
(b) EFL experts
A.6 Consent Form for (a) EFL teacher raters
(b) EFL experts
A.7 Verification of translated documents

Appendix B: Data collection instruments .............................................................. 422
B.1 Observation Protocol
B.2 Questionnaire for (a) EFL teachers
(b) EFL students (*)
B.3 Interview Protocol for (a) EFL teachers (*)
(b) EFL students (*)
B.4 Documents for EFL experts
   (a) University A’s Test Content Judgement Protocol
   (b) University B’s Test Content Judgement Protocol
   (c) University C’s Test Content Judgement Protocol
   (d) Demographic information about EFL experts
B.5 Focus group discussion protocol for EFL teacher raters (*)

Appendix C: Additional resources for data collection ........................................... 451
C.1 Procedure and speaking notes for initial contact with potential participants
C.2 Instructions for raters’ speech sample recording in test rooms
C.3 Checklist of documents to be collected
C.4 Interview transcription template

Appendix D: Examination documents ................................................................. 455
D.1 Assessment criteria, rating scales and scoring sheets at three institutions
D.2 Example of test material for oral examiners (interlocutor outline)
D.3 Speaking skills and discussion strategies introduced in course books
D.4 Scoring sheets for semi-direct assessment

Appendix E: Data management ........................................................................... 463
E.1 Coding scheme of participants and data sources
E.2 Managing recorded speech samples
E.3 Multiple data sources used to examine key aspects in speaking test validation
E.4 Tabulated qualitative data

Appendix F: Examples of preliminary data processing ......................................... 471
F.1 Coding a rater interview transcript in two cycles
F.2 Rough transcript of a speech sample
F.3 Additional excerpts of task-based oral performances
F.4 Occurrence of language functions in speech samples

Appendix G: Sample rating scales for assessing speaking skills ......................... 479
G.1 Sample rating scales from the literature: global and analytic
G.2 Common Reference Levels proposed by the CEFR
G.3 The CEFR-based English Competence Framework adopted in Vietnam (*)

Appendix H: Flowchart of procedures for data collection and data analysis .......... 486

Appendix I: Transcription notation symbols ...................................................... 487

Appendix J: List of tertiary institutions in HCMC (Vietnam) .............................. 488

Appendix K: Original quotes in Vietnamese ....................................................... 489
K.1 Quotes from the Vietnamese press and literature
K.2 Quotes from Vietnamese interviews with research participants

(*) a Vietnamese translation included
GLOSSARY

The following terms will be used as key words in this research study:

- **achievement test**: a test which aims to measure what has been learnt within a course of instruction or up to a given time
- **construct validity**: the degree to which a test can measure what it claims to be measuring
- **content validity**: the extent to which the test approximately provides samples from the domain of knowledge and skills relevant to performance according to the preset criterion
- **inter-rater reliability**: the extent to which two or more examiners would agree with each other on the scores awarded to a single learner
- **rating scale** (also **scoring rubric**): an ordered set of descriptions of typical performances in terms of their quality, used by raters in the rating process
- **reliability**: consistency of measurement across individuals by a test
- **reliability coefficient**: a statistic on a scale between -1 and +1, expressing the extent to which individuals have been measured consistently by a test
- **testing**: the collection of any data that can be used to assess the language abilities of respondents
- **assessment**: the inclusion of more variables when attempting to determine language proficiency and achievement
- **washback**: the effect of a test on the teaching and learning leading up to it
ABSTRACT

English language mastery plays a crucial role in global communication, trading, and cultural exchange. Vietnam has set English as a strategic goal of the national education system for boosting the process of regional and international integration. Vietnamese education has been making a great effort to enhance the effectiveness of English as a foreign language (EFL) instruction with a focus on communicative competence, particularly listening and speaking skills. My study aimed to examine the operational practices of oral assessment as an inseparable component in relation to L2 teaching and learning in Vietnam.

Participants in my study project were EFL majors (candidates) and teachers (examiners) involved in testing English speaking skills at three universities in South Vietnam. My data collection for this empirical research started in late 2015 and continued to early 2016. Data sources included oral test observation, questionnaire surveys, interviews, testing documents, test scores, and EFL content experts’ judgements.

The results highlight a methodological diversity in oral assessment across the institutions in terms of test administration, task design, language content, the process of rating, and the impact of testing speaking. Test scores served more to complete a required unit of study rather than to make reliable inferences about learners’ oral production ability and provide useful feedback for future improvement. Interview tasks did not reflect characteristics of natural conversations when the interlocutor (also examiner) played a predominant role in the assessment context. Several oral test questions required candidates’ theoretical knowledge in order to be answered. These factors had the potential to increase students’ test anxiety and hindered their best performance. Discussion tasks enabled a wide variety of speech functions to be produced and provided opportunities to manage verbal interaction between paired candidates. Interactive speaking revealed students’ weaknesses in a co-constructed oral performance, and a tendency for individual turn-taking to talk about the assigned topic.
My study suggests implications for various stakeholders who could assist to improve the quality of oral testing and assessment in educational contexts. Rater training and double rating are necessary for oral assessment to eliminate unavoidable measurement errors by human raters. My research results indicate a need for more clearly defined assessment criteria and descriptors in the rating scales to obtain higher consistency in assessing spoken English abilities. The recent promulgation of the CEFR-based evaluation framework for L2 proficiency in Vietnam has brought both opportunities and challenges for those who are concerned about enhancing and standardising the national quality of EFL education.