The University Of Sheffield.

Design and Development of a Pronunciation Self-Assessment Checklist for Aircraft Engineering Technology Students in an ESL Higher Education Context

By:

Nurul Ain Binti Md Zulkifly

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Arts and Humanities
School of English

8 September 2023

Author's Declaration

I certify that work in this thesis was carried out in accordance with the Regulations of the University of Sheffield. The work is original, except where indicated by special reference in the text, and no part of this thesis has been submitted for any other academic award. Views expressed in the thesis are those of the author.

Signed: Date: 8 September 2023

Dedication

This work is dedicated to my husband, our parents, our son and our family members.

Acknowledgements

I am grateful to everyone who has made the completion of this thesis submission possible. My great appreciation goes to Associate Professor Dr. Vahid Nimehchisalem, Professor Nigel Harwood, and Professor Ain Nadzimah Abdullah who believed in me. Thank you for your patience, guidance, encouragement and continuous support throughout my doctoral studies. Additionally, I would like to also thank all the staff from both universities, Universiti Putra Malaysia and The University of Sheffield, England, for their dedicated assistance during my studies.

I would like to thank the management of Universiti Kuala Lumpur as a whole and specifically Universiti Kuala Lumpur Malaysian Institute of Aviation Technology (UniKL MIAT) for the opportunity, their understanding and never-ending support.

I am truly grateful to the insightful comments provided by the all the experts which have given me different perspectives on this research. My special thanks also go to my colleagues and students without whose participation and contributions it would not have been possible to undertake this research project.

My deepest gratitude goes to my supportive family for their understanding and patience throughout my PhD journey. To my loving husband who has been there every step of the way, to our son who is my strength and motivator, our parents who kept me in their prayers, our sisters and the girls who have been amazingly kind with their unconditional support and the rest of the family who had put up with me.

As I am surrounded with positive energy, I would like to thank all my good friends for their unconditional love and support. Thank you again everyone who made this journey a memorable one.

ABSTRACT

There is increasing interest in the use of formative self-assessment in the language learning classroom. This thesis presents the design and development of a self-assessment checklist for pronunciation intended as a useful pedagogical tool for use by Bachelor of Aircraft Engineering Technology (BAET) students in an ESL Higher Education context in a private university in Malaysia. The emphasis on pronunciation arises from the importance attached to communications in minimising risk and maximising safety in the aviation industry, where most BAET students will work after graduating, and from the associated priority students and teachers give to pronunciation. Self-assessment is needed for students to become independent learners, a skill which they will need for developing themselves in their future careers. Self-assessment in this study is used in the sense of students developing their understanding of how to self-assess their pronunciation as opposed to self-grading their pronunciation.

A development research approach was adopted, consisting of three phases: design, calibration, and evaluation of the usefulness of the checklist. The approach was chosen to meet the research objectives which were: (1) to design a pronunciation self-assessment checklist, (2) to evaluate the pronunciation self-assessment checklist based on the feedback from teachers and students, and (3) to evaluate the usefulness of the pronunciation self-assessment checklist.

The checklist design was based on the literature review, with changes made after comments from teachers and students in addition to expert validation. In the calibration phase, the checklist was trialled with students (N=50), after which semi-structured interviews were conducted with ten students and two teachers to establish whether the checklist and the guidance it contained were sufficiently clear, and to investigate the feasibility of using the self-assessment checklist in the classroom situation, as measured by the reactions of students and teachers. In the third and final phase, the usefulness of the checklist was evaluated by administering questionnaires developed for this study and validated by experts to students and teachers who had participated in trialling the checklist. Usefulness was measured in terms of the impact, practicality, reliability and validity of the self-assessment checklist. Further expert validation was incorporated into the final design of the checklist.

Overall, the teachers found the checklist to be very useful, as did students (n=20) who returned valid questionnaires. There was evidence that some students had understood and

were able to implement the self-assessment cycle; additionally, there was evidence of both knowledge and regulation of cognition. Some students wanted more explanation of the checklist, or wanted it to be simpler, while others suggested possible ways of embedding pronunciation teaching as well as the checklist into the Aviation English curriculum. Further work is needed to ensure effective implementation of the checklist in the classroom. The thesis offers a novel approach to self-assessment of pronunciation in an ESL context, focusing on students' understanding of their own learning strategies, and highlights the benefits of a rigorous approach by teachers to the use of checklists.

List of Abbreviations

Abbreviation	Meaning
ADDIE	Analysis, Design, Development, Implementation, and Evaluation
ALM	Audiolingual Method
AMM	Aircraft Maintenance Manual
AMT	Aircraft Maintenance Technician
ATC	Air Traffic Controller
BAET	Bachelor of Aircraft Engineering Technology
CAAM	Civil Aviation Authority of Malaysia
CEFR	Common European Framework of Reference for Languages
CF	Corrective Feedback
CLT	Communicative Language Teaching
CPD	Continuing Professional Development
CR ADMIN	Checklist Review Administration
CR FREQ	Frequency Scale
CR GUID	Checklist Review Guidance
CR ITEM	Checklist Review Items
CR LAYOUT	Checklist Review Layout
CR PRO	Checklist Review Pronunciation
CR SEC	Checklist Review Sections
DCAM	Department of Civil Aviation Malaysia
EAP	English for Academic Purposes
EFL	English as a Foreign Language
ESL	English as a Second Language
ESP	English for Specific Purposes
IATA	International Air Transport Association
ICAO	International Civil Aviation Organization
IELTS	International English Language Testing System
IPA	International Phonetic Alphabet
KC	Knowledge of Cognition
L1	First Language
L2	Second Language
L3	Third Language

Abbreviation	Meaning
MAEP	Manual of American English Pronunciation
ME	Metacognitive experiences
OET	Occupational English Test
RC	Regulation of Cognition
RP	Received Pronunciation
SA	Self-assessment
SAPUR	Purpose of Self-assessment
SAST	Standards
SAT	Self-Assessment Tool
SILL	Strategy Inventory for Language Learning
SR	Student Role
SUSA	Students' Understanding of Purpose and Process of Self-assessment
TOEFL	Test of English as a Foreign Language
TR	Teacher Role
TU	Teachers' Use of SA
UniKL	Universiti Kuala Lumpur

Table of Contents

Author's Declaration i

Dedication ii

Acknowledgements iii

ABSTRACT iv

List of Abbreviations vi

1 CHAPTER 1: INTRODUCTION 1

1.1 Preamble 1

1.2 Development research 1

1.3 Contextualisation and background of the study 2

1.4 Researcher's positionality 9

1.5 Problem statement 11

1.6 Rationale 16

1.7 Research objectives and questions 18

1.8 Theoretical framework 19

1.8.1 Metacognition 23

1.8.2 Metacognition and reflection 27

1.9 Contribution to knowledge 30

1.10 Key terms 31

1.11 Thesis structure 33

2 CHAPTER 2 LITERATURE REVIEW 34

2.1 Introduction 34

2.2 Pronunciation 34

2.2.1 Phonology of pronunciation 34

2.2.2 Teaching pronunciation: revolutions or evolution? 36

2.2.3 Nativeness, accentedness, intelligibility and comprehensibility 41

2.3 Pronunciation assessment criteria in high-stakes English tests 42

2.4 Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) 52

2.5 The changing landscape of self-assessment 54

2.5.1 Understandings of self-assessment 55

2.5.2 Self-assessment of pronunciation in English as a Second Language (ESL) speakers 58

2.6 Metacognition and language learning strategies 60

2.7 Checklists 68

2.8 Review of selected studies 70

2.9 Summary 75

3 CHAPTER 3 METHODOLOGY 78
3.1 Introduction 78
3.2 Research design 78
3.2.1 Design phase 83
3.2.2 Calibration phase 84
3.2.3 Evaluation phase 84
3.2.4 Philosophical assumptions 85
3.3 Research process 85
3.3.1 Phase 1: Development of pronunciation self-assessment checklist 85
3.3.2 Phase 2: Calibration of the checklist 88
3.3.3 Phase 3: Evaluation of the checklist 91
3.4 Data collection instruments 92
3.4.1 Self-assessment checklist 92
3.4.2 Semi-structured interviews: design phase 95
3.4.3 Group interview 97
3.4.4 Expert validation: design phase 98
3.4.5 Semi-structured interviews: calibration phase 100
3.4.6 Evaluation of usefulness questionnaires: evaluation phase 101
3.4.7 Expert validation: evaluation Phase 109
3.5 Participants 109
3.5.1 Aviation English teachers 110
3.5.2 Student selection 111
3.5.3 Experts 112
3.6 Data analysis 112
3.6.1 Codebook development 113
3.6.2 Intercoder and code reliability 117
3.7 Ethical considerations 122
3.8 Summary 123
4 CHAPTER 4 RESULTS 124
4.1 Introduction 124
4.2 Design Phase Changes to Checklist v1.0 124
4.3 Design phase findings 132
4.4 Changes to Checklist v2.0 136
4.5 Calibration phase findings 141
4.6 Inspection of a sample of checklists 148
4.7 Evaluation phase findings 151
4.8 Summary 152

5 CHAPTER 5 DISCUSSION, CONCLUSION AND RECOMMENDATIONS 154
5.1 Introduction 154
5.2 Summary of the study 154
5.3 Summary of the findings 156
5.4 Discussion of findings 158
5.4.1 Design stage issues 159
5.4.2 Calibration stage issues 161
5.4.3 Evaluation stage issues 162
5.4.4 Self-assessment in the classroom 163
5.4.5 The practicality of designing and developing checklists 165
5.4.6 Implementation issues 165
5.5 Self-assessment of pronunciation and English for Specific Purposes 166
5.6 Comparison of checklist designs 167
5.7 Development research 171
5.8 Contribution to knowledge 174
5.9 Limitations of the study 175
5.10 Reflections on the research approach 176
5.11 Reflections as a teacher-researcher 178
5.12 Directions for future research 180
5.13 Conclusion 181
REFERENCES 183
APPENDIX A Pronunciation Self-Assessment Checklist v1.0 209
APPENDIX B Interview Schedule (Teacher): Design phase 215
APPENDIX C Structured Group Interview Questions (Student): Design Phase 217
APPENDIX D Pronunciation Self-Assessment Checklist v2.0 219
APPENDIX E Interview Guide (Students): Calibration Phase 224
APPENDIX F Interview Guide (Teachers): Calibration Phase 225
APPENDIX G Checklist Administration Instructions 229
APPENDIX H Evaluation Questionnaires 231
APPENDIX I Expert Reviewer Rating Form 234
APPENDIX J Three Selected Transcripts 237
APPENDIX K Instructions to Second Coder 250
APPENDIX L Codebook 251
APPENDIX M Teachers Evaluation Questionnaire Tables 258
APPENDIX N Participant Information Sheet and Consent Form 260
APPENDIX O Ethics Approval 266
APPENDIX P Pronunciation Self-Assessment Checklist v3.0 267

APPENDIX Q Sample Transcript 272
APPENDIX R ICAO Language Proficiency Rating Scale 277
APPENDIX S Typical English Pronunciation Problems Faced by Students 279
APPENDIX T Visual Representations of Development 285

Table of tables

Table 2-1 TOEFL independent speaking rubric 43
Table 2-2 IELTS pronunciation rubric 44
Table 2-3 OET 2018 47
Table 2-4 ICAO rating scale extract 48
Table 2-5 Pronunciation subscale 51
Table 2-6 Andrade's taxonomy of self-assessment 57
Table 3-1 Research design overview 82
Table 3-2 Steps in calibration phase 89
Table 3-3 Example area to explore with subsidiary questions 97
Table 3-4 Extract from group interview questions 97
Table 3-5 Extract from semi-structured interview (teacher) calibration phase 101
Table 3-6 Components of test usefulness covered in the student questionnaire 106
Table 3-7 Aspects of impact covered in the student questionnaire 107
Table 3-8 Summary of items by domain showing reversed items (student questionnaire) 107
Table 3-9 Numbers of participants by research phase 110
Table 3-10 Teacher qualifications and experience 111
Table 3-11 Students by gender and level of English 111
Table 3-12 Data analysis methods 113
Table 3-13 Example of researcher's assumption in initial coding 115
Table 3-14 Example of amended initial coding 115
Table 4-1 Summary of changes proposed by the teachers and students: Design phase 131
Table 4-2 Summary of experts' suggestions, researcher's responses, rationale and changes 139
Table 4-3 Response ranges in sampled checklists 149
Table 4-4 Spread of sample responses in checklist Section A 149
Table 4-6 Spread of sample responses in checklist Section B 150
Table 4-7 Spread of sample responses in checklist Section C 150

Table of figures

Figure 1-1 Relationship between metacognition and cognition. Adapted from Nelson & Narends, 1996 and Muijs & Bokhove, 2020. 24
Figure 1-2 Simplified metacognitive self-assessment cycle. Adapted from information from Boud and Molloy (2013); Flavell (1979), and Perkins (2008), and figures from Muijs and Bokhove (2020, and Zimmerman and Moylan (2011) 29
Figure 2-1 Format of guidance for raters on assessing frequency (Isaacs et al., 2015, p. 45) 46
Figure 2-2 CEFR phonological scale 53
Figure 3-1 Extract from expert reviewer rating form 100
Figure 4-1 Checklist extract: 'After the speaking activity' 130
Figure 5-1 Speaking self-assessment checklist Jankowska and Zielińska (2015) 168
Figure A-1 Visualisations of checklist development 300

1 CHAPTER 1: INTRODUCTION

1.1 Preamble

The purpose of this development research study was to develop a pedagogical tool for Bachelor of Aircraft Engineering Technology (BAET) students to enable them to develop their understanding of how to self-assess their pronunciation. The study was conducted in a private university in Malaysia which specializes in Aircraft Maintenance Technology and offers courses in Aviation English which are undertaken mainly by Malaysian students. All technical courses are delivered in an English as a Second Language (ESL) environment.

This chapter introduces the development research approach, followed by the context and background of the study that influenced the choice of approach. Then, the researcher's positionality is briefly explained. This is followed by the problem statement which identifies the practical problem as well as gaps in the literature related to assessment and self-assessment of pronunciation. After this, the research rationale, the research aim, objectives, and questions are presented. The theoretical framework of self-assessment within a metacognitive approach is then explained, and an overview of the research process is given. Key terms are defined before the chapter concludes with the organization of the thesis.

1.2 Development research

Development research has been associated with a variety of activities in the field of teaching and learning (Van den Akker, 1999, pp. 4-5). It "directly pertains to instructional development", including the development and validation of a specific instrument to address a problem in a particular context (Richey & Klein, 2005, p. 24). Thus, it can meet

teachers' and students' needs for an instructional product like the checklist in this present study while ensuring the product is based on data gathered from practice as well as theory.

Development studies focused on a context-specific instructional tool often have three or four phases (Richey & Klein, 2005, p. 26). Typical phases in development research can include some or all of the following: design and development, evaluation, validation, model development, model use and model evaluation (Richey & Klein, 2005). The overall process may be iterative, with repetitions of some or all of the phases (Fulcher & Davidson, 2007, p. 84). The process can thus be visualised in a variety of ways, as illustrated in Appendix T. In this present thesis, three phases are used: design, calibration, and evaluation.

Whilst the selected development research approach is described and justified in full in the methodology chapter, it is worth mentioning here that its distinctive features relevant to this present study include the involvement of practitioners and users in repeated prototyping (Van den Akker, 1999, pp. 8-9; Visscher-Voerman et al., 1999, p. 16), and an emphasis on formative evaluation of the product (Van den Akker, 1999, p. 10). The involvement of end users in this way is based on the assumption that they are the best judges of the "value of a product and its functionality" (Visscher-Voerman et al., 1999, p. 21).

1.3 Contextualisation and background of the study

It is asserted that pronunciation is key to effective communication at work; "Pronunciation plays a vital part in employability" (Shak et al., 2016, p. 25). Many studies have stated that Malaysian employers report a lack of English proficiency among the reasons for graduate unemployment (Cheong et al., 2016; Nadarajah, 2021; Nair et al., 2012; Sarudin et al., 2013; Singh, G. & Singh, S., 2008; Ting et al., 2017). Shak et

al. (2016) mentioned that for university students, clear pronunciation is not only important for oral coursework and assessments, but also for "job interviews, proposal presentations, negotiations and other professional instructions" as they move from university to employment (p. 25). Additionally, in an ESL higher education environment like that of the university in this present study, students from diverse linguistic backgrounds need to be able to understand each other as well as communicate successfully with their lecturers. There are 18 first languages spoken in Malaysia in addition to Bahasa Melayu, the language spoken by the majority of the population (Ling, 2020; Rashid et al., 2017). The student intake at the research site in a particular year may include international students as well as Malaysian students, who may converse in their L1 during break times, as mentioned by Kashinathan and Abdul Aziz (2022). Differences among Malaysian student speakers of English, depending on whether English is their L1, L2 or L3, are mentioned by Pillai and Ong (2018, p. 154). Moreover, the quality of English teaching in primary and secondary schools is variable across the country (Kepol. 2017; Rashid et al., 2017), which means that there can be wide differences in English speaking proficiency among students arriving at the university. Burns (2003) summed up the need for clear pronunciation in both employment and the classroom, saying "When communicating in the English language, it is important that the speaker can be understood by the listener and clear pronunciation helps to convey the message even if there are some problems with vocabulary and grammar" (Burns, 2003, p. 5).

It is worth explaining a little more about the particular teaching and learning context of English in Malaysian higher education because in practice it tends to be difficult to distinguish between ESP, English as a Second Language (ESL) and English as a Foreign Language (EFL) in the design and delivery of an English curriculum. Malaysia has historically considered itself an ESL country, and indeed continues to do so (Aziz & Kashinathan, 2021). However, changes over time in the perceived status of English

compared with Bahasa Malaysia as the language of national identity has led to a situation in which the teaching of English in higher education institutions has become closer to TEFL (Ravindran et al., 2022, p. 2). Additionally, a number of students are Chinese or from other countries, and for them, English is definitely a foreign language (Ravindran et al., 2022, p. 2). The limited use of English in the classroom and workplace is confirmed by Hashim and Yunus (2018) who state that in reality English is used in situations which demand it, for example presentations or reports, but that Bahasa Malaysia is used at other times (p. 209).

University students undertaking courses which require specialised knowledge of English for their studies and future careers undertake an ESP course, which varies in content according to their chosen area and which therefore relies heavily on analysis of their particular needs, as indicated by Flowerdew (2014). According to Johns (2014), who cites Dudley-Evans and St John (1998), this means that an ESP practitioner may fulfil a number of roles in addition to teacher, such as curriculum and materials designer (p. 19). In designing and teaching an ESP course, a teacher may have to address students' needs in terms of English for Academic Purposes (EAP), for instance primarily reading or writing, in order for students to be able to succeed in their specialised academic studies, while simultaneously addressing their needs for meeting occupational requirements (Flowerdew, 2014, p. 337). As Feak (2014) mentions, writing skills have tended to dominate ESP curricula due to the direct link between those skills and professional success (p. 34). However, speaking skills have added importance in occupations where interactions with speakers from different linguistic backgrounds have become more commonplace; in situations where a variety of L1 speakers are communicating, it is the speaker's responsibility to ensure their speech can be understood (Feak, 2014, p. 43).

For BAET students, English proficiency is especially significant. Many BAET students seek to work as aircraft maintenance technicians (AMTs), while others will pursue

various careers in the aviation industry ranging from systems engineer to human resources manager. Although they all study in an ESL environment at their technical university, those who wish to become aircraft maintenance technicians are required to achieve a specified level of proficiency on a rating scale determined by the International Civil Aviation Organization (ICAO). The national aviation authority of Malaysia, the Civil Aviation Authority of Malaysia (CAAM), has made language proficiency rating mandatory for aircraft maintenance personnel (DCAM, 2014).

In 2012, the Aviation English lecturers gained experience with the national aviation authority in using the ICAO Language Proficiency Rating Scale to assess the English of aircraft traffic controllers. As a result of this assessment experience, and the Malaysian national aviation authority requirement for assessment by the university of AMT students' language proficiency, the lecturers decided to introduce the ICAO standard of English language proficiency to Bachelor of Aircraft Engineering Technology (BAET) students. It was important for students to know the standard they would be expected to achieve to work in the industry, but most of the teachers felt that giving feedback to students based strictly on the standards could discourage students and therefore started to consider alternative assessment frameworks. After their initial experience of using the ICAO scale when teaching Aviation English classes, teachers and some of their students were asked whether there was anything they would like to be given more attention. The six Aviation English lecturers who had previously been involved in rating aircraft traffic controllers and introducing students to the standards using the ICAO scale stated that students could gain from a greater focus on pronunciation and a number of students similarly expressed a need to focus more on their pronunciation. The emphasis on pronunciation arises from the importance attached to communications in minimising risk and maximising safety in the aviation industry, which is where most of our students will work after graduating. The aviation industry originally focused on communication

between airline pilots and air traffic controllers, using a specific and restricted code of communication. This was extended to improve communication in non-routine situations as well as less formal exchanges between airline pilots and air traffic controllers, and thereafter to air maintenance technicians who need to work in an international environment in which English was typically the lingua franca. The ICAO Language Proficiency Rating Scale was, and still is, the only rating scale specific to the aviation industry. As the literature review indicates, in comparison to other rating scales, the ICAO highlights the importance of clear pronunciation, mainly for reasons of safety. It was therefore decided to proceed with developing a pedagogical tool related to pronunciation.

In 2014, CAAM took the initiative to adopt and amend the requirements set by the ICAO for English language proficiency. This fulfilled the specific maintenance-related requirements for Malaysian aircraft maintenance personnel and addressed the perceived decline in overall levels of English language proficiency among younger aircraft maintenance personnel compared with more senior staff (DCAM, 2014, para. 2.2). Situations where speaking skills, and hence pronunciation, were needed included "briefing, announcements, instructions, discussions, shift hand-over communication, communication with ATC, pilots and other ground personnel" (DCAM, 2014, para. 2.4). The ICAO Language Proficiency Rating Scale would be adopted as a general guide for competency levels, with Level 4 being the minimum required level for the granting or renewal of an Aircraft Maintenance License (see Appendix R for ICAO speaking scale levels and descriptors).

CAAM also specified that assessments of English language proficiency should be conducted by authorised organisations, including listed Approved Training Organisations for aviation maintenance, one of which is the university where this present research was conducted (CAAM, 2021). The university needed to look at how it could develop the

English component of the BAET degree, in line with the statement by White, Kroes and Watson (2000, p. 2) that "a successful AMT training program relies on a system of continuous quality improvement". The university decided that BAET students should be assessed using the framework which was already available for pilots and Air Traffic Controllers (ATCs). There were two main factors that led to this decision: (i) the lack of suitable and available alternative tests, and (ii) the emergence of high-stakes English testing for aircraft maintenance technicians, following the earlier introduction of mandatory standardised high-stakes testing for pilots and ATCs. The Aviation English lecturers at the university therefore started to assess BAET students' speaking skills with the ICAO Language Proficiency Rating Scale.

There were two areas of difficulty, the first of which concerned the course content and assignments. Since the Aviation English courses were intended for a range of occupations in the industry, the overall course content and speaking skills assignments were broader than those maximally relevant for aircraft maintenance technicians alone. As such, the course does not reflect the highly specialized aviation discourse as a form of English for Specific Purposes (ESP) described by Moder (2014), although the requirement for "comprehensible pronunciation, intelligible fluency and delivery" in typical employment situations remains (p. 228). Moreover, few studies had investigated the language proficiency of aircraft maintenance personnel (Drury et al., 2002; Drury and Ma, 2003; Drury and Ma, 2004; Drury et al., 2005) as compared to pilots and ATCs (Alderson, 2009; Alderson, 2011; Cookson, 2011; Farris et al., 2008; Kim & Elder, 2009, 2011; Matthews, 2004; Moder & Halleck, 2009; Tajima, 2004). The final report to the Federal Aviation Administration regarding the language proficiency of aircraft maintenance personnel (Drury et al., 2005) found that limited English on the part of an aircraft maintenance technician or inspector was responsible for between four and ten incidents a year (p. 85). Early detection and correction of language errors was recommended, along with "Regular

testing of AMT's [aircraft maintenance technician] English ability, task assignments recognizing the AMT's knowledge of English and an atmosphere of freedom to ask for language assistance" (Drury et al., 2005, p.95).

The second area of difficulty was the time available for the Aviation English component of the BAET degree; a two-hour session each week for three semesters, each lasting 18 weeks. The time was just sufficient within the aviation-related syllabus to incorporate classroom activities in the four language skills of reading, writing, listening and speaking, all deemed equally important for aircraft maintenance personnel (DCAM, 2014, para. 2.4). The time constraints of the course, along with the general shift towards a more constructivist and learner-centred approach in language teaching (Kaufman, 2004) and the potential for self-assessment in language learning (Jamrus & Razali, 2019, p. 70), led the Aviation English lecturers to feel that students would benefit from the introduction of self-assessment. Self-assessment is "among the main learner-centered practices which would potentially be beneficial in testing as well as learning processes" (Hosseini & Nimehchisalem, 2021, p. 857).

In view of the diversity of spoken language proficiency among students and the likely benefits of self-assessment, the purpose of the pedagogical tool was to enable students to develop their understanding of how to self-assess their pronunciation so that they could identify and attend to their individual needs for improvement. Some typical difficulties with students' English pronunciation are described in Appendix S. At the same time, it sought to address students' expressed dislike of using audio- or video-recordings, as reported to their teachers, to detect their mistakes and improve their pronunciation. The researcher realised that simply urging them to use these strategies was more likely to increase resistance than to bring about change and so decided to use self-assessment to encourage them to take ownership of their learning in terms of the strategies they used. If students considered which learning strategies they used and how well those strategies

were working in terms of helping them improve their pronunciation, they would have the opportunity to adapt their strategies, if not to listen to recordings, then perhaps to try alternatives.

Self-assessment checklists have been produced for language learning within the Common European Framework of Reference CEFR European Language Portfolio (Little & Perclová, n.d.), which has been adopted in Malaysia (Ministry of Education, 2013). This suggests there is support for the potential usefulness of checklists associated with self-assessment. In view of the priority students and teachers gave to pronunciation and the time constraints of the course more generally, the present study concentrates on developing a useful pedagogical tool in the form of a self-assessment checklist for pronunciation.

1.4 Researcher's positionality

As Wellington et al. (2005) stated, "It is impossible to take the researcher out of any type of research or of any stage of the research process" (p. 21) and therefore the researcher explicitly addresses her positionality as a researcher in connection with this project. She is one of a team of six lecturers who teach Aviation English to BAET students at a private technical university in Malaysia. Like most of her colleagues, she is an ESL speaker working in an ESL environment. This present study grew out of team efforts over several years to identify appropriate improvements in teaching and assessment of students' English. Self-assessment was seen as the next area to develop, while pronunciation was seen as a relatively weak area of the speaking skill, which in turn was seen as weaker than the other language skills. This study gave the researcher the opportunity to explore the self-assessment of pronunciation and to develop a pedagogical tool in the form of a self-assessment pronunciation checklist for students.

This led the researcher to consider her position as an insider, not only as a member of a teaching team, but as member of staff with responsibility to the employing university and as someone with a pre-existing teacher-student relationship with some of the participants. Although her understanding of the cultural and educational situation was mostly shared with all of them, she was aware there was a different power relationship involved when students addressed her formally as 'Madam' in class, and that this could especially affect how some student participants might respond.

Being an insider not only facilitated her access to participants, in agreement with Sikes and Potts (2008, p. 3), but also meant that she had real insight into the situation. She hoped that this would help her to contribute to the type of research into real-world classroom-based issues recommended by Rose (2019). However, being an insider can also attract criticism regarding "the extent to which [research] can be considered to be 'objective' and hence 'reliable' and 'valid'" from a strictly scientific viewpoint (Sikes & Potts, 2008, p. 7). Rose (2020) states that, during observations, a researcher can take a role on a spectrum that stretches from "detached observer", perhaps more likely with an outsider, to "complete participant", an insider role (pp. 95-96). According to Rose, "Positionality…needs to be firmly established before conducting observations" (2020, p. 95), with the researcher deciding the role appropriate to the nature of the research.

Positionality also affects interviews; it can be a source of bias and can be "very difficult…or impossible to avoid" (Rose, 2020, p. 119). The researcher in the present study needs to acknowledge where she is aware that her insider role as a teacher may have influenced some of the interview responses, for example where the teacher-student power difference caused some, but not all, students to be maybe less critical of the checklist than they could have been.

Additionally, it is very likely that the analysis and interpretation of data is influenced by the researcher's positionality (Bourke, 2014) and indeed "all aspects and stages of the research process" (Holmes, 2020, pp. 2-3). Positionality exerts an influence on the design and processes involved in any research as well as the ways in which ethical considerations are addressed (Clough & Nutbrown, 2012). However, the researcher's experience was that the nature and stage of the research also affected her positionality, so that she was closer to the detached observer end of the spectrum at some points. Especially when reading the literature, she made a conscious effort to read widely and take on board ideas that did not immediately match her knowledge and experience, incorporating some of them, such as developing metacognitive awareness, into the first version of the checklist.

1.5 Problem statement

Globally, "pronunciation research and pedagogy have long been influenced by two contradictory principles, the nativeness principle and the intelligibility principle" (Levis, 2005, p. 370). The nativeness principle, which stresses accuracy in the sense of sounding as much like a native speaker as possible, has over time given way to the intelligibility principle, which acknowledges that:

> anyone participating in international communication needs to be familiar with, and have in their linguistic repertoire for use, as and when appropriate, certain forms (phonological, lexicogrammatical, etc.) that are widely used and widely intelligible across groups of English speakers from different first language backgrounds. (Jenkins, 2006, p. 161)

A variety of world Englishes is acknowledged; including Indian, Nigerian and Malaysian, among others (Kirkpatrick, 2021), but demands for passenger safety in the aviation industry means that they must be able to understand each other in the work environment.

The ICAO rating scale uses the term intelligibility, defining it in terms of the extent to which accent or dialect interferes with understanding (ICAO, 2010).

Malaysian students have already been learning English, in most cases Malaysian English, for some 11 years before starting university. Nonetheless, it is reported that students with poor speaking proficiency make a wide range of pronunciation errors, especially with consonants, both when reading aloud as reported by Shak et al. (2016), and in discussions (Enxhi et al., 2012). Yazid and Zaiyadi (2017) mentioned similar findings from a pronunciation test. In the context of a study using software based on students comparing their pronunciation to that of a native speaker, Uthayakumaran and Kassim (2018) observed that many tertiary level Malaysian students shared these difficulties, although they were "expected to meet a certain accuracy in pronunciation" (p. 84). However, according to Shak et al. (2016), "striving for intelligibility is the main objective rather than aiming for accuracy (to sound like a native speaker)" (p. 27). Thus, it appears there remains a tension between accuracy and intelligibility in terms of what students and teachers are seeking to achieve.

Uthayakumaran and Kassim (2018) attributed some of the difficulties to the mix of different linguistic communities in Malaysia who had created their own version of English pronunciation or 'Manglish'. "The use of Manglish in daily conversations among students have [sic] led them to believe that a proper English language is being spoken with correct pronunciation skills" (p. 83). Thus, many students are not aware of their mispronunciations.

Other sources of difficulty have been identified, such as language transfer, or interference, from the phonetics of the learner's first language (L1), possibly related to orthography in some cases, and phonological awareness (Derakhshan & Karimi, 2015: Sinha et al., 2009). Darmi (2013) mentioned that the "rhythms, and stress patterns of English" were

also affected by the L1. Fossilisation of errors can occur, "specially errors in pronunciation" (Touchie, 1986, p. 78). Additional sources of difficulty, among others, include anxiety, for example test-related anxiety and fear of negative evaluation (Miskam & Saidalvi, 2019, p. 3), as well as the role of motivation to learn the language, which is influenced by "the social interaction between the teacher and the learner" (Anjomshoa & Sadighi, 2015, p. 135). Thus, there are many possible reasons why students at a Malaysian university may have problems with English pronunciation.

BAET students need to achieve, maintain and ideally exceed the ICAO mandatory standard of spoken English for part or all of their careers in the aviation industry. In this setting, it is important for teachers to provide feedback on the pronunciation errors that interfere with intelligibility. Corrective feedback from teachers on pronunciation has been shown to increase gains from learning (Lyster et al., 2013, p. 22). "Learners benefit from CF [corrective feedback] on both perception and production of pronunciation" (Couper, 2019, p. 43). Typically, in lessons at the research site, teachers correct the most noticeable pronunciation errors, often using repetition as close to the error as possible, for example when a student is responding to the teacher or a classmate.

However, it is not appropriate for teachers at the research site to correct student errors during summative oral assessments which count towards their final semester grades. Feedback is delayed until all students have completed their oral assessment and, for students preparing for a summative oral assessment, feedforward anticipation of the likely outcome would perhaps be more useful. Such oral assessments offer students opportunities to anticipate and self-correct likely mispronunciations through the use of self-assessment. Self-assessment in language learning has moved away from its traditional use as "an alternative means to assess learners' ability" (De Saint Léger, 2009, p. 159) towards more active learner involvement in a developmental learning process and

greater motivation and independence (Hosseini & Nimehchisalem, 2021; De Saint Léger, 2009).

In order to overcome the lack of clarity about expectations regarding accuracy and intelligibility, as well as to increase students' awareness of their mispronunciations and to foster independent learning, it is helpful to provide a set of standards or criteria that explain what students should be aiming for (Andrade et al., 2008; Hosseini & Nimehchisalem, 2021). However, it was not straightforward to identify suitable standards or criteria for assessing pronunciation.

Existing language proficiency rating scales, such as IELTS (International English Language Testing System), TOEFL (Test of English as a Foreign Language), or the ICAO scale in this present study, are sometimes used by teachers to provide guidance about what students can expect in tests. However, a number of scales have been criticised by raters for "the lack of clear and exact wording of descriptors, the wording inconsistency of descriptors across different levels within a rating scale and the length of descriptors within a level" (Zhong, 2019, p. 144). Harding (2017) provided an example of problems with pronunciation assessment using the Common European Framework of Reference for Languages (CEFR) scale, as identified by a focus group of raters. Firstly, he observed that there was a phonological control scale rather than direct mention of pronunciation and that this scale "has been critiqued by researchers as lacking consistency, explicitness and a clear underlying construct" (p. 16). Some of the raters in Harding's study highlighted the "difficulty of assessing pronunciation in isolation from other elements of the speaking construct, particularly fluency but also grammar" (p. 26). Some aspects of the descriptors "appeared to be irrelevant to assessing pronunciation" (p. 20) such as the inclusion of "foreign accent" which was considered "anachronistic" (p. 30) and, moreover, was only mentioned at the lower levels and then presumed to disappear as other elements of pronunciation progressed. In view of these issues, a pronunciation self-

assessment checklist therefore needs to be as clear as possible in defining the construct and the descriptors.

There are two further important considerations, the first of which is usability or ease of use (Harding, 2017; Zhong, 2019) and is similar to what Bachman and Palmer (1996) term practicality. The second is the broader concept of usefulness, which Bachman and Palmer (1996) describe as consisting of "six test qualities – reliability, construct validity, authenticity, interactiveness, impact and practicality" (p. 17). This is explained further in Section 3.4.5.

It has been proposed that student self-assessment has a number of potential benefits in language learning, including heightened awareness, critical thinking and a more student-centred approach (Dlaska & Krekeler, 2008: Jamrus & Razali, 2019; Salehi & Daryabar, 2014; Trofimovich et al., 2016). However, despite many studies of self-assessment in language learning, there remains a gap regarding pronunciation. From an applied linguistics perspective, the gap includes "research targeting second language (L2) pronunciation, which refers here to the linguistic characteristics underlying listener-based global constructs such as accentedness (nativelikeness) and comprehensibility (ease of understanding) in L2 speech" (Trofimovich et al., 2016, p. 122). Trofimovich et al. (2016) suggest that consideration should be given to researching how to reduce the gap between L2 speakers and native listeners, as well as investigating the relative effectiveness of various ways of developing the skills needed for self-assessment. From the perspective of bringing research and teaching closer together, there is a gap in self-assessment of pronunciation in the classroom, which merits little more than a page in a chapter on assessment of pronunciation in the classroom by Isbell and Sakai (2022).

This gap in the literature combines with the present and future needs of the aviation industry, the students (would-be aircraft maintenance technicians and others), as well as

the desire of the English lecturers to enhance the quality of the courses offered by the institution, to create the focus for this study. The development of a pedagogical tool to assist students to self-assess their pronunciation requires a firm foundation with regards to the construct, descriptors, scale, and usefulness of the instrument. The resulting instrument and the learning that occurs as a result of its use will enable students to better prepare for their ICAO pronunciation assessment.

1.6 Rationale

This section explains the choice of pronunciation and a pedagogical tool for student self-assessment to address the problem situation. Initially, the motivation for this study was the need for BAET students who were aiming to become aircraft maintenance personnel to achieve a minimum Level 4 standard in English according to the ICAO Language Proficiency Rating Scale as required by the Malaysian national aviation authority. Aviation English lecturers had previously introduced the rating scale to assess students' oral presentations, mock job interviews and group discussions but found that many students continued to make many errors in pronunciation which caused difficulties in understanding. The Malaysian Examination Council high school English examinations that determine entry to university did not specifically include pronunciation in the scoring of these examinations, which was unfortunate in view of the need identified among Malaysian students for improved pronunciation (Shak et al., 2016). It was therefore decided that more focus on pronunciation at the research site could be helpful. Additionally, the teaching of English in Malaysia had been moving towards a more communicative and learner-centred approach for some years (Ming, 2009; Zakaria & Shah, 2019). This meant it was appropriate to explore whether the ICAO Language Proficiency Rating Scale could be used in the classroom in a more learner-centred way.

The importance of pronunciation is not limited to the local situation. As Pennington and Rogerson-Revell (2019) argued, "pronunciation is a much more important and pervasive feature of communication than is generally recognized" and is "the foundation of messaging" (p. 1) in many "real-world contexts of transglobal and international communication" (p. 2). In response to the growth of transnational communication, multiple varieties of English have developed over time and are increasingly used by non-native speakers with differing linguistic backgrounds (Levis, 2020, p. 320).

The increase in international exchange of personnel has been accompanied by a requirement for equivalency of educational, occupational and language qualifications. In language learning, the Common European Framework of Reference for Languages (CEFR) was developed for use in "planning language learning programmes…language certification…[and] self-directed learning [including] self-assessment" across a range of languages (Council of Europe, 2001, p. 6). The CEFR also aimed to define six levels of language proficiency in positive terms, using 'can do' descriptors for communicative language skills which can be used "for continuous teacher-assessment – or self-assessment" (Council of Europe, 2001, p. 228). The CEFR has since been adopted by countries in continents other than Europe, including its introduction in Malaysia in 2013 (Nawai & Said, 2020, p. 29). This offered an opportunity to explore the potential of 'can do' descriptors in a self-assessment checklist that would be targeted on pronunciation to suit the context of this present study.

The interest in self-assessment arose from the shift in Malaysian education over some years away from a transmission model and towards a constructivist model of teaching and learning (Neo et al. 2007). One of the stated goals of the Malaysia Education Blueprint 2013 – 2025 (Ministry of Education, 2013) was to develop students with "the ability to independently drive one's own learning, coupled with the appreciation of the value of lifelong learning" (p. 2-5). This has led to further developments including the recent

introduction of problem-based learning in the institution where this present research was conducted. All these initiatives require a higher level of student independence than the transmission model of teaching and learning. Student independence in turn requires reflection on learning experiences and learning from mistakes; as mentioned in the context of critical thinking and metacognition, "Reflecting on how one carries things out in practice and analyzing mistakes are ways to encourage success and autonomy in learning" (Rivas et al., 2022, p. 6). Self-assessment is one way of implementing reflection in the classroom.

However, many of the students and some of the teachers at the university where this present research was carried out were not familiar with self-assessment. What was needed therefore was a pedagogical tool to support teachers as well as students in developing their understanding and practice of the self-assessment process.

1.7 Research objectives and questions

The general objective of this study was to develop and validate a pronunciation self-assessment checklist which Bachelor of Aviation Engineering Technology (BAET) students could use to help improve their English pronunciation. Given this, the following specific objectives were proposed for this study:

1. To develop a student pronunciation self-assessment checklist for Bachelor of Aviation Engineering Technology students
2. To calibrate the student pronunciation self-assessment checklist based on feedback from Aviation English teachers and students
3. To evaluate the usefulness of the student pronunciation self-assessment checklist as perceived by students and teachers

Both qualitative and quantitative methods were used to address the objectives.

Research questions

In order to achieve the objectives of developing and validating a self-assessment instrument which students can use to help improve their English pronunciation, the study addressed the following research questions:

1. What criteria should be used to design the student pronunciation self-assessment checklist?
2. What are students' and teachers' reactions to the pronunciation self-assessment checklist?
3. How do students and teachers evaluate the usefulness in terms of impact, practicality, reliability and validity of the pronunciation self-assessment checklist?

1.8 Theoretical framework

The role and uses of self-assessment have changed considerably over the last 50 years as the theories and applications of self-assessment have changed (Panadero et al., 2016a); thus, it is important to be clear about how self-assessment is operationalised in any situation or study. At present, there are several different schools of thought regarding the nature of self-assessment and how it should be carried out, leading one scholar to note that "Without exception, reviews of self-assessment (Sargeant, 2008; Brown & Harris, 2013; Panadero et al., 2016a) call for clearer definitions: What is self-assessment, and what is not?" (Andrade, 2019, p. 1). This present study operationalises self-assessment as formative assessment for learning, in accordance with Andrade and Valtcheva (2009), who state that student self-assessment is about formative assessment rather than determining or contributing to their final grades (p. 12). The purpose of self-assessment is "to generate feedback that promotes learning and improvements in performance" (Andrade, 2019, p. 2), specifically to generate "feedback for oneself from oneself"

(Andrade & Du, 2007, p. 160). Like other forms of feedback, it plays an essential role in formative assessment (Andrade & Valtcheva, 2009, p. 12).

Formative self-assessment reduces the risk of students inflating their grades when those grades do not form part of their final mark (Andrade & Valtcheva, 2009, p. 13). It is important to clarify how the term formative self-assessment is used in this present study, given that there is no single agreed definition of the term. For example, some scholars describe self-assessment as a continuous process with three stages, "self-monitoring, self-evaluation, and identification and implementation of instructional correctives as needed", a process which includes judgements on performance as well as feedback (McMillan & Hearn, 2008, p. 41). Judgements on performance can take different forms depending on their intended use. According to Andrade and Valtcheva (2009), student self-assessment is aimed at future improvement, whereas student self-evaluation is aimed at grading completed tasks (p. 13). Andrade (2019) distinguishes between formative self-assessment and summative self-assessment in terms of progress and performance and the associated types of judgement made by students. Formative assessment may involve judgements "of progress towards specific targets", while summative self-assessment involves "post-task judgments of ability based on performance" and is essentially "self-grading", without feedback (Andrade, 2019, p. 3). Ross (2006) also mentioned that both types of self-assessment include judgements, whether or not they are used to contribute to final grades, although he considered self-assessment in metacognition to be something different (p. 2).

This present study adopts the view of Andrade and Valtcheva (2009) because students need to think about their learning processes, identify areas for improvement and ways of making improvement, rather than be distracted by thoughts of their final grade. The importance of feedback has been stressed by many authors (e.g., Carless et al., 2011; Hattie & Timperley, 2007; Mamoon-Al-Bashir et al. 2016; Shute, 2008). When students

complete a task using self-assessment, it is important to create a feedback-feedforward loop by making improvements the next time they perform a similar task. Some definitions of self-assessment include "self-assessment of one's abilities, processes, and products", all of which depend on "feedback from oneself." (Andrade, 2019, para. 4). All three of these - self-assessment of one's abilities, processes, and products - are engaged in this present study which focuses on helping students to understand how to carry out self-assessment of pronunciation to assist them in achieving at least the required level for the aviation industry. In this particular setting, the pedagogical tool is directed at students learning to self-assess, in other words learning about the process of self-assessment and how to conduct it. This includes students' enquiry into their own learning habits and strategies in terms of their English pronunciation as well as identification of areas for improvement. However, students also need to be aware of the expectations of the aviation industry and indeed other employment sectors for continuous professional development, which is often linked to rising standards and changes to procedures. In their future lives they will have to be aware of these changes and to know where they need to improve. Thus, the inclusion of a required level in the pedagogical tool in this present study means there is reference to criteria and standards, which provide students with guidance on what they are aiming for, as recommended by Andrade and Du (2007) who mentioned that students welcomed such guidance (p. 165) and found self-assessment easier when they were clear about what was expected of them (p. 169).

There are several theoretical perspectives that underpin various models of self-assessment, including "1) cognitive and constructivist theories of learning and motivation, 2) metacognition theory, and 3) self-efficacy theory" (McMillan & Hearn, 2008, p.42). There were three main reasons for selecting metacognition theory as the perspective for this present study. The first reason was that many scholars have reported a close relationship between formative self-assessment and metacognition (e.g., Andrade,

2019; Hattie & Timperley, 2007; O'Malley & Chamot, 1990; Oxford, 1990; Siegesmund, 2016). For example, Andrade (2019) mentions that metacognition is one of the components of the formative assessment process, focusing on how students think about and monitor their learning rather than making judgements about progress towards learning goals (p. 3). Some scholars propose that "self-assessment may assist students to develop metacognitive skills, as they learn to recognise their own abilities and deficits" which will help them to improve (Jessner, 2018, p. 41), or by reflecting on their learning "to target metacognition" (Siegesmund, 2016, p. 205). It is argued that the concepts of metacognition and self-assessment are strongly interwoven (Jessner, 2018, p. 41).

The second reason for selecting metacognition theory was the notable increase in interest and use of metacognition in language teaching and learning in recent years (Haukås et al., 2018; Raoofi et al., 2014), although interest in the role of metacognition in language learning has been evident for some time. For example, metacognitive strategies in language learning, such as identifying opportunities to practise, paying attention, and monitoring production, were highlighted by O'Malley and Chamot (1990) and Oxford (1990). However, the increased interest since then has not been evenly spread across the language skills of reading, writing, speaking, and listening (Haukås et al., 2018). It is recognised that speaking makes "heavy demands on speakers' abilities to use metacognitive strategies" (Zhang et al., 2022. p. 1) and that integrated speaking tasks could be used by teachers to highlight these strategies. The present study seeks to apply this to pronunciation as one aspect of speaking on the basis it will contribute towards overall development of students' use of metacognitive strategies.

The third reason for selecting metacognition theory was its relevance to the specific context in which this present study took place. At national level, for some years Malaysian higher education has moved towards increasing use of a student-centred approach to teaching and learning as it seeks to incorporate soft skills such as thinking

skills, problem solving skills, and lifelong learning as decreed by the Ministry of Higher Education (2007). Over the last few years, teaching staff at the site where this study was conducted have introduced elements of a more student-centred approach in order to boost students' critical thinking and problem-solving skills, most recently by a shift to problem-based learning. Characteristics of student-centred learning include students being active learners, reflecting on their learning, and taking more responsibility for their learning (Lea et al., 2003), all of which are also characteristics of both self-assessment and metacognition.

This present study therefore adopts the metacognitive theory of learning, which has its origins in the work of Flavell (1979), as the main theory underpinning self-assessment. The literature on metacognition is reviewed in more depth in the following section, given its importance to the present study.

1.8.1 Metacognition

Metacognition is referred to less formally as 'thinking about thinking' (Fisher, 1998; Rickey and Stacy, 2000, p. 915). This is illustrated in Figure 1-1. A person uses metacognition to decide (control) which cognitive strategies to use in a specific task, and the success of those strategies feeds back to the metacognitive level where changes of strategy are considered for the next similar task.

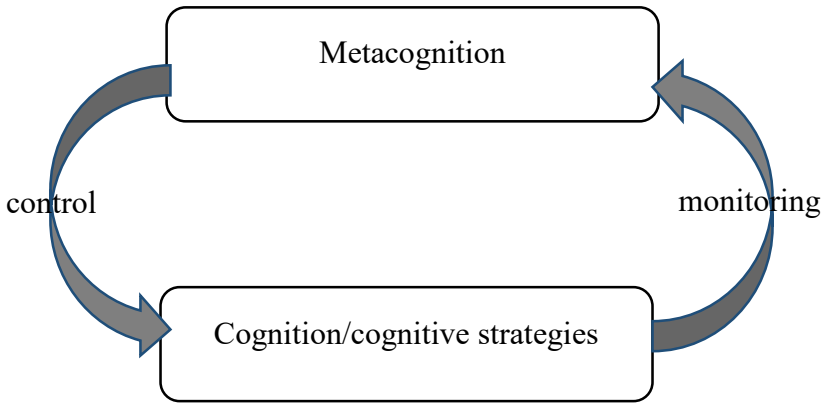


Figure 1-A Relationship between metacognition and cognition. Adapted from Nelson & Narends, 1996 and Muijs & Bokhove, 2020.

According to Haukås (2018), metacognition encompasses "an awareness of and reflections about one's knowledge, experiences, emotions and learning in the contexts of language learning and teaching" (p. 13). Evidently, this applies to pronunciation as a component of language learning.

Although metacognition has a long history (Dunlosky & Metcalfe, 2009; Coşkun, 2018), it has only become a distinct field of research since the 1970s (Tanner, 2012). Prior to that, interest had centred on reflection as a conscious process of thinking about learning. Dewey (1933) described reflection as "active, persistent and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and further conclusions to which it leads" (p. 118). Kolb (1984) clarified that the process should be cyclical and not linear. Schön (1991) explained further that there was a difference between reflection after an experience, or reflection on action, and reflection on what was happening during an experience, or reflection-in-action. Meanwhile, interest in metacognition was gaining ground.

The term metacognition is attributed to Flavell who defined it as "one's knowledge concerning one's own cognitive processes or anything related to them, e.g., the learning-relevant properties" (Flavell, 1976, p. 232). Flavell highlighted the importance of cognitive monitoring, that is, individuals' awareness and observation of their memory, their understanding, and indeed any other conscious thinking processes (Flavell, 1979, p. 106). He explained that this involved "interactions among four classes of phenomena: (a) metacognitive knowledge, (b) metacognitive experiences, (c) goals (or tasks), and (d) actions (or strategies)" (1979, p. 609). He clarified that metacognitive knowledge concerns knowledge about a task and its goal, knowledge of the possible strategies for achieving the goal, and knowledge about oneself (Flavell, 1979, p. 907), such as whether a person learns better by watching someone do something or reading an instruction manual, or believes that other ESL students are better at speaking.

Flavell also explained that metacognitive experiences are experiences which make someone aware of their thinking and provoke an insight or emotional response, for example realising they are reading an article but not understanding it, or that other people do not understand them very well when they speak English. An individual can then select a strategy to achieve the goal of the task. If the individual then thinks about how well their chosen strategy achieved the goal, they are monitoring their cognitive learning. As stated by Flavell (1979, p. 909), "Cognitive strategies are invoked to make cognitive progress, metacognitive strategies to monitor it". Kuhn highlighted that:

> There would seem few more important accomplishments than people becoming aware of and reflective about their own thinking and able to monitor and manage the ways in which it is influenced by external sources, in both academic, work, and personal life settings (Kuhn, 2000, p. 181).

Dinsmore, Alexander and Loughlin (2008) noted that the meaning of metacognition was not always made clear, despite the volume of literature on this subject. They added that this lack of definition was unfortunately often found in "central concepts and constructs, such as knowledge, learning, or motivation", when authors assumed that readers would understand or did not consider the consequences of not being clear (Dinsmore et al., 2008, p. 392). They highlighted that this applied to the use of the terms 'metacognition', 'self-regulation', and 'self-regulated learning'. There were differences in origin and emphasis, which meant the terms should not be used interchangeably (Dinsmore et al., 2008; Muijs & Bokhove, 2020). Azavedo (2020) reported that the variety of definitions and the different ideas, assumptions and processes associated with them created a barrier to reaching a single definition. Since this present study focused on cognitive aspects, the following definition of metacognition given by Kuhn and Dean was adopted: "Awareness and management of one's own thought" (Kuhn & Dean, 2004, p. 270).

Firstly, learners need to be aware, or become aware, of their thoughts. If they are not aware of what they are thinking, they will not be aware of their mistakes or why they are making them, and will be reliant on others to point out their mistakes and tell them how to correct their mistakes. Thus, their learning will be more restricted than if they were aware of their thinking. Perkins (2008, p. 102) proposed "four levels of metacognition: tacit, aware, strategic, and reflective". According to Perkins, awareness starts at the second level, when individuals know that they use different methods of thinking and learning but do not necessarily deliberately think about them or plan which to choose. In language learning, "metalinguistic awareness includes aspects of linguistic knowledge, such as explicit knowledge of … pronunciation rules" (Vold, 2018, p. 68) and students need to be aware of these in order to detect their own mistakes. At the strategic level, individuals will apply their different methods of thinking and learning in an organised way, through using strategies such as decision making and problem solving. At the

reflective level, they will "reflect on their thinking-in-progress, ponder their strategies and revise them" (Perkins, 2008, p. 102).

1.8.2 Metacognition and reflection

There is a close relationship between metacognition and reflection. Muijs and Bokhove (2020) indicated that monitoring and control of thinking processes "includes at least three main components: planning, monitoring and evaluation" (p. 6). Each stage requires reflection on what is being done, and why and how well it is being done. The planning component includes setting the task goal, bringing into play existing relevant knowledge, and choosing which approach and methods to apply. Monitoring involves checking that the approach and methods are being applied, while evaluation looks back at whether the approach and methods were applied as intended and considers what else could have been done. In language learning, Anderson (2008) proposes teachers need to be involved in explaining and modelling five components of metacognition, which he describes as "preparing and planning for learning, selecting and using strategies, monitoring learning, orchestrating strategies and evaluating learning" (p. 100). In performing a particular task or activity, the extent to which planning and selecting strategies are separate may not always be clear-cut, and the orchestration of strategies may be wrapped up in the evaluation phase. Zimmerman and Moylan (2009) offer a model which brings together metacognition and motivation; they use three different terms, namely 'forethought', 'performance' and 'self-reflection', adding various behavioural aspects of motivation, self-control and self-evaluation respectively to these stages. The terms 'performance' and 'self-reflection' have been avoided in the model used in this present study to avoid possible strong suggestions of either high-stakes testing (resulting from the use of 'performance') or reflection about the person rather than the thinking processes involved (from 'self-reflection').

There are other models of reflection, such as Kolb's experiential learning cycle which highlights experience first, then reflection, then conceptualisation or thinking about the theory to understand what happened, and finally applying how things could have been done differently (Kolb, 1984). In contrast, metacognition implies thinking before the experience as well as afterwards. Another widely used model is the reflective practitioner, which was intended to help professionals such as nurses and teachers think about what they were doing and improve their professional practice as well as develop theory (Schön, 1991). While the names and numbers of the stages are different and may contain different elements, they all acknowledge the importance of reflection. Moreover, referring to the importance of reflection, Boud and Molloy (2013) state that opportunities for reflection can be found throughout a task, in the preparation for a task, during the task and afterwards. From the point of view of metacognition, the feedback cycle must include thinking about the thinking and learning processes in the before, during and after stages of an activity as well as the end product. This type of reflective thinking about the stages of a learning task is summarised in Figure 1-2.

Figure 1-2 retains the emphasis on reflection ('thinking about thinking') throughout the process. It can be seen that this includes reflection on each stage of the process, contributing to the central overall reflection that feeds forward into the next task. As Flavell (1979) observed, both cognitive and metacognitive strategies can be called upon to make and monitor progress in thinking and learning while carrying out a task (p. 909). Moreover, "the interplay [between them] continues until the enterprise comes to an end" (p. 909). In other words, "metacognitive reflection" (Muijs and Bokhove, 2020, p. 24) is a continuous process throughout the task, as proposed by Boud and Molloy (2013).

The three main stages in Figure 1-2 are based on those identified by Muijs and Bokhove (2020) and Zimmerman and Moylan (2009) but are titled in relation to metacognition rather than self-regulation.

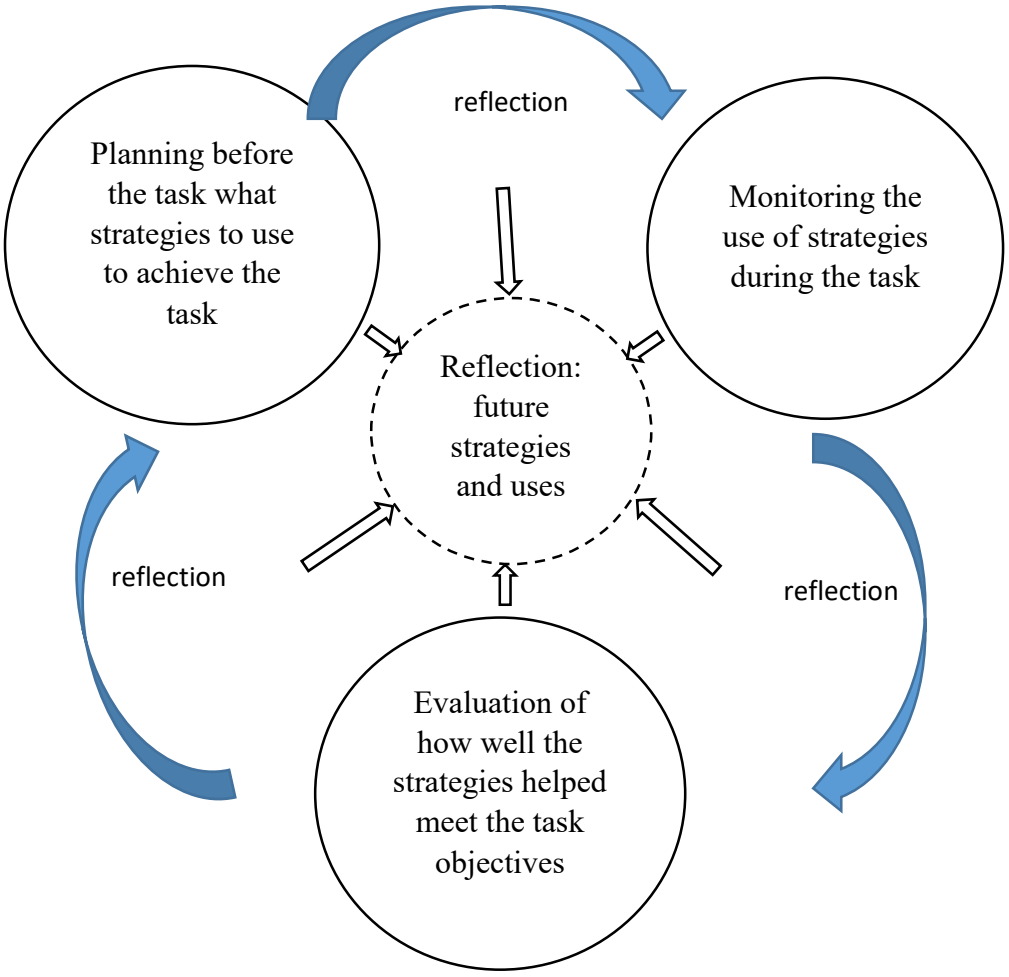


Figure 1-B Simplified metacognitive self-assessment cycle. Adapted from information from Boud and Molloy (2013); Flavell (1979), and Perkins (2008), and figures from Muijs and Bokhove (2020, and Zimmerman and Moylan (2009).

In Figure 1-2, the blue arrows indicate the basic cycle of reflection. These separate stages of reflection are added together in a central body of reflection, as shown by the white arrows, that further develops the metacognitive strategies and knowledge that are available to use when additional similar tasks are carried out.

In this present study, the pronunciation self-assessment checklist was designed to encourage reflection at each of the three stages. In the planning stage, before the activity, students were required to think about how they would pronounce the words and whole phrases and sentences they intended to use in their spoken activity. This would also prompt them to think about the pronunciation they needed to perform the task. They were

then required to think about their pronunciation during the activity with the aim of making them more aware of the thinking processes involved in the second stage. After the activity, the checklist prompted students to reflect on whether their pronunciation had enabled their listeners to understand them.

1.9 Contribution to knowledge

This thesis contributes an addition to knowledge in the following ways:

- It adds to the checklists available for formative self-assessment in learning English as a foreign or second language. In doing so, it addresses an unmet need for a pedagogical tool aimed at raising awareness of key elements of the self-assessment process, namely:
  - assessment criteria;
  - students' own thinking processes; and
  - language learning strategies
- Specifically, it provides a pedagogical tool aimed at enabling students to develop their understanding of how to self-assess their pronunciation by stimulating them to think about their learning strategies in a particular task and to consider alternatives. It highlights awareness of the processes, rather than the product, of self-assessment.
- Whilst the checklist was designed for use in the context-specific setting of English classes embedded within a Bachelor's degree programme for aircraft engineering technology students in an ESL environment, it can potentially be employed in other similar environments.
- The development research approach utilised in this study offered a way of involving end users alongside experts at an early stage and hence maximising their

opportunity to contribute to the checklist, increasing the likelihood that the checklist would be used. Development research has been shown to be a useful addition to the researcher's toolkit.

1.10 Key terms

The key terms related to the study are defined both conceptually and operationally as follows. The conceptual definition is followed by the operational definition.

Pronunciation

Pronunciation includes mastering individual sounds, understanding intonation (the rise and fall of the voice in speaking), rhythm (the flow of words and phrases determined by the relation of long and short or stressed and unstressed syllables) and stress (some sounds in words and some words in sentences are pronounced with greater force or more clarity than others). As mentioned by Isaac and Harding (2017) this means that pronunciation is a combination of both segmental (individual sounds or phonemes) and suprasegmental (intonation and stress) (p. 348). Similarly, Yates (2017) described pronunciation as "the way utterances are articulated" and asserted that in language teaching and learning, focus on pronunciation should be given to "sounds, stress in words and utterances, intonation patterns, pitch, range and variation" (p. 228).

Within the scope of this study, pronunciation is perceived from the perspective of clarity in pronunciation, accuracy in word stress, regular rhythm, natural intonation, and intelligibility.

Self-Assessment

Self-assessment refers to a process "when students judge their own work to improve performance as they identify discrepancies between current and desired performance" (McMillan & Hearn, 2008, p. 40). At the conceptual level, there is no consensus among

scholars whether it is formative, summative or a combination of the two, although there is broad agreement that it "generate[s] feedback that promotes learning and improvements in performance" and should be primarily learning-oriented (Andrade, 2019, p. 2). Self-assessment has been operationalised using a variety of purposes and methods, including the use of pedagogical tools such as self-assessment checklists.

In this present study, self-assessment is operationalised as learning-oriented self-assessment using a checklist inspired by and adapted from The Self-Assessment Checklist for ESL Argumentative Writing released by Nimehchisalem et al. (2014), additionally drawing on other sources related to metacognition and pronunciation. This study focused on developing and validating the checklist which teachers can use to help students to reflect on how they can improve their English pronunciation. The checklist used in this study is divided into three sections, namely: Before the speaking activity, During the speaking activity, and After the speaking activity.

Usefulness

The usefulness of a language test or assessment, including self-assessment, is measured in terms of six qualities, namely "reliability, construct validity, authenticity, interactiveness, impact, and practicality" (Bachman & Palmer, 1996, p. 17). Usefulness depends on the "combined effect" of these qualities and "the appropriate balance among the different qualities…must be determined for each specific testing situation" (p. 18). For the self-assessment checklist in this present study, usefulness was operationalised in terms of reliability, construct validity, impact and practicality. Authenticity and interactiveness relate to the speaking task with which the checklist is used and not to the checklist itself.

1.11 Thesis structure

There are five chapters in this thesis, the first of which has presented the context and the research problem, aim and objectives, along with an overview of the whole research process. The second chapter contains a review of some of the key literature regarding pronunciation, and the criteria for developing and evaluating the checklist. This is followed by the methodology which is described and justified in Chapter 3. Chapter 4 presents the development of the checklist in narrative form, and reports quantitative and qualitative findings from the data. The thesis concludes with a discussion of the development research process as well as the findings and offers recommendations for future work.

2 CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

This chapter reviews selected literature relevant to English pronunciation, self-assessment, metacognition and checklists in order to establish the design criteria for the pedagogical tool. It begins with a review of pronunciation and the historical changes that have taken place in the teaching of pronunciation, including issues of nativeness, accentedness, intelligibility and comprehensibility in international environments. It continues with a discussion of the assessment of pronunciation in high-stakes English tests with particular reference to the ICAO language proficiency rating scale against which aircraft maintenance technicians will be assessed. Next, options for self-assessment are considered, followed by examination of learning strategies in relation to metacognition, before some of the critical factors for developing a self-assessment pronunciation checklist are examined. The chapter concludes by examining selected previous studies in the areas of assessment and self-assessment of pronunciation.

2.2 Pronunciation

The construct of pronunciation has different dimensions. One dimension is the analysis of pronunciation from the perspective of its phonetic elements. Another encompasses socio-linguistic factors such as impressionistic judgments made by individual raters and the need for non-native speakers from different L1 backgrounds to understand and be understood by each other.

2.2.1 Phonology of pronunciation

Pronunciation is a combination of segmental and suprasegmental elements. It "encompasses (1) individual consonant and vowel sounds, commonly referred to in the literature as 'segments', and (2) features that span a larger unit than a single segment, such as word stress, rhythm and intonation, referred to synonymously in the literature as 'suprasegmentals' or 'prosody'" (Isaacs & Trofimovich, 2017, p. 9). These are also

sometimes referred to as 'segmental phonology' and 'suprasegmental phonology', which highlights that there are two distinct types of relationships among speech sounds (Pennington and Rogerson-Revell, 2019, p.4). Segmentals can be further analysed in terms of units such as phonemes, consonant clusters and syllables (Deterding, 2015, p. 76), while suprasegmentals can be further analysed into processes such as linking and intonation patterns (Murphy & Baker, 2015, p. 49). The relative importance of segmentals and suprasegmentals in teaching pronunciation is still debated (Wang, 2020, p. 1). For instance, for teaching beginners, Fraser (2001, p. 52) recommended "basing lessons around words, phrases and sentences, rather than phonemes" and introducing either or both word stress and sentence stress.

However, the construct of pronunciation is not straightforward. As Deterding (2015) mentioned, the pronunciation of some vowels and consonants not only changes over time but varies between British English, American English and other Englishes in use in countries such as Singapore and India (pp .81-82). Malaysian English is also recognised as a distinct variety of English from a pronunciation perspective (Lim, 2014). The English spoken by Malaysian teachers in university classrooms is nevertheless recognised as highly intelligible by international students attending pre-sessional English courses (Teh & Pilus, 2019). The recognition of different forms of English has resulted in changes to the teaching and assessment of pronunciation in terms of how accentedness is perceived (Liu et al., 2020, p. 98). As Lewis and Deterding (2018) mentioned, "Traditionally, pronunciation was usually taught by reference to a native-speaker model, generally Received Pronunciation (RP) British English […] or General American English […]" (p. 161). In today's world, where English is the lingua franca of many non-native speakers in many occupations, it can no longer be assumed that "native speakers own the language" (Deterding, 2015, p. 82).

2.2.2 Teaching pronunciation: revolutions or evolution?

"The history of pronunciation in English language teaching is a study in extremes", according to Levis (2005, p, 369). Pronunciation has sometimes been central to the teaching of English and at other times completely marginalised. Preferred forms of pronunciation have been influenced by the varieties of English spoken in former British colonies and, later, the influence of English for purposes such as international education and careers, as well as "communication among people from *different* first language backgrounds" (Seidlhofer, 2005, p. 339).

Prior to the 1880s, the Grammar Translation method was used mainly for teaching reading, writing and translation of foreign language texts and little attention was given to speaking (Howatt & Widdowson, 2004, p. 151; Murphy & Baker, 2015, p. 37). As interest grew in the late 19th century in learning foreign languages for oral communication, teachers used the target language rather than translation (Hodgetts, 2020, p. 11).

An important early development in the 1880s to the early 1900s was the Reform Movement, which was responsible for "the prioritising of the spoken language over the written" (Setter & Jenkins, 2005, p. 2). It led to the formation of the International Phonetic Association and the production of the International Phonetic Alphabet (IPA) (Murphy & Baker, 2015, p. 38). The IPA was intended to be "capable of representing the full inventory of sounds of all known languages" (Setter and Jenkins, 2005, p. 2) and therefore focused on distinctions between sounds. The International Phonetic Association asserted the importance of phonetics training for teachers and learners alike; "phonetics was viewed as a vital tool for language learning" (Hodgetts, 2020, p. 12). The emphasis in teaching was initially on individual vowels and consonants, listening followed by imitation. Although subsequent changes to teaching languages turned

attention away from pronunciation, the influence of the Reform Movement has never been completely lost (Setter & Jenkins, 2015, p. 2).

The next major shift was the rise of the Audiolingual Method (ALM) in the 1950s, which was underpinned by a behaviourist approach to language learning and relied heavily on repetition (Pennington & Rogerson-Revell, 2019, pp. 120-121). The ALM adopted an "analytic-linguistic approach, explicitly focusing the learner on the building blocks of the phonological system" (Hodgetts, 2020, p. 13), often through "a series of listen and repeat drills" (p. 12). Since the ALM was based on behaviourist psychology, accurate sound production and error correction were encouraged from the outset, beginning with individual sounds (Hodgetts, 2020, p. 13), and typically supported by plenty of imitative drilling in language laboratories (Murphy & Baker, 2015, p. 47). Correct pronunciation was assumed to be that of a native speaker.

Further developments occurred in linguistics on the one hand, and pedagogy on the other. From the point of view of teaching, The *Manual of American English Pronunciation* (MAEP) was "a course text dedicated to pronunciation teaching" first published in 1951 and last revised in 1985 (Murphy & Baker, 2015, p. 46). In the context of teaching pronunciation, contrastive analysis helped to identify L1 interference and direct attention to capitalizing on similarities between L1 and L2 while distinguishing between sounds that cause difficulty (p. 14).

The eighth edition of *Gimson's Pronunciation of English* was published in 2014 (Cruttenden, 2014), since its first edition in 1962, which indicates it has had a lasting role in the teaching and learning of pronunciation. However, as the 1960s progressed, developments in the theory of linguistics and language acquisition challenged the notion that language could be learned by behaviourist principles alone (Hodgetts, 2020, p. 15-16; Pennington & Rogerson-Revell, 2019, p. 121). Chomsky's theory of Universal

Grammar proposed that children could not learn everything by repetition alone but had an innate ability that needed to be activated; this innate ability may not continue into adulthood, but the idea of activating a fundamental ability perhaps influenced new approaches to ESL teaching such as the Silent Way and the Natural Approach (Murphy & Baker, 2015, p. 50). The relative overall importance of pronunciation correspondingly declined (Pennington & Rogerson-Revell, 2019, p. 121).

Meanwhile, interest was growing in contrastive analysis, which could help identify differences between L1 and L2 phonology and "highlight areas where L1 transfer errors were likely to occur" (Jenkins, 2004, p. 109). This differed from a strictly behaviourist approach which assumed all errors required correction in order for learners to develop correct habits. As Hodgetts (2020) observed, "The assumption of the seriousness of an error in terms of intelligibility is obviously of great importance for pronunciation instruction because if the pronunciation error does not interfere with communication, it follows that it is of little importance" (p. 17). This raises questions regarding how pronunciation errors are assessed in high-stakes language tests.

From the 1970s to the 1980s, a major shift occurred as the socio-cultural aspects of language learning and a constructivist philosophy in education led to Communicative Language Teaching (CLT) and a corresponding decline in teaching pronunciation (Levis & Sonsaat, 2017, p. 268). Galaczi et al. (2016) asserted that "pronunciation (phonology) is neglected" in CLT (p. 157). This may be partly due to the importance attached to authentic and meaningful activities in CLT, which encouraged participation and "risk taking" in communication rather than accurate pronunciation (Pennington & Rogerson-Revell, 2019, p. 122). However, the focus on integration of skills "means that pronunciation tasks should, in theory at least, be integrated with other skills (typically reading or listening)" (Hodgetts, 2020, p. 23). The emphasis given to authentic meaningful activities was also evident in task-based language teaching (Hismanoglu, M.

& Hismanoglu, S., 2011; Murphy & Baker, 2015, p. 42), as well as in an approach called DOGME which built classes around materials sourced from learners, with grammar and vocabulary taught as they emerged according to students' needs to communicate about the materials (Meddings & Thornbury, 2009). According to Meddings and Thornbury who developed the approach, DOGME offered "a different way of being a teacher" (2009, p. 21). The shift to a communicative approach meant that it was more important for students to be able to express themselves and understand each other rather than to learn specific items of grammar or try to sound like native speakers (Hismanoglu, M. & Hismanoglu, S., 2011, p. 47).

Since the 1990s, pronunciation teaching approaches and the resources available to teachers and learners have diversified. For example, Patterson and Rogerson-Revell (2019) refer to recent books on phonetics and phonology which include sections on teaching and learning, such as "Roach's (2009) *English Phonetics and Phonology*…[and] Cruttenden's (2014) *Gimson's Pronunciation of English*" (p.414). Moreover, the use of technology to assist the learning of pronunciation is increasing (Levis, 2007; Pennington & Rogerson-Revell, 2019); a variety of textbooks and suggested activities have been published by teachers specialising in this field (Murphy & Baker, 2015, pp. 52-54) and online resources are increasing, as is advice to teachers on how to evaluate them (e.g., Yoshida, 2018). Additionally, technology holds considerable promise for the automatic detection and correction of pronunciation errors (Levis, 2007) but erroneous feedback and equivalence of error detection between humans and automated speech recognition remain key challenges (Pennington & Rogerson-Revell, 2019, pp. 240-270).

It has also been suggested that teacher training has broadened approaches to teaching pronunciation. In Europe, at least, more ESL teachers have qualified through a general teacher training route which promotes ideas and techniques for giving students more choice, encouraging them to reflect on their learning, and consider how they react to

activities such as speaking in class, as well as providing the teachers with techniques for teaching English to speakers of other languages (Hodgetts, 2020, pp. 31-34). This does not necessarily directly assist with teaching pronunciation but makes teachers who have been trained in a constructivist tradition more likely to identify and respond to needs identified in the classroom for more focus on pronunciation. The place of pronunciation in ESL teaching may therefore vary according to the teacher, school or faculty involved, along with the sociocultural and political context.

More generally, the native speaker model, whether American or British, has been replaced in many cases by the communicative needs of non-native speakers from a range of linguistic backgrounds to study and work together. This means that they ideally need exposure to a range of accents and some may need to modify their own pronunciation so that they can be understood by a range of listeners. Since 1995, there has been increasing research into methods and resources for teaching pronunciation (Hodgetts, 2020, pp. 32-37; Murphy & Baker, 2015, pp. 57-58; Pennington & Rogerson-Revell, 2019, p. 402). Some of this research has examined whether teaching segmentals or suprasegmentals is more likely to enhance pronunciation; a review of research indicates there is no clear view on this (Wang, 2020). In terms of what affects the assessment of pronunciation, especially in view of the multiple varieties of English, there has been research into wider issues of accentedness and intelligibility (e.g., Derwing & Munro, 2005; Jenkins, 2002; Levis 2005, 2020; Munro & Derwing, 1995).

Some of the historical developments can be described as revolutions, notably the introduction of the IPA and the shift from a behaviourist to a constructivist approach to language teaching, with the rise of CLT. At the same time, the persistence of *Gimson's Pronunciation of English* and the progression of technologies from the ALM to newer technologies are evidence of evolution. Even the tension between nativeness and accentedness, along with the search to redefine assessment in terms of intelligibility and

comprehensibility, is the result of the evolution, rather than revolution, of the teaching of English pronunciation.

2.2.3 Nativeness, accentedness, intelligibility and comprehensibility

The development of English as a lingua franca means that sounding like a native speaker is no longer considered the aim (e.g., Derwing & Munro, 2008; Isaacs, 2008; Kirkpatrick, 2011). A strong accent may not affect intelligibility or comprehensibility when the words used by the speaker can still be recognised and understood by the listener (Derwing & Munro, 1997). Some scholars distinguish between intelligibility as related to the speaker's production and comprehensibility as related to the listener's reception, while other scholars use one of the terms to include both aspects. Various high-stakes test rating scales use 'intelligibility', although the ratings are unavoidably subjective to some extent because they depend on the listener-assessor as well as the scales (Isaacs & Trofimovich, 2012, p. 477). Isaacs and Trofimovich therefore argue that comprehensibility is the more appropriate concept for assessors to use. Interestingly, the ICAO rating scale uses intelligibility, but defines it in terms of the extent to which accent or dialect interferes with understanding (ICAO, 2010). While raters' assessments can be affected by a shared L2 and some accents may be easier to understand than others, differences among raters' assessments have not been shown to significantly impact the overall grades awarded (e.g., Huang et al., 2016; Winke et al., 2013).

Some aspects of pronunciation may contribute more to ease of understanding than others but as yet there is no consensus on which these are. Thus, Jenkins (2000) reported in her proposal for a lingua franca core of pronunciation that word stress was perhaps not important, whereas many segmental sounds were essential. In contrast, Isaacs and Trofimovich (2012) identified that word stress and intonation were also important (p. 498). With the aim of further clarifying which elements of pronunciation to incorporate into the checklist, the next section examines how the various elements of pronunciation are included and defined in language proficiency rating scales, especially the ICAO scale.

2.3 Pronunciation assessment criteria in high-stakes English tests

The almost total absence of pronunciation scales in the literature as opposed to the scales found in high-stakes English tests explains why it was decided to examine how pronunciation was operationalised in high-stakes tests. Moreover, rating scales play an important role in classroom assessment in general and in this present research, they are specifically important for speaking assessment purposes. Lombard (2017) described rating scales as:

> (…) suitable tools for teaching, learning and assessment; they support transparency in as far as learning expectations, assessment criteria, performance standards and the attainment of these are concerned; they serve as sources for feed forward and feedback; they allow for more objective, reliable and valid assessment; they serve as basis for scholarly dialogue between lecturers and students and even among students; they serve as reflective tools for lecturers and students and they potentially encourage self-regulated and self-directed learning (Lombard, 2017, p. 52).

Thus, they have the potential to inform the design of a self-assessment checklist.

Two of the three tests examined were probably the most widely used internationally for university admissions, namely TOEFL (Test of English as a Foreign Language) and IELTS (International English Language Testing System). The third test, the OET (Occupational English Test), was selected as an example of an internationally used occupational test, which could potentially offer a comparison with the ICAO proficiency rating scale.

The TOEFL independent speaking rubric includes pronunciation within a wider construct of delivery, as highlighted in Table 2-1. The five bands are Advanced (4), High-

Intermediate (3), Low-Intermediate (2), Below Low-Intermediate (1) and 'no attempt' or 'unrelated response' (0). The advanced band descriptor indicates a separation of pronunciation from intonation, which equates to distinguishing between segmentals and suprasegmentals. It also contains elements of fluency and intelligibility. The construct definition is likely influenced by the way in which automated scoring detects and scores these individual elements before combining them into an overall score, as described in Xi et al. (2008). This construct is too broad to be adapted for use in a self-assessment instrument which focuses on an identified need for pronunciation improvement.

Table 2-1 TOEFL independent speaking rubric

SCORE	GENERAL DESCRIPTION	DELIVERY
4	The response fulfills the demands of the task, with at most minor lapses in completeness. It is highly intelligible and exhibits sustained, coherent discourse. A response at this level is characterized by all of the following:	Generally well-paced **flow (fluid expression)**. Speech is clear. It may include minor lapses, or minor difficulties with **pronunciation or intonation patterns**, which do not affect overall **intelligibility**.
3	The response addresses the task appropriately but may fall short of being fully developed. It is generally intelligible and coherent, with some fluidity of expression, though it exhibits some noticeable lapses in the expression of ideas. A response at this level is characterized by at least two of the following:	Speech is generally clear, with some fluidity of expression, though minor difficulties with pronunciation, intonation, or pacing are noticeable and may require listener effort at times (though overall intelligibility is not significantly affected).
2	The response addresses the task, but development of the topic is limited. It contains intelligible speech, although problems with delivery and/ or overall coherence occur; meaning may be obscured in places. A response at this level is characterized by at least two of the following:	Speech is basically intelligible, though listener effort is needed because of unclear articulation, awkward intonation, or choppy rhythm/pace; meaning may be obscured in places.
1	The response is very limited in content and/or coherence or is only minimally connected to the task, or speech is largely unintelligible. A response at this level is characterized by at least two of the following:	Consistent **pronunciation, stress and intonation** difficulties cause considerable listener effort; delivery is choppy, fragmented, or telegraphic; frequent pauses and hesitations.

SCORE	GENERAL DESCRIPTION	DELIVERY
0	Speaker makes no attempt to respond OR response is unrelated to the topic	

Source: Educational Testing Service, 2019

Turning to the IELTS speaking rating scale, this has a clearly defined criterion of pronunciation within the overall construct of speaking. Fluency is a separate criterion, along with vocabulary and grammatical range and accuracy. There are nine bands, three of which (3, 5 and 7) represent stages between others, along with a band 0 for non-attendance, as shown in Table 2-2.

Table 2-2 IELTS pronunciation rubric

BAND	PRONUNCIATION
9	**Uses a full range of phonological features** to convey precise and/or subtle meaning. Flexible use of features of connected speech is sustained throughout. Can be effortlessly understood throughout. Accent has no effect on **intelligibility**.
8	Uses **a wide range of phonological features** to convey precise and/or subtle meaning. Can sustain **appropriate rhythm. Flexible use of stress and intonation** across long utterances, despite occasional lapses. Can be easily understood throughout. Accent has minimal effect on **intelligibility**.
7	Shows all the positive features of Band 6 and some, but not all, of the positive features of Band 8
6	Uses **a range of phonological features,** but control is variable. Chunking is generally appropriate, but **rhythm may be affected by a lack of stress-timing** and/or a rapid speech rate. Some effective use of intonation and stress, but this is not sustained. **Individual words or phonemes may be mispronounced** but this causes only occasional lack of clarity. Can generally be understood throughout without much effort
5	Shows all the positive features of Band 4 and some, but not all, of the positive features of Band 6
4	Uses some acceptable phonological features, but the range is limited. Produces some acceptable chunking, but there are **frequent lapses in overall rhythm**. Attempts to use **intonation and stress, but control is limited**. **Individual words or phonemes are frequently mispronounced**, causing lack of clarity. Understanding requires some effort and there may be patches of speech that cannot be understood.
3	Shows some of the features of Band 2 and some, but not all, of the positive features of Band 4

BAND	PRONUNCIATION
2	Uses **few acceptable phonological features** (possibly because sample is insufficient). Overall problems with delivery impair attempts at connected speech. **Individual words and phonemes are mainly mispronounced** and little meaning is conveyed. Often **unintelligible**
1	Can produce **occasional individual words and phonemes that are recognisable**, but no overall meaning is conveyed. **Unintelligible**

Source: ielts.org

As with other high-stakes English tests, it is rarely possible to access what lies behind words such as 'frequent' or 'wide range', although publicly available information states that "Pronunciation (or phonological) features tested in IELTS or any other high-stake English examination includes the pronunciation of individual sounds, word and sentence stress, and intonation" (www.ielts-mentor.com). However, a review of bands and descriptors for the higher levels of pronunciation (Isaacs et al., 2015) provided insights into the difficulties raters experienced with reaching agreement about definitions and differentiating between bands at higher levels. While comprehensibility and segmental error significantly differentiated among bands 5 to 7, word stress was the only pronunciation-related feature that discriminated between bands 7 and 8 (Isaacs et al., 2015, p. 24). Some participants reported that the effect of pronunciation on coherence and fluency could result in a further reduction in marking, while others felt that they used fluency or lexical resource to help them assign a mark to pronunciations. The notion of 'a full range' was interpreted in terms of nativeness by some, but in terms of intelligibility by others (p. 29). To try to provide clarity about how to interpret 'at times' or 'frequently', the authors provided a guide for each feature to be assessed (p. 45). based on a 'smiley button' scale as illustrated in Figure 2-1.

Figure 2-A Format of guidance for raters on assessing frequency (Isaacs et al., 2015, p. 45)

However, this type of scale does not necessarily resolve the issue that individual raters may well have different views about what is 'frequent' or 'infrequent', and this is likely to be an issue with any test or self-assessment instrument.

The review concluded that future revisions might include "greater definitional clarification" of the term 'phonological features'. Thus, even high-stakes test rating scales may suffer from a lack of precision, although this does not necessarily mean they are not fit for purpose, as universities widely accept both TOEFL and IELTS scores at a specified level for university admission.

It is almost never possible to look beyond the public version of the rating scales used in high-stakes English tests. However, Isaacs et al. (2008) provide a small view of what might lie behind them.

> In the IELTS examiners' version of the scale, this first part of the descriptor is followed by further specification of selected pronunciation-specific features, including, depending on the band level, rhythm, stress, intonation, articulation of individual words or phonemes, chunking, or connected speech. Finally, by the end of the descriptor, there is some statement about the test-taker's ability to convey meaning or to be understood more or less successfully (Isaacs et al., 2008, p.8).

The OET test for healthcare professionals includes pronunciation under the broader construct of intelligibility and assigns fillers and pauses to frequency (Table 2-3). There are six bands, along with a 0 band which indicates 'no response'. It is noticeable that L1 accent is mentioned at all levels, possibly because of the importance of healthcare

professionals being able to limit opportunities to be misunderstood by their patients or their patients' interpreters. This is supported by the use of expressions such as "severe strain for the listener". The descriptor for band 3, "errors in pronunciation/stress/intonation", suggests that these three items are considered to be distinct, but at the highest band (6) 'rhythm' is also highlighted.

Table 2-3 OET 2018

BAND	INTELLIGIBILITY
6	•**Pronunciation is easily understood and prosodic features (stress, intonation, rhythm) are used effectively**. •L1 accent has no effect on intelligibility.
5	•Easily understood. •Communication is not impeded by a few pronunciation or prosodic errors and/or noticeable L1 accent. •Minimal strain for the listener.
4	•Easily understood most of the time. •Pronunciation or prosodic errors and/or L1 accent at times cause strain for the listener.
3	•Produces some acceptable features of spoken English. •**Difficult to understand because errors in pronunciation/stress/ intonation** and/or L1 accent cause serious strain for the listener.
2	•Often unintelligible. •Frequent errors in pronunciation/stress/ intonation and/or L1 accent cause severe strain for the listener.
1	•Almost entirely unintelligible.
0	•Candidate does not provide any response.

Source: OET (2018)

Publicly available guidance for OET test-takers explains clearly that intelligibility is defined as, "The impact of your pronunciation, intonation and accent on how clearly your listener can hear and understand you" (www.occupationalenglishtest.org).

A similar approach is taken in the ICAO rating scale, which was specifically designed for the aeronautical industry and especially for use with pilots, ATCs and, later, aircraft maintenance technicians. Like OET, the assessment of language is related to the occupational setting, in this case through the assumption that tests assume use of "a dialect and/or accent which is intelligible to the aeronautical community" and that this requires

a regional variety of English to be "sufficiently attenuated" to avoid any misunderstanding in radiotelephony communications (see Table 2-4).

The descriptors clearly distinguish between pronunciation, stress, rhythm and intonation, with an emphasis on ease of understanding rather than the amount of strain on the listener, in contrast to the OET descriptors. The additional notes supplied by the International Air Transport Association (IATA) make it clear in the Operational Level 4 band that pronunciation is vital for assisting non-native speakers of English to understand each other.

Table 2-4 ICAO rating scale extract

LEVEL	**PRONUNCIATION Assumes a dialect and/or accent intelligible to the aeronautical community**	**Additional notes from IATA**
Expert Level 6	Pronunciation, stress, rhythm and intonation, though influenced by the first language or regional variation, almost never interfere with ease of understanding.	An Expert Level 6 speaker may be a speaker of English as a first language with a widely understood dialect or may be a very proficient second-language speaker, again with a widely used or understood accent and/or dialect. The speakers' accent or dialect may or may not identify them as second language users, but the pronunciation patterns or any difficulties or mistakes almost never interfere with the ease with which they are understood. Expert speakers are always clear and understandable.
Extended Level 5	Pronunciation, stress, rhythm and intonation, though influenced by the first language or regional variation, rarely interfere with ease of understanding.	Extended Level 5 speakers demonstrate a marked accent, or localized regional variety of English, but one which rarely interferes with how easily understood their speech is. They are always clear and understandable, although, only occasionally, a proficient listener may have to pay close attention

LEVEL	PRONUNCIATION Assumes a dialect and/or accent intelligible to the aeronautical community	Additional notes from IATA
Operational Level 4	Pronunciation, stress, rhythm and intonation are influenced by the first language or regional variation, but only sometimes interfere with ease of understanding	**Operational Level 4 speakers demonstrate a marked accent, or localized regional variety of English. Occasionally, a proficient listener may have to pay close attention to understand or may have to clarify something from time to time. Operational Level 4 is certainly not a perfect level of proficiency; it is the minimum level of proficiency determined to be safe for air traffic control communications. While it is not an Expert level, it is important to keep in mind that pronunciation plays the critical role in aiding comprehension between two non-native speakers of English.**
Pre-operational Level 3	Pronunciation, stress, rhythm and intonation are influenced by the first language or regional variation and frequently interfere with ease of understanding.	**Accent at this Pre-operational Level 3 is so strong as to render comprehension by an international community of aeronautical radiotelephony users very difficult or impossible. It should be noted that native or second language speakers may be assessed at this level in cases where a regional variety of the language has not been sufficiently attenuated.**
Elementary Level 2	Pronunciation, stress, rhythm, and intonation are heavily influenced by the first language or regional variation and usually interfere with ease of understanding	N/A
Pre-elementary Level 1	Performs at a level below the Elementary level	

Sources: ICAO (2010). Publication of Doc 9835-AN/453, Manual on the Implementation of ICAO Language Proficiency Requirements; www.iata.org.

The issue of how to determine and distinguish between frequency measures such as 'usually' and 'frequently' remains, but the use of frequency is consistent, whereas in the OET scale, there is a mixture of frequency measures with 'difficult' and 'easy'. In high-stakes tests, these issues can be resolved to a certain extent through rater training and

moderation. However, a self-assessment instrument may require further clarity, at least agreement between the students and teacher working with it.

ICAO (2010) also provides a set of holistic descriptors to aid raters in making judgements. These reinforce the occupational context, reminding raters to refer to the work setting. They state:

> Proficient speakers shall:
>
> a. communicate effectively in voice-only (telephone/radiotelephone) and in face-to-face situations;
>
> b. communicate on common, concrete and work-related topics with accuracy and clarity;
>
> c. use appropriate communicative strategies to exchange messages and to recognize and resolve misunderstandings (e.g., to check, confirm, or clarify information) in a general or work-related context;
>
> d. handle successfully and with relative ease the linguistic challenges presented by a complication or unexpected turn of events that occurs within the context of a routine work situation or communicative task with which they are otherwise familiar; and
>
> e. use a dialect or accent which is intelligible to the aeronautical community (ICAO, 2010, Appendix 1, p. A-2).

In addition to welcoming the use of intelligibility in high-stakes tests such as TOEFL (Table 2-1), Isbell and Sakai (2022, p. 205) mentioned one alternative pronunciation subscale as having potential for use in the classroom. This is the subscale of the comprehensibility speaking rubric developed by Isaacs et al. (2018) shown in Table 2-5.

Table 2-5 Pronunciation subscale

COMPREHEN-SIBILITY	PRONUNCIATION
5	• Pronunciation is **effortless to understand** • Errors do not interfere with the message • Pitch variation may make the speech sound lively or engaging • Sounding native-like is not expected
4	• Pronunciation requires **little effort to understand** • Errors minimally interfere with the message • Speech may be characterized by too many or too few variations in pitch, sounding disjointed or monotone
3	• Pronunciation requires **some effort to understand** • Errors somewhat interfere with the message (e.g., misplaced word stress, sound substitutions, not stressing important words in a sentence)
2	• Pronunciation is **effortful to understand** • Errors are detrimental to the message (e.g., misplaced word stress, sound substitutions, not stressing important words in a sentence) • Production difficulties may obscure the meaning of a few words
1	• Pronunciation is **painstakingly effortful to understand** • Errors are debilitating to the message (e.g., misplaced word stress, sound substitutions, not stressing important words in a sentence) • Production difficulties may make words sound slurred or indistinct
UR	Unable to Rate. Speaker does not produce an assessable sample of speech (e.g., unresponsive to the task, no articulation of English-like sounds)

Source: Isaacs et al. (2018, Appendix, p. 215-216)

This pronunciation subscale, like that of OET (Table 2-3) has the advantage of being relatively short and easy to read. It has the additional advantage of avoiding reference to L1 or accent and makes it clear that native-like pronunciation is not expected. However, -it shares certain disadvantages with some of the scales previously examined. Different components of pronunciation are mentioned at the various levels, and it focuses on how much effort is required on the part of the listener. In a self-assessment context, the listener is also the speaker, who may not be sufficiently aware of how another listener would perceive the effort required to understand them. Moreover, comprehensibility terms such as 'robotic speech' and 'problematic sounds' are open to interpretation according to the teacher's confidence and perception of their own language ability and accent. Whilst the interpretation may be assisted by training and moderation, still teachers with different

first languages might not agree on which sounds are problematic, nor on perceptions of 'limited' control.

This section has highlighted the importance of clear construct definition and clear descriptors which use consistent language. Taking the view of Pennington and Rogerson-Revell (2019) that pronunciation lies at the heart of meaningful communication through understanding (p. 6), it was deemed appropriate to consider whether the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) could shed light on the matter.

2.4 Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR)

Pronunciation is specifically addressed in the CEFR through its six-level phonological scale, which was reviewed by Piccardo in 2016, in response to growing interest in pronunciation, a desire to move away from the focus on accent (and associated native speaker model), and the need to produce a complete scale that covered all levels (Council of Europe (CoE), 2018, p. 133). The aim was to "provide both a general scale and specific ones in order for teachers/learners to both have a snapshot of their phonological competence and to identify areas for improvement" (Piccardo, 2016, p. 9). The phonological features covered were: articulation, prosody (which included speech rate and chunking), accentedness and intelligibility. The new scale, shown in Figure 2-2, contains three categories, namely overall phonological control, sound articulation, and prosodic features (intonation, stress and rhythm).

There are still problems with the interpretation of words such as 'full range', 'generally', and 'usually', but the format contains the 'can do' statements which are typically associated with the CEFR and which have the advantage of emphasizing the positives in learning progress.

	Phonological control		
	Overall phonological control	Sound articulation	Prosodic features
C2	Can employ the full range of phonological features in the target language with a high level of control – including prosodic features such as word and sentence stress, rhythm and intonation – so that the finer points of their message are clear and precise. Intelligibility and effective conveyance and enhancement of meaning are not affected in any way by features of accent that may be retained from other language(s).	Can articulate virtually all the sounds of the target language with clarity and precision.	Can exploit prosodic features (e.g. stress, rhythm and intonation) appropriately and effectively in order to convey finer shades of meaning (e.g. to differentiate and emphasise).
C1	Can employ the full range of phonological features in the target language with sufficient control to ensure intelligibility throughout. Can articulate virtually all the sounds of the target language; some features of accent(s) retained from other language(s) may be noticeable, but they do not affect intelligibility.	Can articulate virtually all the sounds of the target language with a high degree of control. They can usually self-correct if they noticeably mispronounce a sound.	Can produce smooth, intelligible spoken discourse with only occasional lapses in control of stress, rhythm and/or intonation, which do not affect intelligibility or effectiveness. Can vary intonation and place stress correctly in order to express precisely what they mean to say.
B2	Can generally use appropriate intonation, place stress correctly and articulate individual sounds clearly; accent tends to be influenced by the other language(s) they speak, but has little or no effect on intelligibility.	Can articulate a high proportion of the sounds in the target language clearly in extended stretches of production; is intelligible throughout, despite a few systematic mispronunciations. Can generalise from their repertoire to predict the phonological features of most unfamiliar words (e.g. word stress) with reasonable accuracy (e.g. while reading).	Can employ prosodic features (e.g. stress, intonation, rhythm) to support the message they intend to convey, though with some influence from the other languages they speak.
B1	Pronunciation is generally intelligible; intonation and stress at both utterance and word levels do not prevent understanding of the message. Accent is usually influenced by the other language(s) they speak.	Is generally intelligible throughout, despite regular mispronunciation of individual sounds and words they are less familiar with.	Can convey their message in an intelligible way in spite of a strong influence on stress, intonation and/or rhythm from the other language(s) they speak.
A2	Pronunciation is generally clear enough to be understood, but conversational partners will need to ask for repetition from time to time. A strong influence from the other language(s) they speak on stress, rhythm and intonation may affect intelligibility, requiring collaboration from interlocutors. Nevertheless, pronunciation of familiar words is clear.	Pronunciation is generally intelligible when communicating in simple everyday situations, provided the interlocutor makes an effort to understand specific sounds. Systematic mispronunciation of phonemes does not hinder intelligibility, provided the interlocutor makes an effort to recognise and adjust to the influence of the speaker's language background on pronunciation.	Can use the prosodic features of everyday words and phrases intelligibly, in spite of a strong influence on stress, intonation and/or rhythm from the other language(s) they speak. Prosodic features (e.g. word stress) are adequate for familiar everyday words and simple utterances.
A1	Pronunciation of a very limited repertoire of learnt words and phrases can be understood with some effort by interlocutors used to dealing with speakers of the language group. Can reproduce correctly a limited range of sounds as well as stress for simple, familiar words and phrases.	Can reproduce sounds in the target language if carefully guided. Can articulate a limited number of sounds, so that speech is only intelligible if the interlocutor provides support (e.g. by repeating correctly and by eliciting repetition of new sounds).	Can use the prosodic features of a limited repertoire of simple words and phrases intelligibly, in spite of a very strong influence on stress, rhythm and/or intonation from the other language(s) they speak; their interlocutor needs to be collaborative.

Figure 2-B CEFR phonological scale

Source: Council of Europe (2020, pp. 130-131)

Thus, to a greater or lesser extent, the CEFR, along with the high-stakes English test rating scales examined, recognises the importance of segmentals, non-segmentals and intelligibility (rather than nativeness). However, none of the scales perhaps offers descriptors in the most appropriate way for a self-assessment checklist, either because they are too long, as in the CEFR, or too condensed, as in the ICAO rating scale. It is important for a scale to be easy to use from the perspective of raters (Harding, 2016). In the classroom context, teachers and students should also find it easy to use.

2.5 The changing landscape of self-assessment

This section takes as its starting point Andrade and Du's definition of self-assessment as "feedback for oneself from oneself" (2007, p. 160). This definition is not straightforward, for several reasons. The landscape of teaching and learning continues to change, and this leads to changes in the roles and responsibilities of teachers and learners, along with understandings of self-assessment and feedback. In particular, it raises questions about what exactly is being self-assessed as well as how feedback is given.

The paradigm shift in education from behaviourism to constructivism, along with the corresponding shift from teacher-centred to learner-centred education, has been accompanied by a changing approach to assessment (Ahmad et al., 2020). The role of the teacher has moved away from an authoritative dispenser of knowledge towards a "facilitator of learning", with teachers becoming more like "partners in the learning process" (Dörnyei & Muir, 2019, p. 727). In a similar vein, assessment has extended beyond traditional testing, often with right or wrong answers, and now covers a wide range of tools and techniques, such as observations and learner portfolios, in addition to tests (Umida et al., 2020, p. 134). Umida et al. state that, in assessment, the "process [of learning] is as important as the product" (2020, p. 134). Dörnyei and Muir argue that, within a wider process of motivating and developing students, assessment assists learners' progress towards independent learning (2019, p. 728). Indeed, it has been suggested that self-assessment is central to achievement as well as motivation and development (McMillan & Hearn, 2008).

However, there is no consensus about the nature and meaning of self-assessment. A number of scholars and practitioners have adopted different viewpoints and used a variety of terms in order to clarify their position with regard to the purpose and practice of self-assessment. As Newton (2007) pointed out, "To avoid getting ourselves confused, and

to avoid confusing others, we need to use the language of assessment with greater precision" (p. 158).

2.5.1 Understandings of self-assessment

The literature reveals that, historically, there have been a number of different approaches to implementing self-assessment in the classroom (Black & Wiliam, 2009; Dochy et al., 1999; Falchikov & Boud, 1989; Taras, 2010). There remain a number of quite different conceptualizations and interpretations of what actually constitutes self-assessment. The major distinction is between summative self-assessment, where students' judgement of what they have learned contributes to their final grade, and formative self-assessment, where students' judgements about their learning processes, and possibly also what they have learned, contributes to their further learning (Andrade, 2019).

Andrade and Valtcheva (2009) drew a clear distinction between formative and summative self-assessment, describing the latter as self-evaluation. They stressed the formative nature of self-assessment, emphasizing the importance of reflection and review, and stating that self-assessment is applied to "drafts of works in progress", in addition to judging the extent to which a student's work meets "explicitly stated goals or criteria" (p. 13). They argued that "self-evaluation, in contrast, refers to approaches that involve students in grading their work, perhaps as part of their final grade for an assignment or a class" (Andrade & Valtcheva, 2009, p.13). From the viewpoint of students who are working towards a high-stakes English language test, the distinction in self-assessment between judging the extent to which their work meets criteria and self-grading may often be a narrow one.

The distinction between formative and summative self-assessment on the basis of its purpose is widely accepted (e.g., Black & Wiliam, 1998; Harlen, 2005; Panadero et al., 2016a; Yan, 2016), where the purpose of formative assessment is to assist learning, whereas the purpose of summative self-assessment is to measure what has been learned.

This distinction is typically expressed as the distinction between assessment *of* learning and assessment *for* learning. In summative assessment *of* learning, for example national or international examinations that determine admission to university or a career, reliability is highly important because decision-makers need to be able to have confidence in the results. In contrast, in assessment *for* learning, it is proposed that validity is more important than reliability, where validity depends on how "interpretations and uses of formative assessment results are specified, justified and supported" (Gu, 2021, p. 4).

However, the uses and interpretations are not necessarily straightforward, as shown in a survey of 83 teachers' and headteachers' understanding of assessment *for* learning conducted in New Zealand (Hargreaves, 2005). Participants' views fell into several categories which included teachers giving feedback, monitoring performance against objectives, and informing "next steps in teaching and learning" (Hargreaves, 2005, p. 215). Similar differences in understandings of self-assessment *for* learning exist and it is important to be clear about what is, and why and how it is being assessed (Andrade, 2019, pp. 1-4). Andrade proposed a taxonomy (see Table 2-6) based on distinctions between formative and summative self-assessment, between process and product of learning, and whether or not performance standards were involved.

Table 2-6 Andrade's taxonomy of self-assessment

Competence		**Process**		**Product**	
		Standards		Standards	
		Yes	No	Yes	No
Formative	Task-specific self-efficacy ratings	Judgments of progress toward specific targets	•Traffic lights •Comprehension checks • Self-monitoring; metacognition •Reflective journal writing	•Rubric- or checklist-referenced self-assessment •Self-testing	Open-ended critique of one's own work or understanding
Summative	Post-task judgments of ability based on performance		Post-task judgments of effectiveness of procedures	• Self-grading	• Self-grading

Source: Andrade, 2019, p. 3

Andrade acknowledges that some methods of self-assessment could appear in more than one cell in Table 2-6 and highlights that several relate directly to metacognition (2019, p. 4). Although it is recognised in the table that formative self-assessment is often conducted without incorporating standards, many scholars assert that self-assessment cannot be conducted without some reference to standards or criteria (e.g., Brown & Harris, 2013; Panadero et al., 2012; Panadero & Alonso-Tapia, 2013; Tai et al., 2018; Yan & Carless, 2022). Students need to know and understand the standards and criteria to be able to assess themselves realistically, through involvement in activities that enable them over time to internalise the standards (O'Donovan et al., 2008, p. 215). Some scholars assert that deciding the criteria is an essential first step in self-assessment (Yan & Brown, 2017; Yan & Carless, 2022). Some authors claim that students must be involved in setting the criteria in order for them to be engaged and involved in self-assessment (Tai et al., 2018; Yan & Carless, 2022). However, students who are focused on meeting external examination standards may not be motivated by "co-constructing performance rubrics" (Yan & Carless, 2022, p. 1122). The different perspectives are influenced by whether

self-assessment is viewed as a component of a student's formal education or whether it is for lifelong learning. Boud (2000), for example, not only argues that students should be actively involved in assessment, but also advocates for assessment that meets the needs of the present and future learning needs of students.

Thus, the purpose of any formative self-assessment must be made clear to teachers and students. As Andrade's taxonomy indicates, there is a need to be clear about whether it covers process or product, or both, and the methods need to suit the purpose.

The following section looks at how self-assessment of pronunciation has been implemented in language learning.

2.5.2 Self-assessment of pronunciation in English as a Second Language (ESL) speakers

Self-assessment is becoming more widely used in language learning in general, due to the spread of the CEFR in general, and to the language passport in the European Language Portfolio associated with the CEFR in particular. The language passport requires learners to use a self-assessment grid to provide a summative assessment of their language ability, while the language biography requires completion of formative self-assessment checklists (Ardnt, 2005, p. 325). Self-assessment is supported by evidence in the accompanying dossier. However, with the exception of Dlaska and Krekeler (2008) and Trofomovich et al. (2016), self-assessment of pronunciation has until recently been less well investigated than other forms of assessment of pronunciation (Isaacs & Harding, 2017). Since 2017, nevertheless, the number of studies examining the self-assessment of pronunciation has increased (e.g., Brannen et al., 2022; Jankowska & Zielińska, 2015; Khonamri et al., 2021: Navaie, 2018; Pysarchyk & Nypadymka, 2019; Strachan et al., 2019).

In recent studies, diverse approaches have been taken to self-assessment of pronunciation, including examining different components of pronunciation, reliability, and the effects of

training on various types of self-assessment. Self-assessments have been recorded through varying methods: for example, the use of yes/no checklists, ratings from 'very poor' to 'excellent', counting errors after comparison with a model, and debriefing interviews. Section 2-8 examines a selection of studies in more detail.

In terms of the effectiveness of pedagogical interventions, learners need to be supported to carry out any form of self-assessment and Little highlights that this support may need to be provided gradually, step by step, especially for learners who have mainly experienced the transmission model of teaching (Little, 2005, p. 322). Preparatory support should take account of factors such as teachers' and students' views of teaching and learning, as well as student motivation (Schmidt & Wehmeyer, 2016, p. 75). It is further suggested that students should be given guidance in a range of self-assessment methods that cover affective as well as "cognitive and metacognitive strategies" (Schmidt & Wehmeyer, 2016, p. 77). Nonetheless, it is asserted that, in the classroom context, self-assessment of pronunciation can be incorporated into curriculum objectives, for example in presentations or other speaking activities (Isbell & Sakai, 2022, p. 207).

Teachers can use a variety of self-assessment tools, such as checklists and scales, that may help students to become more aware of how well they are learning and whether they should try out other learning strategies (Isbell & Sakai, 2022). This gives teachers an important role in ensuring students encounter a range of strategies and have opportunities to try them out (Schmidt & Wehmeyer, 2016). In terms of self-assessment of pronunciation, it is said that while tools of speech recognition software are increasingly used to provide learners with an assessment of their pronunciation based on comparison to a norm, software cannot completely fulfil an assessment or self-assessment role for learner development because it lacks the ability to help them decide what to do next to improve further (Dlaska & Krekeler, 2008; Butler, 2023). Dlaska and Krekeler (2008) further assert that "the self monitoring of pronunciation is a central part of any foreign

language learning process" (Dlaska & Krekeler, 2008, p. 507). It has been proposed that teachers who encourage students to self-assess need to assist students to learn how to do it, using "a metacognitively oriented pedagogy" (Haukås, 2019, p. 22). In other words, their teaching should aim to "enhance the metacognitive processes of planning, monitoring and evaluation" (Panadero & Alonso-Tapia, 2013, p. 555). The following section considers metacognition in relation to language learning.

2.6 Metacognition and language learning strategies

This section takes as its starting point the definition of metacognition which focuses on the stages of planning, monitoring and evaluation that have tended to be associated with self-assessment in language learning (see sections 1.8.1 and 1.8.2). However, it is acknowledged that there is no single agreed definition of metacognition across the many fields in which it is researched and used and that more needs to be known about what it is and how it operates in, for example, the stages of planning and reflection in self-assessment (Azevedo, 2020).

Meanwhile, Anderson (2008) has proposed five core components of metacognition that teachers can train students to use. The first of these involves setting learning goals and reflecting on how to achieve them, while the second concerns the choice and application of learning strategies to achieve those goals. It could be argued that the selection of learning strategies is an essential part of the planning stage, i.e., planning how to do it as well as what to do. Another of Anderson's suggested components is combining strategies, which applies equally to the planning and monitoring stages. The remaining two core components are monitoring the use of strategies while carrying out a task and, finally, once the task is completed, evaluating how well the chosen strategies worked.

Studies investigating metacognition have employed a variety of methods to elicit students' thinking about their learning strategies, including think-aloud protocols while

they are performing a task, learning logs or diaries to complete after a task, and inventories of metacognitive awareness or learning strategies. Concurrent think aloud protocols have the advantage of encouraging and requiring learners to talk about their thought processes during a task but are unsuitable for use during a speaking task itself and therefore an alternative approach would be required; for speaking tasks, retrospective think-aloud protocols that aim to access thinking processes from the participant's short-term memory would be appropriate (Gu, 2014, p. 74). Learner diaries have been used in educational research for many years, to record a variety of data including reflective practice (Arndt & Rose, 2023). Checklists are another method for helping students to become more aware of their thinking processes as they work through a list of the steps needed to complete a task (Rowlands, 2007, p. 62).

The Strategy Inventory for Language Learning (SILL) (Oxford, 1990) is probably the best-known tool for eliciting learning aims and strategies in all four language skills, and covers a number of domains in addition to cognitive and metacognitive strategies. These are memory, compensation, effective and social strategies, all of which are beyond the scope of this study. Meanwhile, metacognitive strategies comprise three groups of strategies and eleven items (Oxford, 1990, pp. 18–21). SILL is probably the most widely used tool among language researchers (Amerstorfer, 2018; Tran, 2021; White, Schramm, & Chamot, 2007). An early study among Afrikaans ESL learners in South Africa by Dreyer and Oxford (1996) found that the use of learning strategies based on SILL accounted for some 45% of the total variance in TOEFL scores, with the largest contribution made by metacognitive strategies. Whilst there have been criticisms of SILL in terms of its adaptability to different cultures, its categorisation of strategies and its psychometric properties, as mentioned by Amerstofer (2018) and Papadopoulou et al. (2004), it remains a useful tool for self-assessment (Amerstofer, 2018). However, there are some potential disadvantages for adapting it only for pronunciation. Firstly, there are

relatively few statements which can be applied to pronunciation and they are expressed in general terms rather than linked to a specific task, which would not be suitable for monitoring strategy use. Secondly, the Likert-type scale used in SILL runs from "1. Never or almost never true of me" to "5. Always or almost always true of me"; the use of the word 'true' may suggest that this is a personal trait that may be difficult to change rather than a learning behaviour which may be more variable and adaptable (Mizumoto & Takeuchi, 2010). The 'true of me' labels in a Likert-type scale are more usually found in research on self-efficacy (e.g., Croasmun & Ostrom, 2011). Moreover, the scoring sheet is not needed unless it is used to self-assess use of learning strategies in general.

Within the overall taxonomy of learning strategies, three groups were identified as metacognitive strategies: centring learning, such as paying attention and noticing; arranging and planning learning, for example, setting clear goals and looking for opportunities to work towards achieving them; and evaluating learning by reflecting on progress (Oxford, 1990). The few cognitive strategies relevant to learning pronunciation related to practising sounds, oral repetition of new words, and trying to emulate native English speaker pronunciation.

Other researchers have focused specifically on pronunciation learning strategies (Derwing & Rossiter, 2002; Eckstein, 2007; Osburne, 2003; Pawlak, 2010; Pawlak & Szyszka, 2018; Peterson, 2000). Peterson (2000) conducted an exploratory qualitative study to investigate and classify pronunciation learning strategies. Study participants were 11 adult learners of Spanish who were native English speakers; levels in Spanish ranged from beginner through intermediate to advanced. Six students kept a diary recording every strategy they were using, or had previously used, in learning Spanish pronunciation. Diary data were analysed and strategies identified were added to others found in reviewing the literature. Three students, one from each level, were interviewed about their use of pronunciation learning strategies; the list compiled from the diaries was

used as prompts for clarification or encouraging further thought. Strategies were then classified under 12 headings based on Oxford's categorisation, with smaller tactics combined into a strategy. The 12 headings were:

> representing sounds in memory, practicing naturalistically, formally practicing with sounds, analyzing the sound system, using proximal articulations, finding out about TL [target language] pronunciation, setting goals and objectives, planning for a language task, self-evaluating, using humor to lower anxiety, asking for help, and cooperating with peers (Peterson, 2000, p. 11).

The following strategies were categorised as metacognitive: finding out about TL pronunciation, setting goals and objectives, planning for a language task, and self-evaluating. These were similar to Oxford's categorisation, with the addition of finding out about TL pronunciation. Selecting specific sounds as a learning goal, and recording and listening to oneself as a method of self-evaluation were also added.

Derwing and Rossiter (2002) chose to investigate the use of coping strategies rather than learning strategies among 100 adult immigrants in Canada who were attending full-time ESL classes. Nineteen different L1s were represented and English proficiency levels ranged from low to high intermediate. Individual structured interviews revealed that more than half the participants considered that pronunciation difficulties were the cause of breakdown in communication outside the classroom, highlighting the importance of clear pronunciation to learners. Seven types of coping strategy were identified: "self-repetition; paraphrase; increase in volume; write, spell; slow rate; clear speech; and an "other" category" (Derwing & Rossiter, 2002, p. 159). The authors noted that these coping strategies were not always used appropriately. Many students were unable to pinpoint their pronunciation difficulties and it was therefore proposed that teachers should

find ways of increasing students' awareness of the strategies they used. It was also proposed that ESL teachers should help students to developing coping strategies that would most effectively support their communication in natural language environments. The principles of raising awareness of strategy use and promoting the development of effective strategies are equally important and applicable to the support that learning strategies could provide to the improvement of pronunciation.

Further strategies were proposed by Vitanova and Miller (2002), who adopted an action research approach with ESL university students in the United States who were taking pronunciation classes. They collected students' reflections on their experiences of learning pronunciation, in order to stimulate their awareness of, and engagement in, the improvement in their pronunciation. Data analysis revealed that attentive listening to native speakers and self-correction were two of the strategies that learners found helpful. The authors asserted the importance of teaching students how to learn pronunciation, to enable them to identify where they needed to improve and know the strategies they could use to make the improvement.

The methods employed by Peterson (2000), Derwing & Rossiter (2002) and Vitanova and Miller (2002) offer options for learners to explore their use of strategies some time after speaking. All the authors noted that participants needed supporting or prompting during the process. Osburne (2003) employed a different method which made participants' reflections happen almost in real-time. Higher level ESOL learners were asked to talk autobiographically for 10 minutes about their language learning, then listen to the recording and choose a sentence to repeat with the aim of improving their pronunciation. They were then asked what they had done to try to improve their pronunciation. Data analysis led to the categorisation of eight learning strategies which included focusing on prosodic structure as well as individual words or sounds. Although Eckstein (2007)

criticised the lack of thinking time, this method provided participants with an opportunity to monitor their use of strategies immediately after production. This corresponds as closely as possible to monitoring during a speaking activity, which is the second stage in the metacognitive reflection cycle.

Eckstein examined whether there was a relationship between adult ESL learners' use, specifically frequency of use, of pronunciation learning strategies and scores on an achievement test. A questionnaire was administered to 183 international students attending classes at high-intermediate intermediate, and low-intermediate levels in a university language teaching centre (Eckstein, 2007). There were six possible response categories in the frequency scale, which ranged from 'never' to 'several times a day', including options such as 'about once a week' and 'less than once a month'. Whilst these options might be suitable for students who are attending dedicated language classes, they would be less meaningful for students in an ESL environment who were attending classes once or twice a week as one element of a degree which largely focused on other topics. Particular items which were found to predict test results were adjusting facial muscles while speaking "like opening my mouth wide", noticing other people's mistakes, and asking for help (Eckstein, 2007, pp. 100-101). There was an assumption in the rubric used to score the achievement that the ultimate aim was to sound like a native English speaker, which is not appropriate for speakers of English as a lingua franca.

Interest in pronunciation learning strategies persists; Pawlak, in particular, developed a Pronunciation Learning Strategies Survey consisting of 60 items to which participants responded on a 5-point Likert scale stating how well each item applied to them, from not at all to completely (Pawlak, 2010). Pawlak and Szyszka (2018) have provided an overview of research into pronunciation learning strategies, including applied research aimed at investigating which taught strategies benefit specific components of

pronunciation. The present review of literature regarding language learning strategies has highlighted some potentially useful verbs and items for a self-assessment pronunciation review checklist such as 'noticing' and 'practising', the former a cognitive strategy and the latter a metacognitive one (Oxford, 1990) and 'I look up the pronunciation of new words in a dictionary' (Pawlak, 2010).

The distinction between metacognitive strategies and cognitive strategies is useful because the knowledge of cognition component of metacognition includes knowledge of cognitive strategies, while the regulation of cognition component of metacognition includes monitoring and control of cognitive strategies. Thus, a checklist for self-assessment of pronunciation which is underpinned by a metacognitive approach to self-assessment will need to include both cognitive and metacognitive strategies. The main metacognitive strategies that have emerged from reviewing language learning strategies are: planning for language tasks, seeking opportunities to practise, paying attention, monitoring errors and evaluating and reflecting on own progress (Eckstein, 2007; Oxford, 1990; Pawlak, 2010; Peterson, 2000).

The Metacognitive Awareness Listening Questionnaire (MALQ) developed by Vandergrift et al. (2006) classified metacognitive strategies slightly differently, to include a category of mental translation which contained strategies deemed unhelpful to listening. Four other categories were defined as: problem-solving (inferencing and monitoring inferences, for example detecting errors); planning and evaluation; person knowledge (knowledge of own strengths and weaknesses in listening, as well as anxieties); and directed attention (Vandergrift et al., 2006, 450-451). The questionnaire consisted of 21 items which covered all five categories. Items were organised into strategies used before, during and after listening to a spoken text. A 6-point Likert scale from 'strongly agree' to 'strongly disagree' was used to collect participant responses. In total, 966 participants

from different language and educational backgrounds, and with varying proficiency levels of English from beginner to intermediate-advanced, completed a draft version of the questionnaire. Factor analysis was employed to produce and then confirm a shorter version of the questionnaire, which was completed by a further 512 participants, who were studying either French in Canada or English in Iran. The use of a disagree-agree frequency scale was more appropriate than SILL's 'true of me' scale to the completion of a specific listening task because it avoided the implication of being a more permanent personality trait. Correlation of scores on a listening comprehension test with the questionnaire data indicated that the some 13% of variance in test performance was attributable to metacognition. Whilst none of the items could directly be used in a checklist for self-assessment of pronunciation, it was interesting to note the use of the present tense 'As I listen' in a group of items covering problem-solving, mental translation and evaluation series; this would help to focus participants' minds on the recently completed comprehension task.

The (MALQ) has since been adapted for speaking by Sulistyowati et al. (2022), who selected three categories of metacognition to form the basis for their questionnaire. The three items in the questionnaire which related directly to pronunciation were the following items:

> 13. As I speak, I quickly adjust my pronunciation if I realize that it is incorrect.
>
> 14. After speaking, I evaluate how I speak and try to practice [sic] differently in the future
>
> 19. To increase my pronunciation, I imitate spoken material
>
> (Sulistyowati et al., 2022, pp. 213-214).

This section has revealed a pool of items suitable for use in investigating the use of pronunciation learning strategies and metacognitive awareness in general. However, the instruments are mostly suited to further research into the learning strategies or for research

that leads to particular considerations in teaching pronunciation or one of the four skills of speaking, listening, reading and writing. They are less well suited to raising awareness as a first step towards self-assessment of pronunciation for individuals undertaking a degree in aircraft engineering technology who already study in an ESL environment and for many of whom English is a functional necessity rather than a subject of academic interest. The following section therefore considers checklists, which can include rubrics and scripts among a variety of formats; checklist "is the term used in classroom assessment contexts" (Andrade, 2019, p. 4).

2.7 Checklists

Checklists can support learning in a number of ways. According to Rowlands (2007), checklists make it clear what is expected, such as the sequence of steps to be carried out in a task, or points to be considered while undertaking the task; they do not, however, describe the standards to be achieved. Isbell and Sakai (2022) argue they can be used "for students to rate their own abilities to target features" or to provide an opportunity to reflect on progress and areas for improvement (p. 207). They can be used following instruction and completion of a task as a reminder to students about what they need to do with similar tasks in the future (Ellis et al., 2014). Rowlands (2007) mentions that this can include reminding students of how to tackle a task in language learning, in particular helping them to "internalize new processes" and "scaffold development of independent control of such processes" (p. 66).

Checklists can also act as cost-effective diagnostic tools which can be used to plan subsequent learning (Harder et al., 2015; Schaetzel & Low, 2009). Alternatively, they can be used to evaluate a task after it has been completed, such as a self-evaluation sheet for completion after listening to a recording (Celce-Murcia et al., 1996, p. 405). Indeed, they can meet more than one of these objectives, as in self-assessment checklists that

cover the preparation, execution and evaluation stages of a task, such as reading comprehension (Yoshimura, 2009) or argumentative writing (Nimehchisalem et al., 2014).

There is considerable flexibility and variety in the formats found in checklists, as illustrated by the following examples. Checklists can be stand-alone, such as a list of questions beginning 'Did I…?' or 'Do I..?' to assist L2 teachers to evaluate essential elements of their teaching, using a frequency response of 'in general-this week-today' in which one or all three can be ticked (Dubiner, 2018, p.27). Dubiner's checklist enables L2 teachers to give feedback to themselves by honestly answering the questions and considering the implications, if they have the motivation and opportunity to do so. Other checklists can be intended for use with an analytic rubric, for example for self-assessment of oral presentations (Cañete-Gutiérrez & Inostroza-Araos, 2022). The rubric contains descriptors and levels ranging from 'excellent' to absent', while the checklist asks questions with 'yes/no/what do I need to improve' range of response (pp. 77-79). Others include guidance rather than an analytic rubric to give more detail about each of the checklist items and to encourage students to take more responsibility for their learning by suggesting ways they can achieve each checklist item, for instance the extended guide in Nimehchisalem et al. (2014, pp. 75-80). Some use statements and invite responses in different ways, such as a choice of yes/no or yes/no/not sure, as in the diagnostic writing checklist by Mazloomi and Khabiri (2016) which was aimed at enabling learners to better recognise their strengths and weaknesses over an eight-week period.

The issues involved in decisions about the design criteria of a checklist are important to achieving its purpose. They are likely to be influenced by the classroom context: the constraints of curriculum and timetable, the level of learners, how long they have been learning and using the language, whether they are studying in an ESL environment, and whether or not they are English or Linguistics majors. Additionally, design choices are

likely to reflect the preferred pedagogical approach of the researchers, teachers, or institutions involved. Although these decisions may be made on the basis of teaching experience and knowledge of learners, they need to be capable of justification.

One example of a checklist that brought together the elements of the metacognitive cycle of self-assessment, as well as students' awareness of what they should do and why they should do it, was Nimehchisalem et al. (2014). This checklist aimed to raise awareness of what students needed to before they attempted a writing task, what they needed to do while executing the tasks, and what to check after completion of the task. However, the principle would need adaptation for use with pronunciation, because it is not possible in many cases to think about a spoken task to the same extent as a written task during the execution phase, simply because there is less time available to think due to the intensity of focus on producing the next idea or sentence in real time. The provision of guidance supports students to be more independent in their learning by proposing methods of achieving the criteria in the checklist outside of the classroom.

No checklist, and indeed, no study was found which could quickly and easily be adapted to match the needs of the BAET students at the research site. Selected studies in the following section illustrate the gap in the literature, a gap which is related to the selection of the ICAO descriptors of pronunciation because meeting the ICAO standard is what students work towards and introduction of a different set of criteria may adversely affect their motivation.

2.8 Review of selected studies

The selected studies examine a range of approaches to self-assessment of pronunciation segmentals and non-segmentals, in some instances as a subset of speaking, including the different processes employed. It is noticeable that although there has been an increasing

focus on whether self-assessment leads to improvement, concerns persist regarding the reliability and accuracy of self-assessment (e.g., Ross, 2006; Yan & Carless, 2022).

Dlaska and Krekeler (2008) investigated the reliability of self-assessment of pronunciation skills and the reasons for differences between students' and raters' assessments. Forty-six advanced learners of German who had appropriate knowledge of phonetics were asked to identify errors in their pronunciation of specific sounds in a list of 43 words. They read the list, then listened to a recording of the words by a native speaker, before listening to and repeating each word. The learners then listened to their recorded words, compared their pronunciation of a specific sound with the native speaker's and said whether or not their pronunciation was the same as the native speaker's. The raters then compared the recordings. Although the overall level of students' and raters' agreement was 85%, and students' assessments were 89% accurate, difficulties were mainly attributed to several particular sounds and L1 interference. The authors concluded that similar self-assessments of pronunciation could be useful. However, their method could not easily be applied to Malaysian ESL learners at the research site for three main reasons. Firstly, a focus on segmentals is less appropriate than a focus on prosodic features for people who have been learning and using a recognised world English for a number of years. Secondly, similarity to native speaker pronunciation is less important than intelligibility or comprehensibility in the ICAO high-stakes test. Finally, technical students are unlikely to have, or be given the opportunity to gain, phonetic knowledge.

Another study by Brannen et al. (2022) examined the impact of self-assessment activities on the awareness and accuracy of ESL learners' pronunciation of segmentals. The authors compared experimental and control groups of ESL learners, 29 in total, with a variety of L1 backgrounds, who were enrolled on pronunciation courses equivalent to CEFR levels B1, B2, and C1. Pre- and post-tests were administered; all students recorded

themselves reading 10 short phrases and 25 sentences. Test group students then undertook a self-assessment once week for 10 weeks, recording themselves reading five sentences or a short text according to their level, before listening to a native speaker recording of the same words or text. They next answered questions such as 'Did you pronounce the "th" sounds in the same way as the model?' (Brannen et al., 2022, p. 11). Students in the two higher level groups received written feedback on their self-assessment responses. Experienced native speaker English pronunciation teachers then listened to all participants' recordings of phrases and counted the errors. Some students (N=14) were interviewed after the course about the impact of self-assessments on their awareness and attitudes, as well as their views of their pronunciation. No significant differences were found between pre- and post-tests within each group nor between control and test groups. However, learners considered that the self-assessments helped them become more aware of their errors as well as helping them to improve their English pronunciation (p. 8). This study, like that of Dlaska and Krekeler (2008) essentially involved a comparison of a recording of a student's reading with a recording of a native speaker model. The method is highly suitable for raising awareness and improving aural perception but reading text is a different skill from spontaneous speech production in an authentic task.

A third study investigating segmentals (Gralińska-Brawata, 2022) explored how first-year English philology students (N=17) taking a pronunciation course in a Polish university assessed their progress. A pre- and post-test design employed a list of phrases from a pronunciation textbook, and students responded to two questionnaires. One questionnaire asked about experience of phonetic training and language learning more generally, and the other asked students to note the sounds students thought they had improved, and to state what worked best in improving pronunciation. Students recorded the pre- and post-tests, identified errors in the pre-test then compared both recordings to identify improvements. Additionally, students' and teachers' assessments were compared in

terms of the numbers of errors and improvements detected. Participants reported increased awareness and more than half indicated that 'listen and repeat' activities were a good way to improve pronunciation. The author concluded that "self-assessment may be viewed as a valuable pedagogical tool for helping second language learners to raise their awareness of certain specific pronunciation difficulties and improve their pronunciation skills" (Gralińska-Brawata, 2022, p. 259). However, as in the study by Brannen et al. (2022), the study does not relate to authentic speaking tasks and involves a specialised group of language learners.

Cojo Guatame (2019) also addressed segmentals but adopted an action research approach and a particular task-based language teaching intervention for self-assessment of activities such as short conversations and presentations about familiar real-life topics. Ten students at a Colombian university with CEFR English levels of A1 and A2 aged 17-21 analysed recordings of pre- and post-tests using the IPA and completed questionnaires about their views of self-assessment. They used a checklist to guide peer- and self-assessment in three of nine weekly four-hour lessons between the tests. The checklist covered grammar, vocabulary and fluency as well as pronunciation; students indicated yes or no to assess themselves and their peers. The pronunciation section included:

> Is understandable and can be followed by the other speaker
>
> Uses some variety in the voice (volume, rate, pitch, and rhythm)
>
> Uses stress and intonation
>
> Emphasizes keywords (nouns, verbs, adjectives, adverbs)
>
> Pronounces words correctly
>
> Spelling is accurate and understandable [for phonetic transcription] (Cojo Guatame, 2019, p. 51)

Teachers completed journals after each session. Lesson activities aimed at developing self-assessment skills as well as covering the English course content, while teachers provided ongoing feedback on students' segmental pronunciation difficulties. The study concluded that self-assessment helped to improve segmental pronunciation but that more time was needed. The use of IPA transcription with A1 and A2 level students would not be encouraged by most CLT teachers, nor would such students be encouraged to spend time accessing web resources to learn IPA rather than perhaps listening to authentic native speakers. The expectation that students at this level will use stress and intonation is not based on the CEFR levels shown in Figure 2-2 and could discourage, rather than motivate, some students. The selection of standards and methods appropriate to the purposes of the self-assessment is highly important.

Another action research study was conducted by Jankowska and Zielińska (2015) who developed a self-assessment instrument as a pedagogical tool for 46 student teachers of English in their third year at a teacher training college in Poland. The students had previously undertaken pronunciation practice classes, as well as courses involving assessment, including aspects of self-assessment, and learning strategies. However, the majority of students were initially unable to assess their own short speeches, even with the help of forms they had designed themselves in groups. The authors therefore developed a checklist for advanced level speaking skills. They drew on C1 speaking descriptors from the CEFR (the expected level of achievement), specific items from the college examination requirements, and their experience as teachers. The checklist covered content and vocabulary in addition to pronunciation. Pronunciation items were: "[I managed to]: 7. Speak fluently, 8. Use appropriate intonation, 9. Pronounce "th" correctly, 10. Pronounce final voiced consonants correctly, 11. Pronounce vowels correctly, 12. Use correct stress in words" (Jankowska & Zielińska, 2015, p. 260).

Students graded themselves from 1 to 5, with 5 representing 'very good', and had space to provide examples or comments. The checklist was evaluated and students' use of learning strategies investigated through interviews with a sample of 15 students. A minority of students were not in favour of self-assessment, while the majority found it difficult, leading the authors to conclude that teachers needed to give more time and guidance to those students who needed it before they were ready to use self-assessment. They added that students should be actively involved in the development of a self-assessment checklist, and planned to develop a revised version. Furthermore, they indicated that the accuracy of self-assessments and the extent of agreement between teachers' and students' assessments were areas for further work. Whilst this study used CEFR descriptors at an appropriate level for the participants, it nevertheless illustrates that developing and implementing a pedagogical tool is not always straightforward. Self-assessment may need to be introduced in small steps, with care taken to address possible issues of resistance and inadequate understanding.

2.9 Summary

The evolution of teaching approaches and comparison of high-stakes rating scales' descriptors has shown that although some aspects of pronunciation may contribute more to ease of understanding than others, as yet there is no consensus on which these are. It was proposed that segmentals were vital, but word stress was perhaps not important, in a lingua franca core (Jenkins, 2000). In contrast, it was asserted that word stress and intonation were also important (Isaacs & Trofimovich, 2012). To a varying extent, the CEFR and the high-stakes English test rating scales examined indicated that segmentals, non-segmentals and intelligibility (rather than nativeness) were all important. None of the scales were totally suited to ease of use in the classroom context, although in a self-assessment context, teachers should be able to make students aware of the standards they will be expected to achieve.

Self-assessment was defined as a formative and developmental process which enables students to give feedback for themselves from themselves. In order to do this, they may require standards or criteria to guide their reflections on both their achievement and how they learn, although Andrade (2019) acknowledged that some forms of formative, developmental self-assessment may not involve standards. In any event, learners cannot be expected to suddenly become confident and capable self-assessors; they need support, which may need to be provided gradually, step by step (Little, 2005). Preparation for self-assessment should take account of students' and teachers' views of teaching and learning, in addition to student motivation (Schmidt & Wehmeyer, 2016). Despite the challenges, self-assessment of pronunciation can be integrated with curriculum objectives through opportunities for speaking activities (Isbell & Sakai, 2022). Self-assessment of pronunciation remains important in the context of a growing number of computerised assessment programs based on speech recognition software; ultimately, it is the learner who has to decide what to do in order to improve.

Metacognitive processes were considered in two ways in relation to language learning. The first was the widely recognised understanding of the processes of planning, monitoring and evaluation, and the associated reflection at each stage on how to achieve learning goals. The second concerned the choice and application of learning strategies to achieve those goals. Whilst it was acknowledged that research into language learning strategies involved too many different strategies for use in a pedagogical tool for classroom use, sufficient knowledge of cognitive strategies was essential for learners to be able to monitor their use of strategies and evaluate how well the chosen strategies worked. Relatively few of the strategies identified by language learning strategy researchers applied directly to pronunciation, although three specific items were identified in the metacognitive awareness questionnaire adapted for speaking by

Sulistyowati et al. (2022). These were: self-correction of pronunciation while speaking; attempts to correct mistakes realised after speaking; and imitation of spoken English.

The review of selected checklists and studies did, however, reveal several directly relevant items, in some cases based on diagnostic analysis of learners' needs, in others on teaching materials or teachers' knowledge and experience. Several of these referred to word stress, rhythm, and intonation, although others were more specific or more general according to their particular objectives. It was clear that the majority of studies reviewed did not involve authentic speaking activities, and that there appeared to be a reliance on methods such as listening to recordings or transcriptions, which may not apply in all teaching contexts, nor suit all learner preferences. Comparison of pre- and post-tests did not always indicate significant improvement in pronunciation, nor increased ability to self-assess. Self-assessment may need to be introduced in small steps, with care taken to address possible issues of resistance and inadequate understanding. Whilst there is no single best way to design a pedagogical tool for self-assessment of pronunciation, there are clear criteria to be observed. The purposes must be clear and must be appropriate to the target audience, and the design must achieve the purpose. Design details of a self-assessment tool such as number and length of items, and choice of Likert-type scale, must enable achievement of the purpose, while administration procedures must be appropriate to the research context. A self-assessment checklist cannot stand alone but must be embedded into the context where it will be used.

3 CHAPTER 3 METHODOLOGY

3.1 Introduction

This chapter sets out and justifies the research design, methodology and methods, data collection instruments, procedures used and data analysis methods. The participants are described and an explanation is given of how ethical considerations were handled. This chapter is divided into sections which present the overall research design and its theoretical underpinnings (3.2), the research process (3.3), the data collection instruments (3.4), participants and setting (3.5), data analysis (3.6), ethical considerations (3.7) and summary (3.8).

3.2 Research design

This section explains and justifies the approach taken in this developmental study which addressed the need for an instructional tool to enable English as a Second Language (ESL) students at a technical university to self-assess their pronunciation skills. According to Creswell (2014) the overall research strategy should not only be suited to the research problem but should also take into account "the philosophical assumptions the researcher brings to the study; procedures of inquiry (called research designs); and specific research methods of data collection, analysis, and interpretation" (p. 3).

The research design for the present study was chosen to meet the research objectives which were: (1) to design a pronunciation self-assessment checklist, (2) to calibrate the pronunciation self-assessment checklist based on the feedback from teachers and students, and (3) to evaluate the usefulness of the pronunciation self-assessment checklist. The associated research questions were framed as: (1) What criteria should be used to design the pronunciation self-assessment checklist? (2) What are students' and teachers' reactions to the pronunciation self-assessment checklist? (3) How do students and

teachers evaluate the usefulness in terms of the impact, practicality, reliability and validity of the pronunciation self-assessment checklist?

In order to address the research questions, a development design approach was adopted. Following Richey and Klein (2009, p. xvi), the term 'development research' is used in preference to 'developmental research' in order to avoid confusion with research in other academic fields such as psychology. Development research has been associated with a variety of activities in the field of teaching and learning (Van der Akker, 1999, pp. 4-5). At one level, it seeks to "test theory and validate practice" (Richey & Klein, 2007, p. 1), while at the practical level it has been defined as "the systematic study of designing, developing and evaluating instructional programs, processes and products" that must meet certain criteria (Seels & Richey, 1994, p. 127). The criteria for this present study are defined in terms of impact, practicality, reliability and validity as defined by Bachman and Palmer (1996) (see section 3.4.6).

Development research focuses on the connection between theory, data, practice and knowledge in order to create knowledge which is useful in practice, such as an instructional tool. It is suitable for creating "context-specific knowledge that serves a problem solving function" (Richey & Klein, 2005, p. 24) and as such can meet teachers' and students' needs for an instructional product like a checklist while ensuring the product is based on data gathered from practice.

Typical phases in development research can include some or all of the following: design and development, evaluation, validation, model development, model use and model evaluation (Richey & Klein, 2005). The process may be iterative, with repetitions of some or all of the phases (Fulcher & Davidson, 2007, p. 84). The actual phases and iterations used will depend on the process or product being developed and on the steps required to achieve the objectives of the research. According to Richey and Klein (2005,

p. 26), development studies focused on a context-specific instructional tool may have three or four phases, while others may have more (Kirschner et al., p. 89). The purpose and content of each phase is adapted to the particular study.

The flexibility and adaptability of the three-phase approach is illustrated by the two contrasting examples which follow. Nakatsuhara et al. (2018) conducted a three-phase development study to create a checklist for teachers to use in assessing and giving feedback on students' interactional competence in English. The first phase involved eliciting *Cambridge English: First* examiners' comments on videos of interactional competence. The second phase entailed drafting a checklist which was subjected to expert review, while the third phase consisted of piloting the checklist with four language teachers, followed by further expert review (p. 10). Another study which adopted three phases was carried out by Tracey and Richey (2006), who aimed to develop a model of instructional design which included "the theory and practice of multiple intelligences" (p. 369); they reviewed the theory underpinning both aspects, then constructed a first model and, finally, validated the model using a Delphi approach with four instructional design experts (p, 377).

In order to meet the research objectives of the present study, there were three clear phases which corresponded to the objectives. The first phase covered design of the checklist, which drew on multiple sources from the literature, including Fulcher and Davidson (2007, 2012) and Fulcher and Harding (2022). The second phase involved calibration of the checklist based on end users' reactions. The third phase, the evaluation phase, covered the examination of usefulness of the checklist based on teachers' and Bachelor of Aircraft Engineering Technology (BAET) students' responses to an evaluation questionnaire. Four of the six domains of usefulness proposed by Bachman and Palmer (1996) were examined, namely impact, practicality, reliability and validity. The remaining two, interactiveness and authenticity, were omitted because they refer to the nature of the test

task itself, whereas the checklist in this study could be applied to a variety of tasks (Giraldo, 2019, p. 126).

This present study was situated in the field of language assessment in a teaching context as well as development research and therefore the research design also draws on research methodologies used in language assessment. Whereas it was at first traditional to use quantitative methods and then to adopt qualitative methods (Lazaraton & Taylor, 2017), researchers seeking to answer questions pertaining to new types of language assessment, for example "assessing learners' language development in classrooms", have increasingly used a combination of the two (Tsushima, 2015, p. 105). The combination of qualitative and quantitative approaches can be used to preserve the strengths and reduce the weaknesses in both approaches. A mixed methods approach was chosen to be used in this present study to take advantage of the strength of qualitative research which provides "rich data that is needed to understand even subtle meanings in the phenomenon under focus" (Dörnyei, 2007, p. 127). Qualitative methods typically involve smaller samples than quantitative methods and can therefore lead to findings which are not representative of a wider sample. On the other hand, quantitative methods allow the researcher to gather data more quickly from a larger sample size and thus arrive at a more representative picture of differences in interpretation or perception. However, quantitative methods alone can fail to capture important differences and may not take full advantage of experts' and practitioners' knowledge and judgement in development studies. Furthermore, a mixed methods approach is appropriate for exploring reactions at the individual level as well as at the level of a wider group (Morse & Niehaus, 2016, p. 13), as in this present study.

In the distinct phases of this present development study, the design of the checklist required qualitative methods in order to gather detailed views on the content and clarity from experts, students and teachers. In the calibration phase, qualitative methods were

employed to explore students' and teachers' reactions to the checklist, especially in terms of the clarity and feasibility of the self-assessment tool, after they had had an opportunity to try it out. Information regarding the usefulness of the checklist in terms of an appropriate combination of its validity, reliability, impact, and practicality (Bachman & Palmer, 1996) required a quantitative survey. Since neither quantitative nor qualitative methods alone would meet the research objectives, a mixed methods approach was therefore chosen, shown in Table 3-1.

Table 3-1 Research design overview

RESEARCH OBJECTIVE	RESEARCH QUESTION	ACTIVITY	METHOD
DESIGN PHASE			
1. To design a pronunciation self-assessment checklist.	1. What criteria should be used to design the checklist?	Produce Checklist v1.0	Qualitative method - Literature review.
		Expert validation and data collection from students and teachers	
		Expert validation - (pronunciation and assessment) 2 experts	Qualitative method - E-mail and/or online discussion
		Check clarity or understanding of checklist and extended guide with end users (BAET students and Aviation English lecturers) - Produce Checklist v2.0	Qualitative method - Structured group interview with 8 students - Semi-structured interviews with 2 teachers
CALIBRATION PHASE			
2. To calibrate the checklist based on the feedback from students and teachers.	2. What are the students' and teachers' reactions to the checklist?	Administer/try out checklist	Qualitative method - Online checklist trialled by 50 students and 3 teachers (including researcher) - Semi-structured interviews with 10 students and 2 teachers
EVALUATION PHASE			
3. To evaluate the usefulness of the checklist	3. How do students and teachers evaluate the usefulness in terms of impact, practicality, reliability and validity of the checklist?	Expert validation of checklist 3 experts (Assessment/evaluation) Expert validation of evaluation questionnaire 2 experts (Assessment/evaluation) Administer evaluation questionnaire	Qualitative method - E-mail and/or online discussion Quantitative method - Online questionnaire responses of 50 students and 2 teachers

3.2.1 Design phase

In the design phase, Checklist v1.0 containing the pronunciation self-assessment checklist and extended guide (Appendix A) was designed based on the review of available literature on pronunciation, self-assessment, metacognition and checklists. The checklist and extended guide were then sent for expert validation to two experts in assessment with particular knowledge and experience of the ICAO rating scale (section 3.5.1). The checklist and extended guide were emailed to a further seven international experts in assessment. All the experts were specifically asked to comment on the clarity and completeness of the checklist and extended guide, paying particular attention to the construct of pronunciation.

The expert validation was conducted simultaneously with data collection from students and teachers, in agreement with a rapid prototyping approach, which involves end users throughout the process (Jones & Richey, 2000; Nixon & Lee, 2000). The checklist and extended guide were thoroughly discussed, item by item and in general, with two Aviation English teachers (section 3.5.2) using a semi-structured interview schedule (Appendix B). They were also examined by a group of eight BAET students who took part in a structured group interview. The selection of students who participated is described in section 3.5.3 and the questions used to structure the interview are given in Appendix C. After the detailed discussions with both teachers and students, their comments were analysed and considered along with the comments from the experts who were familiar with the ICAO scale.

Details of the changes made to the checklist and guidance are described in Section 4.1, along with reasons for acceptance and rejection of suggestions. The amended checklist which incorporated guidance from the extended guide (Checklist v2.0) is reproduced in Appendix D.

3.2.2 Calibration phase

The purpose of the calibration phase was to discover whether the checklist would work in practice, as measured by the reactions of students and teachers. This phase was divided into two sub-phases. The purposes of the first sub-phase were to ensure that students were given the opportunity to thoroughly understand the criteria which underpinned the whole self-assessment process, to establish whether the checklist and the guidance it contained were sufficiently clear, and to investigate the feasibility of using the self-assessment checklist in the classroom situation. Checklist v2.0 was administered with three classes of BAET students (section 3.5.2) with the assistance of two teachers in addition to the researcher. The second sub-phase involved semi-structured interviews with the two teachers and a sample of 10 students from those who had trialled the Checklist v.2.0 (section 3.5.2) to further explore any issues of clarity and feasibility. The interview guides for these semi-structured interviews are shown in Appendices E and F respectively and discussed in Section 3.4.5.

3.2.3 Evaluation phase

In the third and final phase of this development research, the usefulness of the checklist was evaluated by administering questionnaires to students and teachers to evaluate the usefulness in terms of impact, practicality, reliability and validity of the self-assessment checklist. First, expert validation of the questionnaires was conducted by two experts in the area (section 3.5.3) who received the evaluation questionnaires and returned their comments by email. The evaluation questionnaires were administered to the teachers and students who had taken part in the calibration phase. Additionally, three further international assessment experts carried out validation of Checklist v2.0.

Further details of the phases of the study are set out in sections 3.3.1, 3.3.2 and 3.3.3.

3.2.4 Philosophical assumptions

Following Creswell (2014), the philosophical assumptions underpinning this study are briefly stated here. There are many worldviews, or paradigms, that can be adopted to suit research studies in the field of education, and "the labels for different approaches are not used in standard ways" (Hammersley, 2012, p. 3). According to Visscher-Voerman et al. (1999), the development paradigm associated with "interactive and repeated tryout and revision" is that of pragmatism (p. 17). Furthermore, Creswell (2014) states that the worldview associated with mixed methods research is pragmatism, because it avoids conflicting views of the nature of truth or knowledge and is rooted in applied research. Creswell notes that pragmatism is concerned with the "consequences of actions" taken to solve "real-world oriented" problems (2014, p. 13). Pragmatism focuses on "what works" and is not limited to any single theoretical stance (pp. 10-11), which means that researchers can select the methods that best suit the problem they wish to address, which is focused on 'what' and 'how' questions rather than 'why'. This present study is practice-based and focuses on the design and perceived usefulness of a pronunciation self-assessment checklist, thus it is situated within the worldview of pragmatism.

3.3 Research process

This section explains in more detail the steps in each of the three phases shown in Table 3-1. It also describes how every step in the whole process gathered feedback and comments on the checklist which was then amended and refined before the next step.

3.3.1 Phase 1: Development of pronunciation self-assessment checklist

In the first phase, the first version of the checklist was designed. This involved defining the purpose of the self-assessment, deciding on the form it should take (checklist or rubric), and specification of the construct of pronunciation to be used. As mentioned by Andrade and Valtcheva (2009, p. 13) and Schmeiser and Welch (2006, p. 308), the purpose and criteria have to be clearly articulated and linked. In this present study, the

purpose was to develop a pedagogical tool which would encourage students to develop their pronunciation skills along with their understanding of self-assessment and their awareness and understanding of their learning processes and actions which would help them to do that. In other words, the purpose was one of self-assessment *for* learning.

The specification of the construct and the design of the checklist were based on the review of literature (section 2.3), the pronunciation component of the ICAO language proficiency rating scale (section 2.3, Table 2-4), and experience of teaching BAET students. An extended guide was added to the checklist in order to explain and justify each item in the checklist. Explanations and illustrative video clips of the terms 'syllable' 'word stress', 'rhythm' and 'intonation' were included in the extended guide following expert appraisal. Full details of the checklist design are given in section 3.4.1, while the extended guide is covered in section 3.4.2. The original version of the checklist with its extended guide is shown in Appendix A.

Expert validation and contributions from end users were sought in parallel. The importance of considering multiple stakeholder perspectives in instructional product development is supported by Richey et al. (2004, p. 1115). It is further supported for prototyping item types in language test development (Nissan & Schedl, 2012, p. 282).

It is also asserted that including the learners' perspectives can enhance the processes of instructional design and development processes (Konings et al., 2014, p.2). The checklist and extended guide were submitted for expert validation to two subject-matter experts in the field of assessment and teaching of pronunciation with particular knowledge and experience of using the ICAO rating scale gained in Malaysia. A further seven international subject-matter experts who were authors of published academic articles or book chapters on the assessment and teaching of pronunciation were also contacted.

Feedback was sought from students and teachers as the end users of the checklist. It was essential that the meaning of the items in the checklist was clear to the end users, and that a guide to completion was provided where additional clarity was required. The layout of the checklist and the process for using it also needed to be clear. Without such clarity, end users would be unlikely to use the checklist. Feedback from two Aviation English teachers was gathered in semi-structured interviews, based on the interview guide in Appendix B which is described in section 3.4.2. Teachers had the opportunity to see the questions, checklist and extended guide before the interviews so that they knew what to expect and could spend some time reflecting upon their evaluations of the checklist and guide beforehand.

Comments on the clarity of checklist items and the extended guide were gathered from eight students in an online structured group interview. The interview guide used to run the session contained broad questions to encourage discussion as well as specific detailed questions, as shown in Appendix C. This was to address the possibility that the students would not engage in active discussion among themselves because the typical Malaysian classroom dynamic is teacher-centred, meaning students only speak when they are addressed directly by their teacher. The students involved had the opportunity to see the questions, checklist and extended guide before the interviews so that they knew what to expect and, like the teachers, could spend some time reflecting upon their evaluations of the checklist and extended guide beforehand. A teacher colleague took notes in order to add to the trustworthiness of the data collection. Demographic information was collected (gender, class and most recent English test result), and students were introduced to the study. Students were welcomed before being reminded of the purpose of the session, then asked a series of broad questions (see questions in section 3.4.3).

The data analysis processes of the group interview, the teacher interviews and the expert validation are described in section 3.6. Differences between end-user and expert

understandings and interpretations of the checklist were addressed before changes were made to Checklist v1.0. Once Checklist v2.0 was ready, the research moved to the calibration phase.

3.3.2 Phase 2: Calibration of the checklist

The purpose of the calibration phase was threefold. The first was to ensure, in accordance with Andrade and Du (2007, p. 160) and McMillan and Hearn (2008, p. 41), that students found the checklist sufficiently clear, practical, and effective for gaining the knowledge of the criteria needed for self-assessment. The second was to provide an opportunity for students and teachers to try out Checklist v2.0, which now incorporated the guidance from the extended guide into appropriate sections within the checklist. The third was to gather feedback from students and teachers on their reactions to the instruments in order to identify where further changes might be required. All three purposes needed to be fulfilled in order to answer the second research question.

The steps in this phase are summarised in Table 3-2.

Table 3-2 Steps in calibration phase

Step no.	Activity
1	Researcher sourced 15 video clips from publicly available presentations to illustrate pronunciation at levels 1 to 6 of the ICAO scale. From these, an ICAO accredited teaching colleague was asked to independently rate 10 clips before the final six were agreed jointly for use as examples of the external criteria for the self-assessment.
2	Researcher produced instructions for teachers on how to administer the checklist.
3	Researcher delivered initial briefing session to teachers and students who watched the video clips to gain a shared understanding of the ICAO levels.
4	Students received Checklist v2.0 via Google Forms on the University of Sheffield system.
5	Students prepared a speaking assignment and completed the first part of the checklist focusing on their individual learning behaviours in the preparation stage (section A of the checklist).
6	Students carried out their speaking assignment and completed the remaining sections of the checklist which were focused on their individual learning behaviours during and after the speaking activity, (sections B and C of the checklist).
7	Researcher conducted semi-structured interviews with teachers and sample of 10 students to elicit their reactions to the self-assessment checklist and process.

Six video clips were selected from a broad range of YouTube videos dealing with various aspects of the aircraft maintenance industry. The speakers had differing abilities in terms of their word stress, rhythm and intonation, as well as differing accents. Clips were specifically chosen for their potential to illustrate differences between the levels of the ICAO scale. It was relatively easy to find clips relevant to the aircraft industry, including the maintenance function, illustrating levels 4 and 5, but challenging to find clips for lower levels and impossible for level 1. The level 1 clip was therefore sourced from examples of beginner English learners. A selection of possible clips at each level was independently rated by a teaching colleague who, like the researcher, was a qualified

ICAO examiner and experienced in using the ICAO rating scale. The teacher was asked to rate the video clips and to select the clearest example of each level, after which the teacher and researcher discussed and agreed the final selection. Links to the selected clips are shown within Checklist v1.0 in Appendix A.

The researcher met with the teachers to agree on a speaking assignment and date and time in the course for trialling the Checklist v2.0 and process. It was agreed to use a group presentation for the speaking assignment. The teachers were given a brief for administering the checklist, explaining that they and the researcher needed to give students the same introduction and instructions to avoid possible misunderstandings by students of what they were expected to do. The instructions are shown in Appendix G.

The researcher, who had undergone ICAO rater training and practised assigning levels to students' work, delivered the initial briefing session to ensure that all three classes received the same information. In this briefing session, the researcher presented the video clips, starting with the clips from the guidance illustrating ICAO levels and then the clips which illustrated stress, rhythm and intonation. When the video clips illustrating ICAO levels were presented, students were asked what differences in pronunciation they perceived between the video clips. Where they had difficulty in detecting the differences, for example in word stress or intonation, the researcher highlighted the differences and explained why they affected the ICAO level. Clips were replayed when students found it difficult to detect the differences; the researcher highlighted the differences and asked if the students could now hear them. The researcher suggested that any students who still found it hard to hear the differences could listen again and shared the links to the video clips with the students.

The researcher produced Checklist v2.0 on Google Forms. The checklist was uploaded onto the university system in two parts. Two links were created to enable students to

access the sections of the checklist, one link (to section A) given for students to complete before they gave their group presentations, and the other link given after their presentations for completion of the remaining sections.

The researcher then conducted semi-structured interviews with the two teachers and a sample of 10 students (selection described in sections 3.5.1 and 3.5.2). As previously mentioned, the semi-structured interview guides for the students and teachers are shown in Appendices E and F respectively and explained in sections 3.4.5. Data from the completed checklists, teachers' assessments and interviews were analysed.

3.3.3 Phase 3: Evaluation of the checklist

This phase involved the examination of the usefulness of the checklist based on teachers' and students' responses to an evaluation of usefulness questionnaire. The main resource for the design of the questionnaire was Bachman and Palmer's approach to "developing a plan for the evaluation of usefulness" (1996, p. 133) and illustrative questions (1996, pp. 151-155 and 280-284). Where individual questions were in some cases adopted or adapted from existing questionnaires, these are described in section 3.4.6. The student and teacher versions of the questionnaire are given in Appendix H.

The first step was to contact experts in the area to find available and willing experts to conduct expert validation. Information about the study, a copy of the checklist, and a copy of the ethical approval were sent, together with information about the ICAO speaking subscale, to two international experts. After taking into account the experts' comments, evaluation questionnaires were administered via Google Forms to the students who had trialled the Checklist v2.0 and the Aviation English teachers involved in the calibration phase.

Of the three new experts in assessment who were contacted for critical review of Checklist v2.0 immediately after trialling was completed, two provided general

comments and one provided detailed comments. Moreover, one of the experts validating the questionnaire also provided comments on Checklist v2.0 as reported in Section 4.4. Comments from the expert reviewers were taken into consideration and further amendments made to Checklist v2.0.

The checklist was refined at each step of the research procedure based on the results from the preceding step, as described in Chapter 4, to ensure that the development process was rigorous, and the final version of the checklist was as robust as possible and ready for implementation.

3.4 Data collection instruments

Data was collected from multiple sources: expert validation, feedback from BAET students and feedback from experienced Aviation English teachers. Expert validation was sought from experts in the fields of assessment and teaching of pronunciation, self-assessment, and evaluation of language assessments with particular reference to the evaluation of usefulness. The instruments were semi-structured interviews based on interview guides, a structured group interview based on a schedule, and questionnaires. The design and use of the checklist itself was the focus of data collection instruments throughout the study and was subject to continual refinement.

The following sections describe the checklist, the data collection methods during the design phase, the data collection methods during the calibration phase, and the evaluation of usefulness questionnaires.

3.4.1 Self-assessment checklist

In agreement with recommended procedures for developing valid questionnaires (Brown, 2001; Dörnyei, 2003), Checklist v1.0 (Appendix A) was developed based on a literature review of relevant books and articles on self-assessment, metacognition, the teaching and assessment, including self-assessment, of pronunciation, and the construct of

pronunciation as operationalised in high-stakes English language tests. Literature on the criteria for developing a checklist as an instructional tool was also examined (see section 2.7).

The checklist was constructed with reference to two main dimensions, namely the metacognitive theory underpinning self-assessment, and the construct of pronunciation employed. The three sections (A to C) were named 'before', 'during' and 'after' a speaking activity, and were informed by the three stages of both metacognitive thinking (Muijs & Bokhove, 2020; Zimmerman & Moylan, 2009) and reflective thinking (Boud, 2013, pp. 12-15), before, during and after a task, as described in Section 1.8. From a learning point of view in this present study, in other words assessment *for* learning, the statements in these sections were about capturing the frequency of certain behaviours related to pronunciation and not about reaching learning milestones such as 'beginner' or 'expert'. In other words, the focus was on the processes of thinking and learning. Thus, the preparation and planning section was separated from the monitoring of learning, with a broader reflection on the experience as the third section.

The types of learning behaviours which featured in the checklist adapted some items from the Strategy Inventory for Language Learning (SILL), version 7.0, for Speakers of Other Languages Learning English Version 7.0 (ESL/EFL) (Oxford, 1990). Adapted items that related particularly to pronunciation were "I practice [sic] the sounds of English" (item 12 in part B), "I notice my English mistakes and use that information to help me do better" (item 31 in part D) and "I practice [sic] English with other students" (item 47 in part F) (Oxford, 1990). In the checklist, the first one was rephrased as "I practise my pronunciation" (item 1 in Section A), the second was revised thus, "I notice my pronunciation mistakes when I am speaking" (item 10 in Section B), while the verb "practise" was used again in item 7 in Section A and item 22 in Section C.

The checklist employed the present tense in Section B ('During the speaking activity'), following the approach found in SILL and in a Metacognitive Awareness Listening Questionnaire by Vandergrift et al. (2006). Vandergrift et al. (2006) combined behaviours for all three stages and used a mixture of simple present tense to indicate habit and present continuous tense to focus on what students were actually doing while listening. Examples of such statements used to identify metacognitive processes include "Before starting to listen, I have a plan in my head for how I am going to listen", "As I am listening, I predict what will happen" (Vandergrift, 2005, p. 80), and "As I listen, I focus on the main words" (Vandergrift et al., 2006, p. 462). The use of the present tense in this way was employed for the 'during' section of the checklist developed in this present study.

The second dimension was the use of the selected construct of pronunciation as the basis for self-assessment of pronunciation. The items in the checklist were pronunciation, stress, rhythm, and intonation as set out in the pronunciation subscale of ICAO descriptors (ICAO, 2004). The items were based on the checklist on the ICAO pronunciation descriptor for three main reasons. Firstly, the construct of pronunciation was clearly separate from fluency and other related constructs. Secondly, a search for suitable alternatives in published studies revealed a lack of the particular focus on pronunciation required for this present study. For example, some focused on performance rather than process (e.g., Cieślicka & Rojczyk, 2017), while some tended to concentrate on the mispronunciation of specific sounds (Brannen et al., 2022; Dlaska & Krekeler, 2008; Enxhi et al., 2012), and others overlapped with fluency (e.g., Ahn et al., 2022; Jankowska & Zielińska, 2015; De Saint Léger, 2009). Finally, the ICAO scale was chosen because it was relevant to the high-stakes performance test that many students would take. The present checklist omitted the reference in the ICAO scale to the extent to which the first language or regional variation interfered with ease of understanding because

pronunciation, word stress, rhythm and intonation were judged to take sufficient account of intelligibility, while accent or first language influence were potential distractions.

For sections A to C, a Likert scale was chosen, because "the method is simple, versatile, and reliable" and widely used in customer service surveys, market research, and academic settings (Dörnyei, 2003, p. 36). Thus, students were likely to be familiar with this type of scale. A similar 'never to always' scale was used in SILL (Oxford & Burry-Stock, 1995, p. 4), although the end points chosen for this present study were 'almost always' and 'almost never' in order to avoid obvious polar opposites. From a theoretical perspective, it has been suggested that avoiding polar opposites leads to a wider range of responses (Wyatt & Meyers, 1987, p. 33), while from a practical perspective, it is highly unlikely that a student will 'always' or 'never' perform a particular action when planning, monitoring or evaluating a speaking activity. Although there is a risk that respondents select the mid-point on a Likert scale by default (Douven, 2018, p. 1203), in this case the midpoint was 'sometimes' because the scale concerned frequency of individual actions rather than opinion. The purpose of the frequency scale was to encourage students to think about how often they used particular learning strategies and to become more aware of this on an individual level. A scale from 1 to 6 could have caused confusion with the levels 1 to 6 on the ICAO scale, and was therefore avoided. Using 4 boxes would have created a forced choice for students who were unsure about how often they used particular learning strategies, which would have worked against the purpose of the three sections, 'thinking about their thinking'.

3.4.2 Semi-structured interviews: design phase

Semi-structured interviews have advantages compared to structured or unstructured interviews. They allow the researcher and participants to benefit from a participant's particular interest or experience because they are more flexible than the strict format of a structured interview. At the same time, they avoid the danger of a researcher losing

control of a completely unstructured interview (Wellington, 2015, p. 142). They are semi-structured in the sense that the researcher has an interview guide which consists of key questions and possible probes on the topics or issues to be discussed (Wellington, 2015, p. 243).

The primary aim of the interviews with teachers was to check that everything was clear and suitable for use with students, while also gathering preliminary information about teachers' views of self-assessment. The interview guide for teachers was piloted with two colleagues who were given Checklist v1.0 and the extended guide to study before the interview. They commented that all the interview questions were clear and did not offer any suggestions for improvement. The interview guide for teachers shown in Appendix C started with an explanation of what self-assessment meant in the context of this present study, followed by asking interviewees to choose one of three statements expressing different perspectives on self-assessment which best represented their own view as an opener to taking a more detailed look at the checklist. This encouraged them to engage with the topic and give more informed responses to the interview questions, as well as providing an insight into their perceptions of self-assessment. Interviewees were then invited to look at the checklist again for a few minutes before proceeding to four broad questions. These were: How clear are the items listed in the checklist?; How appropriate is the language for undergraduate students?; How clear is the layout of the checklist?; and How helpful is the extended guide?. Each of these areas contained further questions for the researcher to use if an important aspect of reviewing the checklist was not covered by the response to one of the broad questions. An example is presented in Table 3-3. The interview schedule concluded with a reminder to summarise what the interviewee had said and to ask whether there was anything else they wanted to add.

Table 3-3 Example area to explore with subsidiary questions

How clear are the items listed in the checklist?	
	Question
Q3	The checklist looks at pronunciation of whole words, stress, rhythm, and intonation. Do you think this covers all the aspects of pronunciation that it needs to?
Q4	You can see there are three sections (if they haven't already commented on this). What is your initial reaction to having these different sections?

3.4.3 Group interview

The aim of the structured group interview, 'to be sure the checklist and guide [extended guide] are clear and understandable', was shared with the students at the start of the session. The explanation of what self-assessment meant was followed by the selection of a statement which best represented their own view, as in the teacher interviews. The broad questions were: How do you feel about assessing yourself?; What are your first reactions to this checklist?; and How helpful do you think the extended guide will be for students completing the checklist?. Again, these were supported by subsidiary questions as illustrated in the extract in Table 3-4.

Table 3-4 Extract from group interview questions

Checklist items	
What are your first reactions to this checklist?	
Q1	What are your views on the items on the checklist?
Q2	What do you think about having the three stages of 'before', 'during' and 'after'?
Q3	What do you understand by the first item in the 'before' group…. Then the same for all the other items in this stage
Q4	In your view, is anything missing here?
Q5	In your view, is there anything here that does not need to be here?

The broad questions were intended to promote discussion among students, while the more detailed questions would offer an alternative approach if students were reluctant to participate, as often happens. In the researcher's experience of teaching these students, many of them are uncomfortable to have a discussion among themselves when a teacher is present, probably because they are used to responding directly to the teacher and only

when directly addressed with a question from the teacher. Despite the problems that were faced due to poor internet connection, the collected data turned out to be useful.

3.4.4 Expert validation: design phase

According to Olson (2005), "Expert reviews are frequently used as a questionnaire evaluation method but have received little empirical attention" (p. 295). When Olson examined the evaluation of two surveys by six expert reviewers, she found that experts could identify questions which would lead to inaccurate or absent data, despite "substantial disagreement… in the ratings across expert raters for both studies" (p. 303). However, their value in evaluating "the validity and practicality of the key components of …prototypes" has long been recognised (Nieveen, 1999, p. 133). As Tessmer (2013) acknowledged, expert reviews are an essential component of formative evaluation in development research (pp. 13-15). A Delphi method consisting of a panel of experts has been used in applied linguistics research in various ways, such as to create an instrument, or agree a definition of a concept (Sterling et al., 2023, p. 3). Whilst Sterling et al. (2023) see the potential benefits of involving stakeholders along with experts at an early stage (p. 9), others such as Richey and Klein (2014) see expert reviews as part of formative evaluation at more than one stage of development (p. 1108).

Copies of the checklist and extended guide were sent together with a copy of the ethics approval to two experts in the use of the ICAO scale for teaching and assessment. The covering email asked them if they were available and willing to carry out an expert review within a specified timescale, making it clear that they could ask for any additional information about the study they felt they might need, and could contact the researcher by email or telephone. One expert, who requested a telephone conversation to discuss the purpose and context of the checklist, made very few comments following the conversation. The other expert offered comments without additional knowledge of the context and it was clear from the difference between them that more information should

have been provided to set the context; it should have been made clear how the checklist would be introduced and used.

More contextual information was therefore provided in the covering email sent to a further seven international experts in the field of the assessment and teaching of pronunciation. The one positive response contained a request for more detail about what the researcher needed; this led to a search for examples of how scholars and doctoral students had dealt with requesting such information from experts.

The clearest example found in the search for previous uses of expert reviews was White (2011) whose doctoral thesis contained a clear explanation of what was required (pp. 49-50) as well as a rating form whose format could be adapted for this present study (pp. 193-200). White's documentation stood in contrast to vague email requests on the one hand and, on the other, relatively restrictive forms seeking conformity with education outcomes such as compliance with curriculum and learning objectives in a specific context. The complete final version sent to experts is shown in Appendix I, with a short extract presented in Figure 3-1 for illustration.

Expert·Reviewer·Rating·Form¶

Pronunciation·Self-Assessment·Checklist·for·Aircraft·Engineering· Technology·students¶

Student·instructions:·In·the·following·sections,·students·are·asked·to·click·on·the· frequency·level·of·each·criterion·as·follows:·1·=·Almost·never;·2·=·Rarely;·3·=· Sometimes;·4·=·Often;·5·=·Almost·always¶

¶

Expert·Reviewer·Instructions:·For·each·item·please·indicate:·¶

A.·How·relevant·the·item·is·to·the·processes·of·self-assessment·of·pronunciation.·Please·place·a· checkmark·in·the·appropriate·box.¶

1·=·not·at·all·relevant·2·=·slightly·relevant·3·=·moderately·relevant·4·=·highly·relevant¶

B.·Please·indicate·items·which·should·be·removed·or·reworded·by·placing·a·checkmark·in·the· appropriate·box.·If·you·wish·to·suggest·rewording,·and·have·time·to·do·so,·please·use·the·space· in·the·'reword'·column·to·do·so.·¶

Note:·Frequency·levels·are·greyed·out·to·limit·visual·distraction.¶

SECTION·A:·Before·the·speaking·activity·¶

Relevance¤				Evaluative·criteria¤	Frequency· level¤	Remove¤	Reword¤
·····Before·the·speaking·activity:¤							
1¤	2¤	3¤	4¤	1.·I·practise·my·pronunciation.¤	¤	¤	¤
1¤	2¤	3¤	4¤	2.·I·choose·words·which·I·can·pronounce·easily.¤	¤	¤	¤
1¤	2¤	3¤	4¤	3.·I·check·on·the·pronunciation·of·difficult·words.¤	¤	¤	¤
1¤	2¤	3¤	4¤	4.·I·pronounce·the·words·clearly·in·English.¤	¤	¤	¤

Figure 3-A Extract from expert reviewer rating form

3.4.5 Semi-structured interviews: calibration phase

The semi-structured interview guide for the students (Appendix E) explored the clarity and feasibility of the checklist from the students' viewpoints. They were asked in general how they felt about the clarity and feasibility of all stages of the process, highlighting the briefing session and video clips, the ease of using the checklist, and the suitability of the checklist as a classroom learning tool, including its administration via Google Forms and links. The researcher noticed that in some of the first interviews with students her usual role of a helpful teacher meant that she rephrased questions before giving students time to think or occasionally prompted them with a possible answer. However, she corrected this in the later interviews, rephrasing only where clarity was sought by the students rather than automatically.

The semi-structured interview guide for the teachers (Appendix F) was built around a series of statements which invited them to gauge their responses in terms of ease and helpfulness of the checklist and briefing session, the distinction between ICAO scale levels in the video clips, students' ability to self-assess their level using the checklist, and administrative aspects. Examples of the statements are shown in Table 3-5.

Table 3-5 Extract from semi-structured interview (teacher) calibration phase

Self-assessment process?	I'd like to talk about the process first and then go on to talk about the checklist, okay?
Q1	Thinking about the whole process… Statement/screenshare 1 Please choose ONE of the statements which is the closest to your own feelings: 1: I found the self-assessment process easy to do. 2: I found the self-assessment process difficult to do. 3: I found some of the self-assessment process easy to do, and some of the process difficult to do. Please explain your choice. Follow up: 'You said […]. Tell me more about that'. Use 'Go on' if they stop halfway and look at me.
Q2	Next, I'd like you to think about the briefing session for students at the start of the process. Statement/screenshare 2 Thinking about the briefing session. Choose ONE of the statements which is the closest to your own feelings: A The briefing session was very unhelpful. B The briefing session was unhelpful. C The briefing session was helpful. D The briefing session was very helpful. Please explain your choice. Follow up: If they suggest changes are needed, ask 'What changes would you like to see?' Then 'Why do you think that change is important?' If more than one change, ask about each one.

3.4.6 Evaluation of usefulness questionnaires: evaluation phase

The evaluation questionnaires in Appendix H were adapted from questionnaires used with previous checklists (Nimehchisalem and Mukundan, 2013; Nimehchisalem and Mukundan, 2015; Nimehchisalem et al., 2021). Bachman and Palmer (1996) explained the usefulness of a language test, which they mainly associate with performance (assessment of learning for decision-making purposes) as comprising reliability, construct

validity, authenticity, interactiveness, impact and practicality (p. 18). Reliability is related to tests in terms of consistency of scores between one version of a test and another as well as consistency among rater scores, and is generally measured statistically. However, if the pronunciation SA checklist is to support student learning, teachers would hope to see higher self-ratings given over a period of time. Construct validity can be defined in terms of how fully the checklist covers the construct of and skills involved in pronunciation. Authenticity relates to the correspondence between a test task and the use of the target language in real-life situations (Bachman & Palmer, 1996, p. 23). The speaking activity used with the checklist is treated as a learning experience rather than a test task and so authenticity has been excluded from the evaluation of usefulness of the checklist. Interactiveness has also been excluded because it involves the interaction between the nature of the test task and a test taker's language knowledge and ability (Bachman & Palmer, 1996, p. 25), whereas the checklist focuses on students' exploration of self-assessment and the strategies they use to develop their pronunciation. In this context, the evaluation of usefulness is closely related to impact rather than interactiveness. Both impact and practicality are important because the checklist is intended to support self-assessment. In addition to the domains, the underlying principles of constructing the questionnaire were the number and order of statements, the inclusion of reversed items and the use of Likert scale responses.

Dörnyei (2009) mentioned that using several items for a single aspect or domain of a scale can "maximize the stable component that the cluster of individual items in a scale share" (p. 25). Scales constructed in this way are multi-item scales. Dörnyei (2003) further suggested that, "Generally speaking, it is risky to go below 3–4 items per sub-domain" (p. 25) and recommended a minimum of "4 items per subarea" (Dörnyei, 2003, p. 34).

Additional sources were found in order to identify aspects of reliability which related specifically to ESL speakers' language ability, namely Dörnyei (2003), Jebb (2021), and Medina-Díaz and Verdejo-Carrión (2020). For example, the definition of the domain of reliability incorporated readability and the items were subjected to two readability tests available online, namely the Gunning fog index and the Flesch-Kincaid Grade Level Studies test, as recommended by Jebb *et al*., (2021). The Flesch-Kincaid Grade Level Studies test returned a score of 62.75, which is interpreted as being easily understood by 13-15 year old American students and approximately B2 Upper intermediate on the CEFR grid. Reliability also incorporated the comprehensibility of the instrument items and instructions in terms of the "vocabulary, grammatical structure, language and format of items [being] suitable for the student" (Medina-Díaz & Verdejo-Carrión, 2020, p. 267). The domain of validity focused on the construct of pronunciation as defined in this present study, based on the ICAO scale descriptors for pronunciation, stress, rhythm and intonation (Table 2-4): pronunciation of whole words, word stress, and overall rhythm and intonation. For the domain of impact, as suggested by Bachman and Palmer (1996, p. 137), and recommended by expert validation, the researcher drew up a "list of specific influences we would want the test to have and minimum acceptable levels for those" (p. 137). The list consisted of: metacognitive awareness; separated into knowledge of cognition and regulation of cognition; future use by individual students/teachers and end user satisfaction, linked to the affective domain of learning and perceived benefits for teaching as well as overall satisfaction.

The number of questions was limited in order to minimise the risk of reducing the response. Regarding the ordering of questions, the student and teacher questionnaires adopted Dörnyei's approach to arranging items in an evaluation questionnaire involving a multi-item scale: "What I usually do is take 4-5 content areas that are related to each other and then mix up the constituent items randomly" (Dörnyei, 2003, p. 60).

One reversed item was included in each of the domains (reliability, validity, impact and practicality). There are different views in the literature about the inclusion of reversed items, as mentioned by Józsa and Morgan (2017, p. 9). According to Józsa and Morgan (2017), Likert scales typically include some reversed items; some use an equal number of reversed and non-reversed items, while other scholars are opposed to using them. Reversed items can provide a check on whether questionnaire respondents are paying proper attention, and are therefore used to exclude questionnaires which show conflicting responses. Conflicting responses may indicate a lack of attention to completing the questionnaire, which can make some of the responses unusable for analysis.

A pragmatic approach was adopted in selecting 5-point Likert scales. Although 4- or 6-point scales can force respondents to avoid the neutral position, and Dörnyei stated a personal preference for 6-point scales (2009, p. 28), this present study used a 5-point scale to allow individuals who were genuinely undecided to express that view freely, as indicated by Johns (2005) and Weijters et al. (2010). For example, respondents may have wanted more experience of using the checklist before deciding whether they agreed or disagreed. Without a mid-point, respondents could have been forced into selecting a measure of agreement or disagreement when they genuinely neither agreed nor disagreed. Weijters et al. (2010) state that this can lead to "negative affective reactions" resulting in higher levels of disagreement, (p. 11) and "misreponse (MR) to reversed items" (p. 5).

The questionnaire items are shown along with the domains in Tables 3-5 to 3-7 (student) and Appendix H (teacher). In addition to using Bachman and Palmer's questions for logical evaluation of usefulness as guidance for designing questions (1996, p. 150), other sources of questionnaire items were employed. One such source, Nimehchisalem et al. (2021, p. 25), illustrates how items were adapted for inclusion in the present questionnaire. For example, the first item "I found it easy to work with the scale" was adapted to "I find it easy to work with the checklist" in the student questionnaire, and to

"The checklist is easy to work with" in the teacher questionnaire. The second item in Nimehchisalem et al. (2021, p.25), "I will use this scale to correct my own students' written works," became "I will use this checklist when I prepare for speaking activities" for students, and "I will use this checklist with my own students" for teachers. Some items such as "Using the checklist, I paid more attention to my pronunciation" (item 3, student questionnaire) were created to try to identify the impact of this particular instructional tool, while others related to practicality in terms of usability or ease of use at the research site, such as "I can fit the time the checklist takes to use into the class time I have available" (item 11, teacher questionnaire).

Two experts in the field of the evaluation of language assessment were contacted for validation of the questionnaires. An expert reviewer rating form was prepared asking reviewers to judge the relevance of each item and to suggest removal or rewording of items as appropriate. The rating form was emailed together with copies of the student and teacher questionnaires, the domain tables, the checklist, the ethics approval for the study, a covering email which set the context for the study, and the evaluation of usefulness questionnaires. Amendments were made to the specification of the impact domain following comments from one expert that the relationship should be clearer between each item and the aspect of impact to which it referred. The final versions of the evaluation of usefulness questionnaire are contained in Appendix H; the student versions of the components of test usefulness, aspects of impact, and summary of items by domain showing reversed items are shown in Tables 3-6 to 3-8.

Table 3-6 Components of test usefulness covered in the student questionnaire

Item	Text	Component
1	I find it easy to work with the checklist.	Practicality
2	I will use this checklist when I prepare for speaking activities.	Impact
3	Using the checklist, I paid more attention to my pronunciation.	Impact
4	The checklist fully covers the aspects of pronunciation I need for speaking activities.	Validity
5	I will encourage other students to use this checklist for a speaking activity.	Impact
6	Because of the checklist, I will have more interest in using the ICAO Language Proficiency Rating Scale as guidance.	Impact
7	All the terms in the checklist are clear and easy to understand.	Reliability
8	I found this checklist difficult to work with.	Practicality
9	The checklist follows a logical order.	Reliability
10	The checklist has made me think about different ways to improve my pronunciation.	Impact
11	This checklist has made me more likely to assess my pronunciation from time to time in accordance with the ICAO Language Proficiency Rating Scale.	Impact
12	I found some of the items in the checklist difficult to understand.	Reliability
13	I will use this checklist to review speaking activities.	Impact
14	Pronunciation is clearly defined in this checklist.	Validity
15	There are important elements of pronunciation missing from the checklist.	Validity
16	This checklist can be used with speaking activities in my technical engineering classes.	Reliability
17	The checklist is too complicated for me to use.	Practicality
18	Overall, the checklist will help me understand the ICAO/aircraft industry pronunciation standard.	Impact
19	The checklist helps me to assess all aspects of my pronunciation (whole word, stress, rhythm, intonation).	Validity
20	Using this checklist makes me worry too much about my pronunciation.	Impact
21	The checklist can be completed quickly.	Practicality
22	The checklist is easily accessible through Google Forms.	Practicality
23	What changes do you think can be made to the checklist to improve it?	All

Following expert validation, the relationships between specific aspects of impact and the items were made explicit and are shown in Table 3-7.

Table 3-7 Aspects of impact covered in the student questionnaire

Item no.	Text	Aspect of impact
2	I will use this checklist when I prepare for speaking activities.	Regulation of cognition
3	Using the checklist, I paid more attention to my pronunciation.	Regulation of cognition
5	I will encourage other students to use this checklist for a speaking activity.	Student satisfaction (Affective domain of learning)
6	Because of the checklist, I will have more interest in using the ICAO Language Proficiency Rating Scale as guidance.	Student satisfaction (Affective domain of learning)
10	The checklist has made me think about different ways to improve my pronunciation.	Knowledge of cognition
11	This checklist has made me more likely to assess my pronunciation from time to time in accordance with the ICAO Language Proficiency Rating Scale.	Knowledge of cognition
13	I will use this checklist to review speaking activities.	Regulation of cognition
18	Overall, the checklist will help me understand the ICAO/aircraft industry pronunciation standard.	Knowledge of cognition
20	Using this checklist makes me worry too much about my pronunciation.	Student satisfaction (Affective domain of learning)

The items associated with each domain of usefulness, along with reversed items, are summarised in Table 3-8.

Table 3-8 Summary of items by domain showing reversed items (student questionnaire)

Domain	Items (R indicates reversed item, e.g., R of 7 = reversed item of item 7)
Reliability	7, 9, 12 (R of 7), 16
Validity	4, 14, 15 (R of 4), 19
Impact	2, 3, 5, 6, 10, 11, 13, 18, 20 (R of 3)
Practicality	1, 8 (R of 1), 17, 21, 22

The scoring and analysis of the evaluation questionnaires was performed as follows. With the exception of the free response item, the first step was to reverse the scores of the reversed items, so that 1 was converted to 5, 2 to 4, and so forth, before all scores for individual students were input to SPSS (version 28). Each student's responses were inspected for items with conflicting responses to reverse items and those questionnaires removed from the calculation. Responses to each item ranged from 1 to 5 on the Likert scale, which meant that for a single questionnaire the minimum score was 1 multiplied by the number of positive items plus 5 multiplied by the number of reversed items. The maximum score was 5 multiplied by the number of positive items plus 1 multiplied by the number of reversed items. The total of item scores was then calculated along with the mean for that questionnaire.

To assign the mean score from a single questionnaire to one of three categories of usefulness (useless, moderately useful, and very useful), cut-off points were assigned for a Likert scale from 1 to 5 by using terciles, a standard statistical method, to give an interval range of 1.33. This gave the following score categories: useless from 1 to 2.33; moderately useful from 2.34 to 3.66; and very useful from 3.67 to 5.0.

The questionnaire was piloted with eight students. The students were interviewed after completing the questionnaire and asked if any items were unclear, difficult to understand or difficult to answer. Two students said they found the reversed items difficult to understand and answer, and were confused by them being "the other way round", needing too much thinking, and not being sure if they had answered them correctly. However, the researcher felt the need to retain the reversed items in order to maintain content validity (Weijters & Baumgartner, 2012) as well as helping to ensure that students were more likely to complete the questionnaire carefully. After excluding questionnaires with conflicting responses, Cronbach's alpha values were calculated. For the individual domains, values were as follows: the reliability domain which consisted of 4 items (α =

.885), the validity domain, also with 4 items (α = .887), the impact domain which consisted of 9 items (α = .927) and the practicality domain of 5 items (α = .710). The overall Cronbach's alpha for the 22 items was α = .963 which indicates high internal reliability.

Following the pilot, from the total of 50 students, the remaining 42 were sent a link to the evaluation questionnaire and 32 responses were received, a response rate of 76%. Seven questionnaires were excluded from analysis because of conflicting responses with reversed items. The overall Cronbach's alpha was 0.905, while for the individual domains, the results were: reliability 0.642, practicality 0.652, validity 0.796 and impact 0.837. It has been shown mathematically that "Cronbach's alpha underestimates true reliability" (Sijtsma, 2009a, p. 170) and "there is no clear and unambiguous relationship between alpha and the internal structure of a test" (Sijtsma, 2009b, p. 169). Thus, the results of the questionnaires presented in Section 4.6 were considered within a broader consideration of usefulness as described in the Chapter 5 discussion.

3.4.7 Expert validation: evaluation Phase

Additionally, three further assessment experts responded to a request for validation of Checklist v2.0, with the documentation updated with changes as reported in Section 4.4.

3.5 Participants

This section describes the sampling strategy before providing information about the three main groups of participants and their involvement in each stage: students, teachers and experts. The pronunciation self-assessment checklist was designed for undergraduate BAET students who were studying in a particular university. The university specified the Aviation English syllabus for those students, most of whom would work as air maintenance technicians after graduating. The population consisted of the students and their teachers involved in delivering Aviation English classes. Participants in

development research typically include some combination of designers, developers, students, teachers, and experts (Richey & Klein, 2004, p. 1115). This can be considered a particular case of what Creswell terms purposeful sampling (2014, p. 32). In purposeful sampling, the inclusion criteria are defined, based on "who can best help you understand the central phenomenon you are exploring" (Creswell, 2014, p. 76). The individuals who could best contribute to the development of a classroom-based pronunciation self-assessment checklist were the end users (students and teachers) and experts who could validate the checklist.

The sample sizes were: three of the six teachers (including the researcher); fifty students in total from three classes, one class taught by each teacher; and a total of seven experts.

Table 3-9 Numbers of participants by research phase

Phase	Type of participant	Number	Activity
Design	Expert in area	2	Validation of Checklist v1.0
	Teacher	2	Semi-structured interview
	Student	8	Structured group interview
Calibration	Teacher (including researcher)	3	Trialling the checklist
	Student	50	Trialling the checklist
	Teacher	2	Semi-structured interview
	Student	10	Semi-structured interview
Evaluation	Expert in the area	3	Validation of Checklist v2.0
	Expert in the area	2	Validation of evaluation of usefulness questionnaires
	Student	8	Piloted evaluation questionnaire
	Student	32	Returned questionnaire
	Teacher	2	Returned questionnaire

3.5.1 Aviation English teachers

The researcher spoke to the other five Aviation English teachers, explained the research and what their participation would involve for themselves and their students, then invited them all to indicate their willingness to take part and their availability. Two of the teachers responded positively, one of whom was a qualified ICAO examiner. They took part in semi-structured interviews in both the design and the calibration phases, as well as administering the checklist in class in the calibration phase.

Table 3-10 Teacher qualifications and experience

Teacher	Level of qualification	Major	Teaching experience
T1	PhD	English	23 years
T2	MA	World Literature	8 years

3.5.2 Student selection

Fifty students in total from the three classes gave their consent to participate in the study. The sampling was purposeful because the checklist was designed for use with BAET students at the research site, and student participation depended on the willingness of their teachers to allocate class time for the briefing session, the speaking assignment, and completion of the checklist and evaluation questionnaire.

All fifty students were in their first year, 8 in the second semester and 42, who had previously completed a diploma, in the fourth semester. All were aged between 18 and 20 years old. Table 3-11 presents their characteristics by gender and level of English.

Table 3-11 Students by gender and level of English

	Level of English					
	A	A-	B+	B	C	Totals
Male	23	11	3	2	1	40
Female	4	2	2	2	0	10
Totals	27	13	5	4	1	50

For practical reasons, in order to avoid undue difficulties with timetabling and organising the group interview outside study hours, as well the difficulties of off-campus learning during the COVID-19 pandemic, the smallest class consisting of eight students was selected for the group interview in the design phase.

All fifty students completed Checklist v2.0, after which volunteers were invited to be interviewed in the calibration phase and ten responded positively. The volunteers came from all three classes, included female as well as male students and a representative range of English ability.

3.5.3 Experts

A total of seven experts contributed to the study. Four were invited because they were established researchers in the area, two in the design phase for validation of the checklist, and two in the evaluation phase for validation of the evaluation questionnaire. Three others responded to a request for checklist validation which was sent to a total of seven experts identified through a search of literature for academics with expertise and potential interest in the area.

3.6 Data analysis

Data collected from the structured group interview and semi-structured interviews were recorded and transcribed before analysis. The data were analysed qualitatively, based on development of the codebook, including intercoder reliability, described in Sections 3.6.1 and 3.6.2 respectively.

Student questionnaires were analysed with descriptive statistics using IBM Statistical Package for the Social Sciences (SPSS) Statistics 22. Teacher questionnaires were analysed separately as there were only two of them and the questionnaire items differed from those in the student version of the questionnaire.

The data analysis methods are summarised in Table 3-12.

Table 3-12 Data analysis methods

INSTRUMENT	DATA ANALYSIS
Demographics questions on level of spoken English	Information about sample
Structured group interview	Records were kept of decisions made, reasoning behind the decisions and agreed actions (changes to Checklist v1.0) based on student feedback and comments on the clarity and feasibility of the checklist. Thematic analysis of data pertaining to students' views regarding more general aspects of self-assessment was carried out with coding based on structural codes, theory-based codes and data-driven codes.
Semi-structured interviews	Thematic analysis was used with coding based on structural codes, theory-based codes and data-driven codes.
Evaluation questionnaire	Quantitative analysis of responses (frequency, mean, calculation of usefulness).

3.6.1 Codebook development

'A code in qualitative inquiry is most often a word or short phrase that…assigns a…salient, essence-capturing…attribute for a portion of language-based or visual data. […] Just as a title represents and captures a book or film or poem's primary content and essence, so does a code represent and capture…primary content and essence.' (Saldaña, 2009, p.3). DeCuir-Gunby et al. (2010) report that there are several ways in which codes can be developed (p. 137). One method, which they describe as data-driven, involves codes emerging from the raw data (DeCuir-Gunby et al., 2010, p. 137). However, as Byrne (2022) points out, it is unlikely that codes simply emerge in this way, but rather that they are influenced by the researcher's interpretation of the data (p. 1397). Nonetheless, these codes have their origin in the data. A second way in which codes are developed originates in the aims and objectives of a particular study, which can be termed structural coding (DeCuir-Gunby et al., 2010, pp. 137-138). Saldaña (2013) explains that structural codes are like labels, allowing easy retrieval of all data relevant to a particular point of enquiry (p. 84). The third option is to develop codes from concepts or theory which already exist, termed theory-driven codes by DeCuir-Gunby et al. (2010, p. 137).

Saldaña (2013) mentions that researchers need to have a flexible approach to code development, as it is usual for coding schemes to evolve through a cyclical process (p. 37). In similar vein, DeCuir-Gunby et al. (2010) note that a researcher who is developing structural or data-driven codes returns a number of times to inspect and consider the data, whereas a researcher who is developing theory-driven codes has to repeatedly return to the theory (p. 138).

In this present study, all three types of code were applicable to the data collected. The first set of codes comprised structural codes which were directly related to the specific questions asked in the review of the checklist at the design stage and later in the calibration phase. The second set consisted of codes from the theory on which the design of the checklist was based, namely self-assessment and metacognition, in line with DeCuir-Gunby et al. (2011) who state that "Codes are generated from the theories that guide the research" (p.141). The third set of codes, or data-driven codes, was developed from other data provided by participants, was not covered by the first two sets of codes and which did not fit easily into the theory. The codebook itself was set out under four headings: theme, code name, definition and sample data. Intercoder and code reliability procedures led to minor changes (Section 3.6.2) and the final version of the codebook is shown in Appendix L.

The process of developing codes and the codebook began with dividing the transcripts into the smallest possible chunks of meaning and allocating a word or phrase to the content of each chunk, as well as notes of thoughts about what a broader category might be. The researcher took care during the data analysis process to identify codes and themes from the transcripts, regularly pausing to check whether she was imposing her values and understanding on what participants had said, and whether she had accurately represented their meaning. She chose not to draw up a list of possible codes or themes at the start of the data analysis process in order to minimise her personal views, coding several

interviews based on participants' own words before starting to draw up a codebook. During these early stages of data analysis, she was aware of moving from an insider to more of an outsider position, revising the coding of data, and the codes themselves, several times and constantly checking that she was thinking back to the theory as well as the collected data.

An example from S3 calibration phase interview is shown below, where the researcher made an assumption that the student could actually self-assess his pronunciation rather than considering how this might relate to the theory underpinning the checklist.

Table 3-13 Example of researcher's assumption in initial coding

Excerpt	Code	Notes
It's OK, I feel very good after I have done it because when I take a look at the questions of before, during and after speaking, the speaking activity, I can reassess and re-evaluate my speaking, intonation, how I speak in rhythm, so I can improve myself in the future when I speak…er…during the speaking activity.	Student was able to self-assess his pronunciation ability.	

In the first review of codes, this chunk was divided into three to unpack the ideas and code them separately with a closer association with self-assessment, its purposes and processes, and with student independence, as shown in Table 3-14.

Table 3-14 Example of amended initial coding

Excerpt	Code	Notes
It's OK, I feel very good after I have done it because when I take a look at the questions of before, during and after speaking, the speaking activity, I can reassess and re-evaluate my speaking, intonation, how I speak in rhythm, so I can improve myself in the future when I speak…er…during the speaking activity.	SA to identify language proficiency ability.	Student independence
I can reassess and re-evaluate my speaking, intonation, how I speak in rhythm	SA review	Student understanding
so I can improve myself in the future	Use to improve	Purpose of SA

After all transcripts had been coded in this way, codes were compared within and across transcripts. Two further examples which related to the excerpt in Table 3-14 were:

> (a) "I have to think about what...like…how I do things like the pronunciation stuff and that...so I have to take some time, trying to think back, like what do I do if I get some words I don't know how to pronounce or I don't understand words, do I even look it up" (S10).

This was initially coded as 'metacognitive awareness' with a note 'student knowledge' because thinking about the underlying theory increased as more transcripts were coded.

> (b) "Yeah, because when I… when I speak sometimes I don't assess myself unless if I… I look after the… the video after I do the speaking activity" (S8).

This was initially coded as 'metacognitive monitoring' because thinking about the underlying theory continued to increase as work progressed.

Coding of similar chunks in other transcripts were 'SA process' (in S9 and S10), and 'metacognitive awareness' (S12 and S13). Theory codes were also being developed and refined, with various categories under metacognition, such as 'metacognitive strategies' and 'metacognitive awareness', being clarified and redefined as 'knowledge of cognition', 'regulation of cognition' and 'metacognitive experiences' following several readings of the relevant theory. All the above examples were finally coded as 'regulation of cognition' (RC), which was defined as 'Remarks about planning and/or monitoring and/or evaluating and/or reflecting on what they do or have done, as well as comments on the processes involved in any or all of these' and was the final code assigned to all the above chunks of text.

Comparison and refinement of codes was carried out twice more before the first draft of the codebook was produced, after which overlapping areas of theory and data-driven codes were removed. For example, 'student understanding of self-assessment' as a data-driven code was separated from theory codes 'purpose of self-assessment' and

'standards'. Once the codes were separated and defined, intercoder and code reliability procedures were conducted.

3.6.2 Intercoder and code reliability

This section describes the procedures used for determining intercoder reliability and the reliability of codes, the final stage of developing the codebook according to DeCuir-Gunby et al. (2011, p. 146). Intercoder reliability is a numerical measure of the extent to which two or more coders agree on the codes and how they are assigned in a particular set of qualitative data (Kurasaki, 2000, p. 179). It has been increasingly used in qualitative data analysis to improve the rigour and transparency of the analysis (O'Connor & Joffe, 2020, p. 3). The procedures chosen should reflect the nature of the project and the data (O'Connor & Joffe, 2020, p. 6).

From the fifteen transcripts of interviews in this present study (two interviews with each of two teachers, ten student interviews, plus a group interview), three were selected as a sample. They were chosen to give a sample of participants and type of interview. These were one teacher interview from the design phase, the group interview with eight students from the design phase, and one student interview from the calibration phase.

A second coder was identified after consideration of several colleagues outside the research team as well as the two teachers involved in the research. Whilst some scholars suggest the coder should be external to the research, this could have raised issues of ethics and data protection, as mentioned by O'Connor and Joffe (2020, p. 6). The most suitable available second coder was one of the teacher participants, who offered the specific advantages of knowing the context and learner characteristics in addition to being a qualified ICAO examiner. The teacher had no knowledge of the theories that underpinned the present study, nor had they been involved in the development of codes. There is a precedent for the selection of an appropriately qualified teacher (Cheung & Tai, 2021, p. 1169).

The sample that was given to the second coder was already divided into chunks by the researcher, as suggested by Campbell et al. (2013) and O'Connor and Joffe (2019), who assert that the emphasis is on consistency of coding rather than on the consistency of segmenting text. The second coder then coded one transcript first in order to carry out an informal check on code patterns to identify any obvious problems with code definitions or interpretations to allow the coding frame to be refined before commencing the formal independent coding with the larger subset of data (O'Connor & Joffe, 2020, p.6). One of the transcripts from the calibration phase data was used for this.

Two meetings were held with the second coder. The first session, which lasted about an hour, was dedicated to an overview of the process, clarified any questions about code definitions and confirmed that the coder's input would be helpful and valued. As the second coder was to complete the process independently, the second session was held after the first transcript was coded, to identify and resolve any issues.

At the first meeting with the researcher, the second coder was given the codebook, the instructions for coding and one transcript. The instructions to the second coder are shown in Appendix K. The second coder was asked to code the S3 student calibration phase transcript for two reasons: it was a typical student transcript and contained a greater range of codes than the group interview or teacher interview. At the meeting to compare and discuss the codings of this transcript, 5 of the 18 codes differed. Differences in coding were: CR ADMIN (checklist review administration) vs CR SECTION (checklist review section), CR SECTION vs CR ADMIN, SUSA (student understanding of self-assessment) vs CR GUID (checklist review guidance), SUSA vs RC (regulation of cognition).

Two unresolved differences were due to the fact that the items could have been multiple coded, whereas the instructions to the second coder only allowed for single coding. They

were related to the distinction between checklist administration (the links) and checklist sections, for example:

> "No, everything is clear, 'cos when you click …er…you click the first time before the speaking activity, before we speak we speak in [unclear - our activity?] you give us the…er…er…what you call it? the Section A, so before, we have to answer before [unclear] and then after we speak, you give us the Section B and the section C, so it's very…er…very helpful and very easy to follow, for me at least." (S3)

The researcher coded this as CR ADMIN (checklist administration) because it referred to links to the checklist, while the second coder used CR SEC (checklist review section) because it referred to sections of the checklist.

In the second example, "I like the before, during and after speaking activities" was coded as CR SEC by the researcher and CR ADMIN by the second coder, who chose CR ADMIN because the word 'section' was not used, and the 'before, during and after' could refer to when the links were given.

These two examples did not affect data analysis and interpretation because both aspects were addressed in the development of the checklist.

One difference of opinion was easily resolved. The following two chunks were coded as student understanding of self-assessment (SUSA) by the researcher and checklist review guidance (CR GUID) by the second coder:

> "Er…I think you do it inside class with the students. Like…yeah like section A until section C they can look at independently."
>
> "For me, no…I need a professional guide."

The codebook was quite clear that 'guidance' only referred to the guidance contained in the checklist, whereas these statements implied that guidance came from the teacher until students could complete the checklist independently. In the light of this, the second coder agreed that these chunks should be coded as SUSA.

Additionally, it was agreed not to code very general overall comments at the end of the transcript: “Yeah, it’s more than enough, yeah”, “Yeah. It’s not that long”, and “No er I think the checklist is all good you know”.

One change was made to the codebook, when, at the end of the meeting, the researcher asked if the second coder had any other questions about any of the codes. The second coder asked what would happen if a teacher talked about students’ roles as learners, having observed that ‘Students talk about their roles as learners’ would not allow a teacher’s comment to be coded. Thus, the existing entry in the codebook was amended to ‘Interviewees talk about students’ roles as learners’ to enable coding of comments from teachers as well as students. The final version of the codebook is shown in Appendix L.

The second coder then coded two more transcripts of T2 teacher design stage and GI group interview, after which the percentage agreement of coding was calculated. This gave percentage agreements of 81% and 91%, an average of 84.6%; the high figure of 91% percentage was due to the number of transcript chunks related to responses about checklist items. Taking the average figure of just under 85%, this was sufficient to give the researcher confidence in conducting the full data analysis (Cheung & Tai, 2021, p. 15).

For T2 (teacher design stage), 5 of the 26 codes differed. The differences are illustrated in the following examples.

> “No...mmm…I did have a few questions like well when I’m going through it. Like how do the students…how can the students actually identify if they are speaking with the wrong rhythm or the right rhythm, or how are they going to be able to see that, because if they feel like…like…like the example I gave you just now, they feel like they’re OK but to our ears…like…something is wrong, but how do they see that? how do they like analyse that on their own?” (T2)

The researcher coded this as SAST (self-assessment standards) because the teacher was talking about whether students could hear the difference between right (correct standard)

and wrong rhythm. Meanwhile the second coder selected ME (metacognitive experience) because the statement referred to how students might feel when assessing their pronunciation. The codebook makes it clear that ME is about somebody's own learning and experience, whereas here the teacher was talking about students. In the light of this, SAST was the more suitable code.

The second example was:

> "Mmm…By giving examples…more examples, like for example I will give an example of the checklist and then explain to them if I write 'almost never' and then if I write, if I choose 'almost never', 'rarely' or 'sometimes', what would the end […] result be". (T2)

The researcher selected CR GUID (checklist review guidance) with the advantage of knowing the question that was asked about how the guidance could be improved. The second coder chose CR FREQ (checklist review frequency scale) since the teacher stressed the frequency scale as an example of additional guidance. Without the question, the second coder lacked the context of the chunk given. Hence, this chunk should be coded as CR GUID.

The third example was:

> "Yes. I believe, so it is in according to the order, like you start with the frequency 1 to 5 and then definitions, of course we need to know like the definition first before you let them…" (T2).

The researcher coded this as CR GUID because the teacher was talking about the frequency and definitions in the checklist which form part of the guidance. However, the second coder identified this as CR LAYOUT (checklist review layout), perhaps because this chunk included definitions as well as the frequency scale that had led to the coding of CR FREQ in the previous example. According to the codebook CR LAYOUT was defined as 'the font and spacing of the layout are easy to read, or not, and/or on whether there is not too much or too little information', so CR LAYOUT may have been a default choice by the second coder. For the analysis the code CR GUID was used because the definitions

of stress, rhythm and so forth were included under 'and/or the wording of guidance items and instructions' in the codebook.

The final step in developing the codebook was to determine the intercoder reliability and the reliability of the codes. In order to establish intercoder reliability, the approach used was to calculate a basic proportion of agreement. Miles and Huberman (1994) suggest calculating reliability as the number of agreements divided by the total number of agreements plus disagreements (p. 64). A reliability of 0.8 or better is generally considered acceptable (DeCuir-Gunby et al., 2011, p. 9). The present research focussed on determining reliability in terms of how the codes represented the interview data in terms of the checklist review codes, codes from theory and data driven codes. The reliability percentage was calculated using Microsoft Excel.

In this development study, the analysis as a whole is presented in Chapter 4 as a narrative of the process, with the quantitative analysis included towards the end of the process.

3.7 Ethical considerations

Ethical approval for this research was granted by the University of Sheffield (Reference Number 046390, Appendix O). British Educational Research Association (BERA) guidance (2011, p. 5) states that "all educational research should be conducted within an ethic [sic] of respect for the person, knowledge, democratic values, the quality of educational research and academic freedom". The informed consent of all teacher and student participants was obtained. The researcher gave them the information they needed about the research to decide whether to participate (information sheet and consent form; see Appendix N). They were informed that their participation was voluntary and that if they so wished they could withdraw at any time. Permission was also sought and obtained to use demographic data such as years of experience and age.

3.8 Summary

This chapter has presented and justified the selection of the mixed methods approach adopted in this development study. It has outlined the phases and steps in the research process and provided details of the participants and setting of the research. The data collection methods have been described, namely: expert validation, structured group interview, semi-structured interviews, and evaluation of usefulness questionnaires. The instruments and procedures, including procedures for intercoder reliability, were explained in detail followed by data analysis and ethical considerations.

4 CHAPTER 4 RESULTS

4.1 Introduction

This chapter reports the development of the checklist through the design, calibration and evaluation stages, showing how comments from experts, students and teachers were used to design, define, and evaluate the checklist. Additionally, it presents end users' views of self-assessment and their experience of trialling the checklist that was gathered from the interviews in the design and calibration stages. Unredacted quotes from students are given in the language they used in order for the reader to gain a sense of their overall speaking ability.

4.2 Design Phase Changes to Checklist v1.0

This section describes the changes made to the checklist in the design phase (Checklist v1.0, Appendix A) based on comments from teachers, students and experts. It proved useful to gather comments from the different sources because they contributed a range of perspectives. The contributions from the experts are presented first, followed by the contributions from the end-users (teachers and students). Some of the items were modified while some suggestions were rejected.

The experts proposed a variety of changes from differing viewpoints. Expert E2 commented that the instructions at the beginning of the checklist did not directly address the objective to students but stated 'The objective of this checklist is for students to…'. This criticism was addressed by separating the different components and adding 'you': for example, 'This checklist is to help you assess your own pronunciation' and 'Along the way, you will find guidance to help you complete the checklist'. The same expert questioned the need for the solicitation of demographic information which appeared at the beginning of the checklist. However, it was considered important to collect minimum demographic information so that the researcher could identify different types of

respondents if needed during data analysis, for example if comparisons were required by class or level of English.

The same expert raised the issue of tenses used throughout the checklist and suggested that all the items should be expressed in the past tense if the checklist was to be used for self-assessment purposes. This assumed that the student would use the checklist after performing a speaking activity. However, Section A (the 'before' section) was designed for use while students were preparing for a speaking activity, and the use of the present tense in Section B (the 'during' section) was designed to make the experience feel immediate again, reliving the experience and helping to promote reflection-in-action. The original choice of tenses was justified based on examples in the literature (Oxford, 1990; Vandergrift et al., 2006) and therefore the use of the present tense was maintained.

A suggestion from E2 that the item 'I pronounce the words clearly in English' should be reworded to read 'I try to pronounce each and every word clearly in English' was adopted. The researcher agreed with E2 that using the word 'try' would give average or weaker level students more opportunities to feel included and encouraged in the process of completing the self-assessment checklist. For the domain of impact, as suggested by Bachman and Palmer (1996, p. 137), and recommended by expert validation, the researcher drew up a "list of specific influences we would want the test to have and minimum acceptable levels for those" (p. 137).

'I self-correct my pronunciation during the presentation' was refined to 'I self-correct my pronunciation whenever I mispronounce'. It was highlighted by E2 that students could only self-correct if they knew their pronunciation was wrong and it would be impossible to self-correct if they were unaware it was wrong. The change was therefore made by the researcher.

A major concern raised by both E1 and E2 centred on the use of technical language and assessment standards. On one level, the terms 'stress', 'rhythm' and 'intonation' were said to need more explanation. For example, E2 wrote a comment on the item 'I took note of where I spoke with the wrong rhythm', saying "I don't think students know about rhythm in pronunciation. Rewrite the statement by describing what you mean instead"; similar written comments were made on the statements about 'stress' and 'intonation'. This concern was considered at length but ultimately rejected on the basis that the literature states that teaching these technical elements of English requires practical examples, exercises and practice (e.g., Celce-Murcia et al., 1996; Levis & McCrocklin, 2018). Explanations in written form were unlikely to be helpful. Accordingly, this concern was addressed by expanding the researcher's briefing session for students to include explanations and examples of the technical terms, in addition to including the key words in the extended guide, as described in sections 3.3.1 and 3.3.2.

The briefing session allowed students the opportunity to seek clarification of technical terms, while enabling the researcher to ensure students had the understanding they needed to be able to assess their learning strategies and themselves. A suggestion to include a description of the extended guide itself in the introduction to the checklist was addressed by the inclusion of the guidance in the checklist.

Expert E1 questioned whether there should be more guidance for Section B ('During the speaking activity') which contained guidance on the use of the frequency scale in this section but did not contain the 'why' and 'how' guidance included in the other sections. There was no suggestion about what guidance might be needed, rather that its absence was noted, and was perhaps an omission on the part of the researcher. No action was taken by the researcher at this stage because it was considered that, in contrast to sections A and C (Before and After the speaking activity) when students had time to practise beforehand and reflect afterwards, they would be unlikely to refer to the extended guide

for section B during their speaking activity; at best this would be challenging and distracting for the students, and it is highly unlikely they would be able or willing to use it.

The major change was the incorporation of the content of the extended guide into the checklist as the Google Forms version of the checklist was produced, after which the extended guide was simply referred to as 'guidance'. Some participants suggested that all the guidance should come first "so that students can see the bigger picture and they can see what is that idea of doing this assessment" (S8), and to increase the probability of students using it (T1). Other participants proposed it should relate to the section to be completed: for example "I think it should before, they know maybe like where they are at now and then proceed with the next one" (S5), and "So once the students are done with that section then they can move on to the next section, in other words easier, so they can concentrate section by section" (T1). The researcher's experience of teaching BAET students supported the proposal to place sections of the guidance immediately before the checklist section to which it referred. The decision to place the guidance just before the section to which it referred was in the end determined by the use of Google Forms to administer the checklist, "because the format in Google form, when we try to squeeze everything in one page, so it won't look good" (S7). This prompted the researcher to take care to ensure clarity of page content and achieve an optimal balance of content on different pages. Two students raised the issue of unnecessary scrolling up and down, one saying "First, I didn't understand what were stress, but when I scroll down, there's definition. Maybe you can put the stress definition, the syllable definition up top" (S3). The students' comments were very helpful regarding the layout and use of Google Forms; they had more experience of using Google Forms than the teachers or researcher.

The majority of changes related to ensuring the clarity and readability of the checklist, the instructions, items and guidance. Teachers suggested that instructions would be better

given in bullet form because students are used to this, and the researcher duly changed the instructions from sentences in paragraph to bullet point format. This format is recommended in Web accessibility guidelines (Accessibility Guidelines Working Group, 2022). In addition, using a format familiar to students would help to keep their focus on pronunciation self-assessment.

Suggestions which improved the readability of the checklist were accepted. To reduce redundant words in agreement with Accessibility Guidelines Working Group (2022), the researcher removed verbs from items in Sections A, B and C and replaced them with an introductory phrase, as proposed by teachers: for example "like 'before the speaking activity', that is the key sentence 'I pronounce the words' and the second one 'I choose the words I can pronounce easily'… instead of 'when preparing for speaking'" (T1). The original and revised versions of items in Section A are shown below.

SECTION A: Before the speaking activity

ORIGINAL VERSION	REVISED VERSION
Evaluative criteria	Evaluative criteria
	Before the speaking activity
1. I practise my pronunciation before a speaking activity.	1. I practise my pronunciation.
2. When preparing for a speaking activity, I choose words which I can pronounce easily.	2. I choose words which I can pronounce easily.
3. I check on the pronunciation of difficult words before a speaking activity.	3. I check on the pronunciation of difficult words
4. When I am preparing, I pronounce the words clearly in English.	4, I pronounce the words clearly in English.
5. As part of my preparation, I stress the words accurately in English.	5, I stress the words accurately in English.
6. During preparation, I speak English with a regular rhythm.	6, I speak English with a regular rhythm.
7. Before a speaking activity, I practise speaking English with a natural intonation.	7. I practise speaking English with a natural intonation.

This change was consistent with readability guidelines such as those in the Australian government Style Manual (www.stylemanual.gov.au). A summary of accepted changes and rejected suggestions is presented in Table 4-1.

There were several reasons why some suggestions were rejected. A proposal from T1 to use actual student performances to illustrate ICAO levels was rejected because the researcher had not been granted ethics approval for this. Another suggestion from T1 was that students would find the frequency scale less confusing if the end points were 'never' and 'always'. After some consideration, the researcher decided to retain 'almost never' and 'almost always' as the end points of the frequency scale in preference to the suggested 'never' and 'always' on the basis that absolute extremes could force students who perhaps did something once or twice into a response which did not reflect accurately what they did. This was supported by Wyatt and Meyers (1987), who found that avoiding polar opposites generated a broader spread of responses. Additionally, 'almost never' allows students with very limited experience to feel a little less negative in responding.

One teacher mentioned that statements in Section C should be rearranged to avoid influencing responses to the statements which followed 'I reviewed a recording of my speaking activity for self-improvement'. Since this statement came before others such as 'I listed down the words I mispronounced', it was felt that students would think that responses to other statements should be based on the recording and would therefore not respond to them. However, the researcher concluded that rearrangement would not have improved the logical flow of ideas and thus would not have improved clarity or readability. Each statement was independently based on the introductory statement which is 'After the speaking activity', as shown in Figure 4-1.

SECTION C: After the speaking activity

Please circle (1-5) to indicate the frequency level of each criterion according to the key given.

1 = Almost never
2 = Rarely
3 = Sometimes
4 = Often
5 = Almost always

Evaluative criteria	Frequency level
After the speaking activity:	
16. I reviewed the recording of my speaking activity for self-improvement.	1…2…3…4…5
17. I listed down the words I mispronounced.	1…2…3…4…5
18. I took note of the words that I stressed inaccurately in English.	1…2…3…4…5
19. I took note of where I spoke with the wrong rhythm.	1…2…3…4…5
20. I took note of where my intonation caused problems for my listeners.	1…2…3…4…5
21. I listened to correct examples of pronunciation in English.	1…2…3…4…5
22. I practised speaking correctly after listening to examples of pronunciation in English.	1…2…3…4…5

Figure 4-A Checklist extract: 'After the speaking activity'

Two additional statements were suggested by students. The first suggestion 'I often speak English. I always speak English at home or in classroom' was rejected because, although this could help to improve speaking, it did not focus on learning strategies and self-assessment behaviours related to components of pronunciation. The second suggestion was to add something like 'How often do I use a complex word in my speaking? During my presentation, how often do I engage with complex words in my speaking?'. This was rejected because it fell under the heading of vocabulary and not pronunciation.

Table 4-1 presents a summary of the changes proposed by the teachers and students during the design phase, together with the researcher's response, the rationale for actions taken, and an indication as to whether the proposed changes were suggested by teachers, students or both.

Table 4-1 Summary of changes proposed by the teachers and students: Design phase

Suggestion	Researcher's response	Rationale	Source (T = teacher, S = student)
Overall changes			
The Google Forms version of the checklist should ensure pages are not too full.	The clarity of page content and the amount of content on different pages were optimally balanced when the checklist was transferred to Google Forms.	This option helps to ensure the checklist is easier to read and hence facilitates completion by the students.	S
The extended guide should be incorporated in the checklist itself, as close as possible to the section to which it referred.	The guidance was incorporated at the start of the section it referred to.	This option reduced the need for scrolling up and down (a distraction) when completing the checklist.	S, T
Specific changes			
Instructions would be better given in the bullet form because students are used to this.	The instructions were changed from sentences in paragraph to bullet point format.	Bullet points are recommended in Web accessibility guidelines. In addition, using a format familiar to students would help to keep their focus on pronunciation self-assessment.	T
Sections A, B and C should be introduced by a key phrase, e.g. 'Before the speaking activity' for Section A, to reduce repetition of 'preparing' in individual items.	A key sentence was used before each section and words repeated in different items within the section were excluded.	The removal of redundant words is recommended in Web accessibility guidelines. This makes the checklist more economical.	T
Rejected			
Video clip examples should use actual student performances.	This could be considered in the future.	The researcher did not have ethics approval to do this in the present study.	T
Frequency scale should use 'never' rather than 'almost never' to distinguish more clearly between 'almost never' and 'rarely'.	The use of 'almost never' and 'almost always' is supported in the literature. It allows students to respond more freely and avoids the potential full negative effect of 'I never do this'.	The absolutes of 'never' and 'always' may produce a narrower range of responses (Wyatt & Meyers, 1987) and force students who have done something once or twice into a response which does not reflect accurately what they do.	T

Suggestion	Researcher's response	Rationale	Source (T = teacher, S = student)
Statements in Section C should be rearranged. The first statement, 'I reviewed a recording of my speaking activity for self-improvement', could lead students to feel that if they had not reviewed a recording, they could not respond to the statements which followed.	All the statements that follow the key statement 'After the speaking activity' can be responded to equally and independently of the others. The format is the same as in the first two sections.	Rearrangement would not have improved the logical flow of ideas.	T
The definitions in the extended guide are sufficient, so the links for definitions are not necessary and should be removed.	The links are available for students who want to listen, or listen again, to samples illustrating key words to improve their understanding.	Giving examples of stress, rhythm and intonation is recommended for teaching these concepts in ESL (e.g., Celce-Murcia et al., 1996; Levis & McCrocklin, 2018).	T
The extended guide should include brief explanations of Sections A and C immediately before the relevant section.	It was accepted that more explanation was needed. However, teachers highlighted that students read bullet points more easily than sentences. Moreover, students highlighted the need to avoid making the checklist too long, as well as the importance of brevity in the layout on Google Forms.	The explanation of these sections of the checklist would be given verbally by teachers in two briefing sessions, one a week before, and the other immediately after the speaking activity, according to a prepared script. This would also allow students to ask for clarification of the explanations.	T
An additional statement was suggested to include 'I use complex words' in Section B, 'During the speaking activity'.	The statement showed the student was engaged with the checklist development but the suggestion did not directly relate to the stress, rhythm or intonation of the complex word.	This suggestion fell under the heading of vocabulary and not pronunciation.	S

4.3 Design phase findings

Students and teachers both thought that the concept of pronunciation as specified in the checklist was complete, without unnecessary words or items. Teachers confirmed that "we would usually focus on the intonation and then all of the criteria that you have

mentioned… we do not need like extra questions or extra criteria" (T2), while students commented more generally; for example, "I think about the content is well-chosen, very focused on few specific things about self-assessment" (S2). Teachers expressed concern about whether students would understand or remember the terms; as T1 said, "I know the stress is to be like this for example but the student might think that the stress is different" (T1). The same teacher added "if…we use this regularly every time they want to do any speaking activities, we give them this, maybe they will understand, maybe they can remember enough, in other words". As one student expressed it, "we need to see the bigger picture first" (S8). These comments highlighted the issues of explanation, shared understanding of assessment criteria, and repetition and reminders of criteria, all of which needed to be taken into account in trialling, and later implementing, the checklist.

All participants were happy with the number and order of sections; typical responses were "For the section is nice and perfect" (S6) and "The amount of questions in each section is perfect, not too much, it's just the right amount" (S5), and were supported by several other students. Similarly, students were satisfied with the number of items and their clarity, as illustrated by "the questions are all very straightforward, and simple, compact and just perfect" (S5), although teachers suggested starting each section with an introductory sentence and reducing the words in the following sentences, as described in Table 4-1.

The teachers had not used Google Forms, but the students were quick to suggest the importance of layout, which included minimising the need to scroll up and down, and making sure the amount of information on each page was neither too much nor too little (S3, S7). Consideration of the layout led to the inclusion of sections of the guidance just before the sections of the checklist they referred to (S2, S5), with definitions of linguistic terms at the very beginning (S3).

One student also suggested that administering the checklist during a class would be more effective than sending the link separately, "because some people may get annoyed because there so many questions and just tick, tick, tick and you will not get accurate information" (S6). The researcher and teachers agreed with administration in class from the perspective that this would ensure students completed the checklist.

Turning to existing perceptions of self-assessment, students and teachers were asked to choose one of the following three statements which best represented their personal view of self-assessment.

1. Self-assessment helps students to become independent learners; this is a useful skill that can help them in their careers in the future.

2. Self-assessment may or may not be helpful or necessary, depending on how it is used and if there is enough time to do it.

3. Self-assessment is not necessary; teachers provide all the assessment that students need.

Five of the eight students chose statement 1, while three chose statement 2; no students chose the third option. The main reasons for choosing the first option were that students needed to be independent learners, that they knew themselves and their strengths and weaknesses better than anyone else, and that they needed the skills of self-assessment in their future careers and life. From the perspective of developing independent learning, one stated:

> "It is nice to have their teachers or lecturers to help them doing the self-assessment, but at the same time you know yourself best and you know your weaknesses and you know how to strengthen yourself" (S5).

Some students made a connection between taking responsibility for their own learning and developing their professional competence when they entered the world of work. One student recognised that self-assessment skills could assist with making the transition to

employment and went even further, suggesting that knowing oneself could improve relationships with other people:

> "It's really different, in the classroom and with the outside world. The situation is really different so with the self-assessment that they receive in the classroom they can use in the future to help them to improve their skills towards the other human being outside there. Not only to our teachers, but to others person" (S7).

The three students who chose the second statement explained that they needed lecturers to help them identify what they should improve, and that some students would not understand without help. S1 for example stated:

> "As a student, our job is to learn, and we need guidance to learn and that's where the lecturer's role comes in, you know... because students doing self-assessment may not be accurate because sometimes we don't know where we did wrong" (S1).

Another student recognised the broad spread of ability within a class:

> "Sometimes some students understand and some students not understand because it hard. That's all" (S6).

The spread of ability, knowledge and confidence was exemplified in the contrasting ways in which students referred to the checklist, some repeatedly referring to it as a questionnaire even after being gently corrected, some searching for the correct word, and others confidently referring to it as a checklist.

In contrast to students who commented that they knew themselves better than anyone else, one student commented "I don't [self-assess] during and after because I don't know myself" (S8).

The teachers were divided in their opinions, one choosing the first statement and one choosing the second. However, the teacher who chose the first option had not tried to use self-assessment with students, but instead had "a sort of discussion … when they are listening to my lecture… sometimes we do talk about that" (T2). The other teacher felt that students would not notice what a teacher would notice and always gave feedback

after a speaking assessment to tell students what they had done well and what their mistakes were. These comments informed the initial briefing session that the researcher delivered to all three classes, ensuring a good explanation of the terms 'rhythm', 'stress', and 'intonation'.

Students also valued the explanation of the ICAO scale and its descriptors, as illustrated by S5 and S8:

> "It's really like, amazing, because when you see this checklist, you get to see and you feel that English is actually very important and the fact you have to meet up with this language proficiency rating scale. I understand how much like English is important" (S5).
>
> "It's very thoughtful to have that scale" (S8).

This view was supported by other students who mentioned that they needed to improve their pronunciation in terms of their career in the aviation industry "because it is so dangerous" (S6) if they say a particular word but a work colleague hears it as a different word.

4.4 Changes to Checklist v2.0

This section describes the changes made to Checklist v2.0 in the evaluation phase based on comments received from experts after the calibration phase. It also reports students' and teachers' reactions to the questionnaire, through the semi-structured interviews conducted immediately after trialling Checklist v2.0, and their evaluation of the usefulness of the checklist as elicited through the questionnaires.

As with Checklist v1.0, some suggestions were accepted while others were rejected. There was a suggestion from E3 for the demographic section to include non-binary gender categories (which contradicted the suggestion from E1 in the design phase to exclude gender recognition altogether). In the Malaysian context, only male or female gender recognition is culturally acceptable, and additional categories could have deterred some

students from completing the questionnaire; hence only male and female categories were retained, while the suggestion from E3 was rejected.

E3 mentioned that the item '…I choose words which I can pronounce easily' and the associated guidance make sense if the goal is to get an expert rating for a speaking activity, but commented that many of the words students must learn to say in English are not easy words, and the goal is to learn to say all the words they need in English, "all, easy, middling, or difficult". Although this was a valid point, the checklist asks students to look at their learning strategies and behaviours, so some may choose words that are easily pronounced while others may look up the pronunciation of words which are more difficult to pronounce. BAET students will generally have in mind the goal of getting good marks for their presentation in the final examination, so if repeatedly choosing words which are easy to pronounce results in a student losing marks for vocabulary in a practice assignment, the checklist gives them an alternative learning strategy for dealing with the pronunciation of more difficult words, namely 'I check on the pronunciation of difficult words' (item 3 in section A).

E3 also questioned why the following four items were combined in the guidance: (i) I pronounce the words clearly in English, (ii) I stress the words accurately in English, (iii) I speak English with a regular rhythm, and (iv) I practise speaking English with a natural intonation. The researcher's reponse was that students are expected to practise any or all of these as needed before a speaking assignment, and so the four can be combined, as practising what they need will assist them to self-correct during a speaking activity.

To allow ease of reading the guidance by the students, E4 advised keeping the pattern of items consistent. For example:

Item	Why & How
Before the speaking activity:	
1. I practise my pronunciation.	*Why?* • *You will have more confidence during a speaking activity.* *How?* • *Practise with friends or record yourself.* **corrected to:** Why? • **To develop** confidence during a speaking activity. How? • **By practising** with friends or listening to my recorded speech.

This suggestion was accepted; for ease of reading, all reasons were changed to start with 'to' and all ways of doing it were changed to start with 'by'.

Referring to Section B, the 'during speaking activity' section, E4 asked about the frequency level for the statements. "What if their answer is 'Always'? Can this be 'Almost always or always?'". However, this was considered alongside earlier points about the absolutes of 'always' and 'never' leading to a narrower range of responses and whether changing the frequency scale in this section would introduce the potential for confusion with the scale used in Sections A and C.

Moreover, the measurement scale descriptions for Section B statements were amended to remove mention of the numbers of mistakes as recommended. This avoided students getting distracted with identifying the number of mistakes they have made during the speaking activity.

A summary of expert comments received after the calibration phase, the researcher's response, and their acceptance or rejection is presented in Table 4-2.

Table 4-2 Summary of experts' suggestions, researcher's responses, rationale and changes

Suggestion	Researcher's response	Rationale	Expert
Specific changes			
Accepted			
[In section A] I practise my pronunciation. Why? To develop confidence during a speaking activity. How? By practising with friends or listen to my recorded speech. *Keep the patterns of items consistent.	Consistent wording/word patterns will make it easier for students to read.	Consistent wording reduces load on working memory (Dunham, Lee & Persky, 2020).	E4
[In section B] Counts for the measurement scale descriptions for Section B are problematic when the length of the speaking activity varies. Students are likely to count 5 or 30 regardless of the length of the activity.	The measurement scale descriptions were amended to remove mention of the numbers of mistakes as recommended.	Removal of the counts prevents students from getting distracted with counting the number of mistakes they have made during the speaking activity.	E3
Rejected			
To include in demographic information categories representing non-binary gender categories.	Contradictory suggestion to E1 comments in Design Phase to exclude gender from demographics in the checklist.	In the Malaysian context, which only accepts male or female gender recognition, including non-binary categories could cause rejection of the checklist.	E3
This guidance is quite useful for instructional purposes, and self-directed learning. However, it risks cuing the "right" answers for the self-assessment, making social bias more of a problem than usual.	Many instructional checklists are intended to cue the right answers so that as many students as possible can get things right. Moreover, Sections A and C ask students *how often* they use certain learning strategies, and not if they always use them. Students will need to be reminded of the purpose of the checklist and the associated need for them to be honest with themselves.	The checklist is for instructional purposes only, so the guidance is appropriate.	E3

Suggestion	Researcher's response	Rationale	Expert
The item '…I choose words which I can pronounce easily' in the checklist and the guidance make sense if the goal is to get an expert rating for a speaking activity. But many of the words students must learn to say in English aren't easy words, and the goal is to learn to say them all, easy, middling, or difficult.	As mentioned by E1, 'pronounce easily' is used to encourage weak students to attempt the checklist. This checklist asks students to look at their learning behaviours, so some may choose words that are easily pronounced while others may look up the pronunciation of words which are more difficult to pronounce.	BAET students will generally have in mind the goal of getting good marks for the presentation. Students are also encouraged to acquire the vocabulary they need.	E3
It's not clear to me why these 4 items are combined in the guidance, nor how the answer to Why? is related to each one. I pronounce the words clearly in English. I stress the words accurately in English. I speak English with a regular rhythm. I practise speaking English with a natural intonation.	Students are expected to practise any or all of these before the speaking activity. The 'why' and 'how' (reason and methods) refer to all of them.	Repetition of the 'why' and 'how' for each statement will add to the checklist length and possibly bore students.	E3
[Section B] What if their answer is "Always"? Can this be "Almost always or always?"	The absolute of 'always' can force students who have done something once or twice into a response which does not reflect accurately what they do. 'Almost always' or 'always' could introduce confusion, either in Section B itself, or with the other sections.	Retaining the simpler of the options, 'almost always', may reduce load on working memory (Dunham, Lee & Persky, 2020).	E3
'I try to' is not needed in the statement 'I try to pronounce each and every word clearly in English'. Students either pronounce correctly or they do not.	E1 proposed using 'try to' to encourage weaker students, emphasizing effort rather than outcome; the E1 proposal was accepted.	"Assessment for learning should recognise the full range of achievements of all learners" (Arnold, 2022, p. 4).	E3
'I pay attention to speaking English with a regular rhythm' – why is 'pay attention' included here?	Paying attention is one aspect of metacognition so needs to be retained.	Literature on metacognition refers to paying attention as a prerequisite for regulation of cognition (e.g., Rueda, Moyano & Rico-Picó, 2023).	E3

The changes in the frequency scale for Section B, 'During the speaking activity', were as follows. The items to which the frequency levels refer are:

Evaluative criteria	Frequency level
During the speaking activity:	
9. I am careful when pronouncing the words in English.	1 2 3 4 5
10. I notice my pronunciation mistakes when I am speaking.	1 2 3 4 5
11. I self-correct my pronunciation whenever I mispronounce.	1 2 3 4 5
12. I try to pronounce each and every word clearly in English.	1 2 3 4 5
13. I stress the words accurately in English.	1 2 3 4 5
14. I pay attention to speaking English with a regular rhythm	1 2 3 4 5
15. I speak English with a natural intonation.	1 2 3 4 5

The deleted words are shown as strikethrough in the scale, while replacement words are indicated in bold:

Frequency	Description
1 = Almost never	You are not careful when pronouncing words during a speaking activity. You ~~can make people~~ **attempt to** understand you anyway, maybe by changing words, or gestures or repeating what you say.
2 = Rarely	You **are rarely careful with your pronunciation** ~~attempt to be careful when pronouncing words~~ during a speaking activity~~, but not for all the words. It could be only 5 words in a 15 minute speaking activity~~. The word choice would be based on your preference (e.g., **easiness** ~~difficulty~~ of pronunciation).
3 = Sometimes	You attempt to be careful **with your pronunciation** when pronouncing words during a speaking activity, but not for all the words. The word choice would be based on your preference (e.g., **easiness** ~~difficulty~~ of pronunciation).
4 = Often	You are careful when pronouncing as many words as possible throughout a speaking activity, regardless of how difficult they are to pronounce, although you still make mistakes.
5 = Almost always	You are careful when pronouncing almost all the words throughout a speaking activity regardless of how difficult they are to pronounce.

4.5 Calibration phase findings

The findings from the semi-structured interviews that followed the trialling of the checklist did not lead to any changes to the checklist. However, they revealed teachers' and students' reactions to the checklist after they had experienced using it. Additionally, they shed light on some of the issues that would need to be addressed prior to implementation. Teachers reported that they found the checklist easy to use because it was structured and all the instructions were clear. They stated that the briefing session

was a helpful introduction and that repeating the briefing session on the day the first link (to Section A) was given to students helped to ensure understanding. In fact, they commented that the whole process worked well. The teacher's role would be to look at the checklist together with the students, inform them of the purpose and importance of the checklist, and then use the videos to illustrate the differences in levels in terms of the pronunciation. It was suggested by T2 that the teacher should stay with the students "just in case they have any questions, anything that they don't understand, like for example terms or any words that are that they are not familiar with".

It was suggested by T1 that it would be beneficial to give more examples to students, ideally drawn from students' own mistakes, if they were willing to share them, and that perhaps video clips could be taken from a good student after the first speaking assignment to illustrate correct pronunciation. This suggestion indicates that T1's interest and imagination were stimulated by participating in the study. However, the suggestion raises several issues, one of which is how a teacher would respond if students were unwilling to share their mistakes. If students were reluctant to share mistakes in this way, their resistance could make them more likely to resist the practice and potential benefits of self-assessment. A teacher who was unfamiliar with self-assessment or had objections to it, perhaps because they perceived the teacher as expert, could use students' resistance to justify their own. In practice, it would be much easier and quicker for the teacher to note examples of errors from made by all students during speaking assignments, and to highlight four or five which most affected intelligibility, then drill them several times. The immediate aim of the checklist is to encourage individual students to look up correct pronunciation of the words they need for their speaking assignments, which may differ from one student to another, and to promote feedforward rather than feedback in order to prevent errors from occurring.

Reflecting on using the checklist, T1 suggested peer assessment, practising the correct pronunciation of real-life mistakes, the teacher and students working through the checklist together, and, in any event, giving students more pronunciation practice. However, the checklist as it stands is not designed for peer assessment, although it could be used after completion to initiate discussion with a peer or small group of peers about what strategies they employed and whether those strategies worked. In fact, a number of students indicated that this happened informally after the tryout; they mentioned during the interviews that they had spoken to peers about how they tried to improve their pronunciation. They asked if they had been right to do so, to which the researcher responded that they had done something really useful. The discussions following the tryout among some students about the checklist and process was a sign that completing the checklist stimulated thinking processes; they were thinking about their thinking. Moreover, seeking information from others is a recognized metacognitive strategy that assists individuals to calibrate their self-assessment by providing an additional source of feedback.

It was proposed by T1 that students should be required to complete all sections of the checklist in a single session, either side of the speaking assignment, since students would be likely to forget things. Certainly, one student shared this view:

> "I feel like it's an interesting procedure which however, I, I think the, the space in between the two questionnaires and the interviews is too long, just I forget some of the stuff…as we are all human beings we, we tend to forget sometimes so the work that you do the question, the first question there, then the discussion. After that the second question and then we proceed immediately with the interview. So the, the results will be more accurate" (S12).

Although this could reduce potential benefits, such as being able to change learning strategies while preparing a speaking activity, and allowing a more considered reflection after the activity, the point that students are likely to forget is an important one. A more effective way of helping them to remember could be to ask at the end of the process

whether the strategies they had used in preparing the speaking activity had helped them during the activity itself, and how well they thought their strategies had worked when they reviewed them; questioning them in this way would also help to reinforce their learning. Both T1 and T2 commented that students would need more reminders, as well as repeated use of the checklist, to help them remember, which indicates that they were considering how best to use the pedagogical tool. Before implementation, it would be essential for teachers to meet as a team to discuss and reach agreement on how to effectively employ the checklist in the classroom.

Six students identified specific impacts of the checklist, notably an awareness of the metacognitive self-assessment cycle. Referring to the overall effect of the before, during and after stages of the cycle, students stated the process had made them re-assess their rhythm and intonation (S3), see their mispronounced words and correct them in the future (S11). It helped them know more about their pronunciation, and made one student realise he only reviewed his speaking activity on occasions when it was video recorded (S8). Three students identified that, for them, some parts of the cycle were more challenging than others. They specifically referred to Section B, 'During the activity', including one who could be said to have had a metacognitive experience, stating:

> "I can connect the questions with the forum that I presented the other day very well. And then after the forum I had to […] answer some of the second part right? So yeah, […] and while answering the second part, I realized it. I realized more, more it opens up. More things to me that that that I have to focus on, that I didn't realize when I was answering the first part" (S2).

The other two students who highlighted Section B expressed difficulty in trying to assess themselves while they were in the middle of speaking, reporting that it was a challenge to find the right balance of focus; but as one of them said "I think it's [Section B, 'during the speaking activity']…it's a great part to be there" (S5). Some students found themselves thinking, possibly for the first time, about what strategies they used to find

out how to pronounce a word and thus adding to their knowledge of cognition, as illustrated by a description of completing Part A, 'Before the speaking activity':

> "realizing how I have to take some time, trying to think back, like what do I do if I get some words I don't know how to pronounce, or I don't understand words, do I even look it up, maybe, I'm not sure" (S10).

Two students recognised there were times when they needed to correct mistakes they made in the speaking assignment by afterwards returning to the dictionary to listen to the pronunciation of more difficult words or to syllable stress (S2, S8). Identifying some pronunciation problems as a result of completing the checklist led three other students to say they would seek help from teachers or peers. Seeking help is considered a metacognitive strategy; this suggests another way in which the checklist supported individual understanding of self-assessment.

In contrast, one student reported needing more time to understand the self-assessment process (S10), while several expressed the view that some other students might find it difficult to understand the process or complete the checklist. Two students stated how they felt about the experience, one saying they felt very good because they could see where they could improve in the future (S3). However, the other student said they felt relieved when the process ended (S14), which suggests that anxiety may be easily aroused in students who are working towards external high-stakes qualifications and who therefore see the self-assessment process as irrelevant to their immediate needs.

Students and teachers alike found the briefing session and video clips helpful, T2 commenting that they helped the teacher. "Even me as a lecturer I mean I, I get to see like clearly you know how to explain to my students clearly" (T2). Many of the students were more aware of the effect of incorrect rhythm and intonation on intelligibility, as well as, in their words, the importance of English and the ICAO requirements. One student expressed it thus:

> "Basically, from the video, [students] pay more attention and they will automatically compare their speaking and this speaking from the video as well" (S13).

Three of the ten students commented on their level while talking about the video clips, a reaction which is perhaps unavoidable when criteria or standards are presented to them and when they are aware they are working towards a high-stakes assessment. Their reaction raises an interesting question about how teachers deal with this in the context of formative assessment.

Some students missed the briefing sessions and did not watch the video clips, or only watched small parts of them. One student (S11) was of the opinion that only a native speaker could achieve the highest level on the ICAO scale and a new clip was therefore needed for that level, despite the fact that it had been explained that in the aviation industry, the important thing was to understand and be understood by other non-native English speakers. The same participant interpreted ICAO level 5 pronunciation as having a very good vocabulary and grammar, someone who often communicates in English at that level, and ICAO level 4 as being understood by others although perhaps communicating half the time in their L1. Two students equated ICAO levels with speaker accent or nationality, one saying:

> "I can differentiate level 6 and level 5 because when I watch level 5 it was Russian people and automatically I can set in my mind oh the Russian people pronounce this word this way, it should be pronounced that way if the people was an English people, yeah. So yeah it's something like that, just as slightly different between the level 5 and level 6 speakers" (S11).

Opinions were offered about Europeans being better at languages and Malaysians being unable to reach those standards, despite the fact that some students in the year group had studied in the United States and were fluent in American English, a fact that was recognised by some participants.

These comments implies that more time may be needed in the briefing session to discuss each of the video clips more fully in relation to stress, rhythm and intonation, and to find other opportunities to remind students of what the criteria mean.

The ten student interviewees and their teachers found all sections were equally easy to administer and understand, although it was noted that "It's not that difficult to fill [fill in the checklist], but well, you need to, you need to think to fill" (S14). One student also commented that although his own classmates could all understand easily, there might occasionally be a student who "had difficulties to understand a question in English, and [who] might need the teachers or lecturers to explain them or give them example in Malays [sic]" (S11). The same student also mentioned that it was important to feel self-confident about doing the self-assessment, "then you may do the assessment much better". In the light of these last two comments, this student may have needed more reassurance about the purpose of the checklist and how the results would be used by the students themselves. Remarks such as these, like the impromptu discussions among students mentioned earlier, demonstrate the potential in the classroom for exploring and negotiating the meaning of self-assessment, and for providing another source of information which students can use to calibrate their self-feedback.

Regarding practical considerations of access to and use of the checklist, there were no problems accessing the Google Forms Checklist v2.0 during the trial. The researcher checked the links were working and created new links, if necessary, before each session. There were, however, problems with the internet on several occasions, as well as some issues of non-compatibility with mobile devices.

Interviewees were also asked how long the checklist took to complete; on average, 30 minutes was more than enough, although some students required more time. Teachers were slightly concerned about the time needed for students to complete the checklist,

maybe 30 minutes in total; this indicates the pressure they experienced to cover the prescribed syllabus in a limited amount of time. Teachers emphasised the need to ensure that students already fully understood what to do before completing the checklist, so that both parts could be administered within a single session. As mentioned earlier, administration in a single session may not be the most effective way to encourage feedforward, highlighting once again the importance of discussing and reaching agreement on how to implement the checklist.

Although the calibration phase interviews did not lead to any changes to the checklist, they highlighted a number of important issues regarding differences in students' understanding of the criteria and of the concept and process of self-assessment which would be important for implementation.

4.6 Inspection of a sample of checklists

Before the beginning of the evaluation phase, while awaiting expert validation of the evaluation of usefulness questionnaire, the researcher took the opportunity to examine the responses in a sample of completed checklists to see whether students had used the full range of the frequency scale or whether they had avoided the endpoints. Every fifth checklist was examined in this way, ten in total. None of the selected checklists showed responses clustering in the middle three scale points. This not only reassured the researcher regarding the choice of endpoints as 'almost always' and 'almost never' but also provided a degree of reassurance that checklists were likely to have been completed in a thoughtful way. Four or five points on the frequency scale were used in all ten checklists, suggesting that students were responding thoughtfully rather than simply ticking the central boxes. Three students used the full range, while the remaining seven used four of the five points. The ranges of responses and number of students are summarised in Table 4-3.

Table 4-3 Response ranges in sampled checklists

RANGE OF RESPONSES	NUMBER OF STUDENTS
From 1 to 5	3
From 1 to 4	3
From 2 to 5	4

The spread of responses within the ranges was also examined, as summarised in Tables 4-4 to 4-6. As expected, the great majority of responses fell into the 'sometimes' and 'often' categories. However, there was no obvious pattern of boxes being ticked semi-automatically, as shown in the tables. Two interesting points in particular can be seen in responses to Section A, 'Before the speaking activity' (Table 4-4): all but one of the students indicated they chose words which were easy to pronounce 'almost always' or 'often'. In contrast, all but one 'almost never' or 'rarely' referred to the ICAO scale as guidance.

Table 4-4 Spread of sample responses in checklist Section A

STRATEGIES SECTION A, 'BEFORE'	FREQUENCY*					TOTALS
	1	2	3	4	5	
	Number of students					
1. I practise my pronunciation.	0	1	4	3	2	10
2. I choose words which I can pronounce easily.	0	1	0	6	3	10
3. I check on the pronunciation of difficult words.	0	2	4	3	1	10
4. I pronounce the words clearly in English.	0	0	5	4	1	10
5. I stress the words accurately in English.	0	1	5	4	0	10
6. I speak English with a regular rhythm.	0	0	6	4	0	10
7. I practise speaking English with a natural intonation.	0	1	3	6	0	10
8. I refer to the ICAO Language Proficiency Rating Scale (LPRS) as guidance for my pronunciation for my speaking activities.	4	5	0	1	0	10
TOTAL RESPONSES	4	11	27	31	7	80

*Key: 1 almost never; 2 rarely; 3 sometimes; 4 often; 5 almost always

A similar pattern was seen in responses to Section B, 'During the speaking activity', as shown in Table 4-5. Careful pronunciation of whole words appears to be the most commonly used strategy in the sample, and more than half indicated they noticed and self-corrected their mistakes while speaking.

Table 4-5 Spread of sample responses in checklist Section B

STRATEGIES SECTION B, 'DURING'	FREQUENCY					TOTALS
	1	2	3	4	5	
	Number of students					
9. I am careful when pronouncing words in English.	0	0	1	9	0	10
10. I notice my pronunciation mistakes when I am speaking.	0	0	2	7	1	10
11. I self-correct my pronunciation whenever I mispronounce.	0	0	3	6	1	10
12. I try to pronounce each and every word clearly in English.	0	0	0	8	2	10
13. I stress the words accurately in English.	0	0	6	4	0	10
14. I pay attention to speaking English with a regular rhythm.	0	0	7	3	0	10
15. I speak English with a natural intonation.	0	0	5	5	0	10
TOTAL RESPONSES	0	0	24	42	4	70

Key: 1 almost never; 2 rarely; 3 sometimes; 4 often; 5 almost always

A similar pattern was also seen in responses to Section C (Table 4-6), although responses were more evenly spread across 'rarely', 'sometimes' and 'often', indicating that some students spent less time reviewing their performance after the task than they did preparing for or concentrating during a speaking task.

Table 4-6 Spread of sample responses in checklist Section C

STRATEGIES SECTION C, 'AFTER'	FREQUENCY					TOTALS
	1	2	3	4	5	
	Number of students					
16. I reviewed the recording of my speaking activity for self-improvement	1	3	4	2	0	10
17. I listed down the words I mispronounced.	2	5	2	1	0	10
18. I took note of the words that I stressed inaccurately in English	2	1	4	3	0	10
19. I took note of where I spoke with the wrong rhythm.	0	6	2	2	0	10
20. I took note of where my intonation caused problems for my listeners.	0	1	4	4	1	10
21. I listened to correct examples of pronunciation in English	0	0	3	4	3	10
22. I practised speaking correctly after listening to examples of pronunciation in English.	0	0	3	4	3	10
TOTAL RESPONSES	5	16	22	20	7	70

Key: 1 almost never; 2 rarely; 3 sometimes; 4 often; 5 almost always

At this stage, the inspection of responses gave no more than brief insights into how often students employed particular strategies, but there was sufficient variability in the numbers of responses along the frequency scale to suggest that students were giving considered responses to the items.

4.7 Evaluation phase findings

A total of 25 valid completed evaluation of usefulness questionnaires were returned by students to the researcher. Both teachers also returned a valid completed questionnaire. All responses fell into the moderately useful category (5 student responses) or very useful category (20 student responses and 2 teacher responses), as shown in Table 4-3.

Table 4-3. Evaluation of usefulness by category and number of students and teachers

Category of usefulness	Number of students	Number of teachers
Very useful	20	2
Moderately useful	5	0
Not useful	0	0

An open-ended question at the end of the questionnaire invited suggestions for improving the checklist. Both teachers stated no changes were necessary, one saying "I think the checklist is acceptable, clear and well-organized. Hence, no improvement needed" and the other teacher was of a similar view: "I think there are no changes needed for this checklist".

Responses from students were more directed to speaking activities and to the teaching of pronunciation than to the checklist itself. These were: "Provide phonetics on how to pronounce difficult words including the meaning for the context", "Include more terms for example AMM, CAAM, UNiKL etc.", "More Aviation terms", "Maybe an emphasis on types of pronunciations for example British and American styles", and "More details and specific about how to pronounce and present in the speaking activity", "Give more

exercises on pronunciation", "Practice students to reduce monotone", and three more suggestions that more practice speaking should be included in the classroom.

Further suggestions for use in the checklist itself were: "Can have more reference materials in case some students may not understand the concept", "Provide more videos on how to pronounce certain things" and "Add more reference materials such as video clips to each of the category". One student wanted use of colour to highlight which points were more important.

Suggestions directed at improving the introduction and use of the checklist included "Make sure we can properly understand the contents of the checklist", meaning that extra time and care should be provided for students who need this. In the light of comments in the design and calibration phases that the checklist items were clear and easy to understand, proper understanding of the contents must mean proper understanding of self-assessment.

Students' suggestions also indicated that further thought was required to embed pronunciation teaching as well as the checklist into the Aviation English curriculum. This point is picked up in the discussion chapter.

4.8 Summary

This chapter has narrated the development of the checklist through to its third version. Overall, the teachers found the checklist to be very useful, as did 20 of the 25 students who completed valid questionnaires. Their range of views has been reported, together with their ideas for further improvements as the checklist was implemented. In at least some cases, student remarks provided evidence that they understood and were able to implement the self-assessment cycle, and there was also some evidence of both knowledge and regulation of cognition. Student understanding of the self-assessment

process could be strengthened by adopting the expressed need for more briefing and practice.

5 CHAPTER 5 DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This final chapter summarises the development of the checklist and responses to the research questions before discussing the results in the contexts of self-assessment in the classroom and development research. The purpose of the checklist was to develop a pedagogical tool for ESL BAET students to enable them to develop their understanding of how to self-assess their pronunciation. Specifically, it aimed to encourage them to take ownership of their learning in terms of examining their use of learning strategies and reflecting on whether those strategies were working in the sense that they were helping students to improve their pronunciation.

5.2 Summary of the study

The purpose of the checklist in this present study was to develop a pedagogical tool for ESL BAET students to enable them to develop their understanding of how to self-assess their pronunciation. Specifically, it aimed to encourage them to take ownership of their learning in terms of examining their use of learning strategies and whether those strategies were working in terms of helping them improve their pronunciation. The checklist was designed, developed and validated through a developmental research process consisting of three stages, namely design, calibration, and evaluation.

In the design stage, the checklist was inspired by the CEFR self-assessment grid (Council of Europe, https://coe.int, n.d.), the use of a Likert-type frequency scale in SILL (Oxford, 1990), the Metacognitive Awareness Listening Questionnaire (Vandergrift et al., 2006), and an argumentative writing self-assessment checklist (Nimehchisalem et al., 2014). The checklist drew on a review of literature in the fields of pronunciation, self-assessment, metacognition and the theory underpinning the approach to self-assessment. Adaptations of the original models included selection of the construct of pronunciation to

be used and items appropriate to pronunciation rather than writing, speaking or all aspects of language learning. An extended guide was developed to separate the self-assessment of the language learning strategies used by students from the explanations of why and how the learning strategies could assist them to improve their pronunciation.

The checklist was sent to two experts in the field of ICAO rating scale assessment, after which the wording of one item and instructions to students was changed. A decision was made to retain the ICAO scale and share with students an understanding of what they were working towards. The use of the present tense in the section 'During the speaking activity' was retained in agreement with examples from the literature (Oxford, 1990; Vandergrift et al., 2006). Teachers and a sample of students (n=8) were then interviewed about the clarity and relevance of checklist items, and the qualitative interview data were used to remove unnecessary repetition and reorder some of the items. Parts of the extended guide were incorporated into the checklist at the start of the checklist section to which they referred and recommendations from students regarding Google Forms layout were used to develop an online form of the checklist (Checklist v2.0) for trialling.

In the calibration phase, a briefing session was presented by the researcher to teachers and students, introducing the purpose of the checklist in addition to the terms needed to understand the items (syllable, stress, rhythm and intonation) and the ICAO scale descriptors illustrated by video clips. The video clip selection was carried out by the researcher together with a colleague who assessed the clips independently before agreeing the final selection. The checklist was administered to students (n=50), after which semi-structured interviews were conducted with teachers (n=2) and a sample of students (n=10) to elicit their reactions to the self-assessment checklist and process. There was consensus among interviewees that all sections were equally easy to administer and understand, with some students commenting that they needed time to think about what strategies they used while completing the checklist.

In the evaluation phase, students and teachers completed an evaluation of usefulness questionnaire which assessed usefulness in terms of reliability, validity, impact and practicality. After expert validation, the questionnaire was piloted with students (n=8) and Cronbach's alpha was calculated to be α =.963. The questionnaire was administered to the remaining students (n=42) and the returned valid questionnaires (n=25) indicated that students found the checklist very or moderately useful. Further minor amendments were made to the guidance in the checklist following additional expert validation.

5.3 Summary of the findings

The first research question asked *What criteria should be used to design the student pronunciation self-assessment checklist?*

The criteria used to design the checklist were based on the literature review, together with general principles for overall design such as clarity of wording and layout. Criteria based on the literature review were taken from the appropriate construct of pronunciation to be used, the underpinning theory and operationalisation of self-assessment, and elements of metacognition and associated learning strategies.

The selected concept of pronunciation was adopted from the pronunciation subscale of ICAO descriptors: pronunciation, stress, rhythm, and intonation (ICAO, 2004). A developmental and formative approach was taken to self-assessment, with the aim of assisting students to experience and understand some of the thinking processes involved in this type of self-assessment. The underpinning theory of metacognition was used to structure the checklist. The checklist sections were organised according to three widely recognised stages of the metacognitive self-assessment cycle: planning (before), monitoring (during) and evaluation (after the activity). Individual items were adapted from SILL (Oxford, 1990) and the Metacognitive Awareness Listening Questionnaire (Vandergrift et al., 2006), among other sources. The clarity and feasibility of the checklist

were assured through expert validation, semi-structured interviews with teachers and a structured interview with a group of students.

The second research question asked *What are students' and teachers' reactions to the pronunciation self-assessment checklist?*

Regarding students' and teachers' reactions to the pronunciation self-assessment checklist after they had experienced using it, the interviews at the end of the calibration phase established the clarity and feasibility of all sections of the checklist, as well as the briefing session with video clips, from the end users' viewpoints. Participants were also asked about the suitability of the checklist as a classroom learning tool, along with any other comments they wished to make about the actual experience. All the students interviewed (n=10) found that the checklist was easy to access and complete, although one found the thinking involved was challenging. The teachers (n=2) and the researcher found the checklist easy to administer but the teachers expressed some concerns about whether students would remember, and in some cases understand, what it was about. Several students and both teachers thought that all sections of the checklist should be made available in a single session, immediately before and then immediately after the speaking assignment. Administering all sections in one session would minimise the possibility of students forgetting from one week to the next what they had done, the content of the section 'Before the speaking activity' and their responses about how often they employed particular learning strategies.

In the researcher's view, this rather contradicted the idea of reflection during the preparation stage on the learning strategies students used while preparing a speaking assignment. This indicated that these students had not fully understood the purpose of the self-assessment, although they had grasped the importance of pronunciation and of understanding the checklist. Several students and both teachers expressed the need to go

through the briefing and explain the checklist at least twice before students actually completed it. The teachers thought the checklist could be a useful additional tool because the structure and instructions were clear, but were still considering how best it could be implemented.

The third research question asked *How do students and teachers evaluate the usefulness in terms of impact, practicality, reliability and validity of the pronunciation self-assessment checklist?*

The results of the evaluation of usefulness questionnaires in terms of the impact, practicality, reliability and validity of the checklist showed that students and teachers found the questionnaire to be moderately useful (n=5) or very useful (n=20). Moreover, the majority of students who were involved in piloting the questionnaire (n=6) also found the checklist to be very useful. One of the teachers proposed that it could be worth developing similar checklists for other areas of English, initially other areas of speaking such as fluency, despite having expressed concerns about the time required for checklist completion.

Suggested improvements related more to the teaching of pronunciation and having more opportunities to practise speaking than to the checklist itself. It can be inferred from this that the checklist certainly raised awareness of the importance of pronunciation and stimulated students' thinking about their learning needs.

5.4 Discussion of findings

This section discusses issues that emerged in the design and development stages, the issue of classroom self-assessment more generally, the practicality of developing checklists, and the implications for implementation of the checklist.

5.4.1 Design stage issues

Two issues arose in the design stage, namely whether the ICAO scale descriptors were suitable for use with students, and whether the technical terms such as rhythm and intonation should be explained in words in the checklist. These issues were underpinned by assumptions that a checklist should be self-explanatory and stand alone and, to a lesser extent, that the ICAO scale descriptors were an element of the teachers' and assessors' expertise. Checklists which appear to stand alone, such as aviation checklists, list each step to be taken to ensure safety. However, the pilots or AMTs who use them have years of training, knowledge and skill, so the checklists do not really stand alone. Moreover, airline safety is also dependent on engineering design and production quality, indicating that a wider team and more extensive process of teaching and learning underlie the checklists. In the very different context of heart surgery, checklists alone have not prevented adverse events, and changes needed at the system level have been identified (Raman et al., 2016). In the context of teaching ESL writing, the argumentative writing checklist was accompanied by an extended guide which explained the reasons and methods for carrying out each of the main 23 items (Nimehchisalem et al., 2014, pp. 75-80). However, it is logical when explaining stress, rhythm or intonation in spoken English for the teacher to use exemplars, modelling and exercises to fully convey the meaning of the terms, rather than to try to give written explanations.

The second issue of using descriptors from the ICAO high-stakes test scale is interesting. It is argued that teaching to high-stakes tests has an adverse washback effect on students (Tsagari & Cheng, 2017). The negative washback is attributed to various causes which include heightened student anxiety, teachers teaching to the test as a result of which learner-centred learning increasingly gives way to teacher-centred teaching, and learning becomes more focussed on memorisation (Au, 2007; Zhao et al., 2016). However, it has been pointed out that the washback effect is not simple but results from a combination of

individual and external factors, including the influence of family members and institutional expectations, among others (Dawadi, 2021; Shih, 2007). It has been proposed that high-stakes testing can have a positive washback effect and there are ways in which test developers and teachers can increase the positive effect (Cheng et al., 2015). Quite separately from the debate surrounding the washback effect, it can be argued that students have a right to know the standards against which they will be measured or which they will be expected to achieve. Their ability to meet specified standards may determine whether they can reach their goals in life, whether their goals are realistic, and, by implication, how much effort may be involved. In the context of language learning, goal-setting is important to encourage intrinsic as well as extrinsic motivation (Rose et al., 2019, p. 895). As Moss et al. (2021) express it in the title of their report, this is about high standards rather than high stakes. Students may set small step-by-step goals according to their current knowledge and ability but knowing the ultimate goals of their learning can assist them in setting the smaller goals which lead students in the right direction. Teachers and assessors may be reluctant to give up their role as decision-maker on behalf of the student, perhaps because knowledge is power, or perhaps because they are concerned about the changes and uncertainties in their future role. T1 indicated this by stating that students did not understand like the teacher did before quickly adding that it would be the teacher's role to help them understand.

The resistance to sharing high stakes standards may be linked to the longstanding criticism of self-assessment in terms of its lack of reliability, which arguably limits its usefulness in terms of assisting student achievement (Brown & Harris, 2013). Reliability of self-assessment is typically measured in terms of comparisons between, for example, teachers' and students' judgements or between the results of two versions of a test. Reliability was not examined in the present study although it could be examined in the future by comparing students' judgements of their level with teachers' judgements, for

instance after training teachers to rate pronunciation followed by training students to rate in the same way, based on teacher-student discussions of descriptors and rating of exemplar video clips.

Brown and Harris (2014) recognise that self-assessment is an important life skill (competence) (p. 23) and have proposed that problems of reliability can be avoided if students participate in designing criteria at an appropriate stage in a self-assessment curriculum that begins early in school life and gradually builds towards accurate self-assessment. However, this is not practicable in many courses or classrooms, raising the question of whether self-assessment should then be employed at all. The position of Andrade (2019) and Isbell and Sakai (2022) is that self-assessment is a reality and will not disappear. As the majority of the students in the structured group interview mentioned, self-assessment is important and necessary for their future lives, and therefore teachers should address it in one way or another. Making sure that students properly understand assessment criteria is a necessary first step in self-assessment and an important part of the teacher's role in scaffolding learning (Panadero et al., 2016b).

5.4.2 Calibration stage issues

Following on from the design stage issues, the issues that emerged from the calibration stage concerned students' understanding of self-assessment and the criteria. The teachers' perceived need to explain the checklist at least twice before students actually completed it may be attributable to their familiarity with the transmission model of teaching. Other techniques, such as questioning students as well as inviting and promoting discussion, could help individual students to pinpoint areas of difficulty in understanding. It was also noticeable that some students interpreted the criteria in terms of accent or nationality, while at least one student missed the briefing session and did not access the video clips from the guidance. All students need to understand and internalise the pronunciation criteria in order to benefit from the self-assessment. Additionally, the

teachers and some of the students seemed more concerned with the least amount of time needed to complete the checklist than with extracting maximum value from the exercise, although both teachers then made creative suggestions about how to find more classroom time to exploit the checklist and its benefits to the full. It was evident that more than half of the students interviewed did not link their experience of completing the checklist with future improvement in their pronunciation. Students did not have the opportunity to complete the feedback cycle by preparing for a second speaking activity, and, because the actual speaking activity followed too soon after the preparation for it, students lacked the opportunity to generate their own feedforward. This is not a major failing of the present checklist design and development study which focused on the instrument, but is a key concern for implementation, as discussed in section 5.4.6. Carless and Boud (2018) and Yan and Carless (2022) highlight that self-assessment requires individuals to be able to make sense of and use the information from the process to adjust and improve their learning strategies. Teachers need to know that students have the skills to do this, and to be able to model the process for students to be able to understand how to do it before trying it for themselves.

5.4.3 Evaluation stage issues

Although the majority of the students who returned valid completed questionnaires reported that the checklist was very useful, some had not grasped the underlying purpose of self-assessment, in the sense that they recommended ways of improving the teaching of pronunciation but had not used independent learning to try to fill the gaps they identified. Moreover, the teachers had no previous experience of self-assessment, either as teachers or learners, and had only recently started to make the transition towards a more constructivist approach to teaching. This could make it more difficult for them to make students aware of the value of self-assessment, without which students could perceive self-assessment as an extra and unimportant task (Jamrus & Razali, 2019, p. 66).

These issues raise concerns regarding implementation. Isbell and Sakai (2022) warn that "forms of assessment are not menu items to be picked at the spur of a moment" (p. 199). Students need a clear understanding of why self-assessment is being introduced, along with repeated opportunities to be able to monitor their progress towards meeting clear criteria (Boud, 1995). Without practice and support, they are less likely to develop the skills needed to be able to assess themselves (Boud, 1995). Thus, the first step in implementation would need to be professional development for teachers, not only a seminar but also personal experience of the challenges as well as the benefits of self-assessment, in order for teachers to provide the necessary support. There are a number of self-assessment instruments suitable for English language teachers to use in the context of a professional development intervention, such as the European Profiling Grid (https://egrid.epg-project.eu/) or the British Council self-assessment tool for teachers contained in the annex to Borg and Edmett (2018).

5.4.4 Self-assessment in the classroom

There is strong support for student self-assessment in the classroom (Leahy et al., 2005; Andrade, 2019). Self-assessment offers several benefits for teachers; it can save time and is flexible (Isbell & Sakai, 2022, p. 207). The checklist provides opportunities for students to reflect on the strategies they use to check and produce correct pronunciation, and how well those strategies work in practice. Thus, it provides a first step in enabling students to assess their progress towards a particular standard, in agreement with Boud and Falchikov (2006). As a pedagogical tool, the checklist did not require students to award themselves a grade and thus avoided the criticisms and difficulties associated with the accuracy and reliability of self-assessment. Moreover, if the checklist succeeded in its aim of making students more aware of whether their existing pronunciation learning strategies were achieving the outcomes they desired, it is probable that at least some of them would change their pattern of strategy use. For example, the data in Tables 4-4 to

4-6 show that completion of a full cycle, or repeated cycles might have led to fewer students choosing easy words, while more students might have reviewed their speaking activity more frequently. Differences such as these indicate progress; they are not a sign of unreliability. At the same time, the checklist did succeed in raising awareness of the importance of pronunciation among most of the students. Raising awareness in this way can in turn lead students to modify their judgement of their ability (Isbell & Sakai, 2022).

Additionally, there are several ways in which the checklist could be used as feedforward practice, in line with the types of practice identified in the literature review by Sadler et al. (2023). The briefing before the checklist explains the ICAO descriptors for pronunciation and provides video clips illustrating the various levels of the scale as well as clarity with reference to understandings of whole word pronunciation, stress, rhythm and intonation. Aviation English teachers at the research site usually give students a formative assessment opportunity to practise a speaking activity before summative assessment of speaking assignments takes place. They tend to form an overall impression of speaking, including content of presentation and fluency, and provide selective feedback. When teachers are clarifying the task requirements for the formative assessment, they could introduce part of the briefing session to encourage students to use the vocabulary they need (rather than choosing words they can pronounce easily) and look up pronunciation if necessary. They could deliver the whole briefing session and administer the first section of the checklist, 'Before the speaking activity', for students to monitor the pronunciation aspect of their preparation. Moreover, the whole checklist could be used as formative self-assessment, followed by students noting what they will do differently, if appropriate, when preparing for the summative assessment. In a review of the literature on feedforward, Sadler, Reimann and Sambell (2023) observe that the aspirations and claims of feedforward found in the literature is not matched by practical examples of how teachers support the process (p. 315). Training students in how they

self-assess, giving them feedback on their use of self-assessment processes, and scaffolding their learning about self-assessment is needed to support students' effective use of self-assessment (Panadero & Alonso-Tapia, 2013; Panadero et al., 2016b; Panadero et al., 2019). It is important to maximise the appropriate use of a pedagogical tool when considerable time and effort has been invested in creating and validating it.

5.4.5 The practicality of designing and developing checklists

There is a balance to be found between the amount of time needed to develop and validate a pedagogical tool and the amount of time a teacher might have available for such development. The checklist in this present study was developed because the Aviation English teaching team had previously carried out an extensive search to find either a suitable assessment or self-assessment tool without success. The opportunity to develop a self-assessment instrument as part of a research project occurred at around the same time as interest in self-assessment was increasing at the research site. Additionally, the research site had moved to blended learning because of the COVID-19 pandemic and then to problem-based learning; these approaches require students to have many of the same skills as self-assessment. One of the teachers who participated in the study wondered if it would be possible to adapt the checklist to fluency and other aspects of English, in view of the time already invested and the perceived usefulness of the end product. Such possibilities are likely to depend on how the pronunciation self-assessment checklist is implemented.

5.4.6 Implementation issues

Issues of student understanding of the process and the criteria, together with the thinking skills they need to be able to close the feedback loop as well as to effectively feedforward to their next speaking assignment, have been highlighted. These issues suggest that a further study is required before implementation to compare the ways in which they could be addressed, not simply considering the issues highlighted in section 5.4.4 but using

research to provide an evidence base for the decision taken. The checklist in this present study was developed through research but its implementation in the classroom has yet to be addressed. A number of the key steps identified by Panadero et al. (2016b) remain to be planned, namely: teaching students how to apply the criteria, for example by modelling; providing feedback on how the students have understood and completed their self-assessment, and assisting them to use what they learned to make improvements, in addition to ensuring enough time is allowed afterwards for reflection and revision of work. It is vitally important for students to gain useful feedback from external sources as well as themselves in order to develop their self-assessment skills. This could be done in a number of ways, for example feedback from the teacher or peer discussion, and it would be useful to know which was the most likely to bring most benefit to the students, as suggested in the section on future directions for research.

5.5 Self-assessment of pronunciation and English for Specific Purposes

The pronunciation self-assessment checklist may be useful in some of the teaching and learning situations in ESP which were described in section 1.3. Whether ESP courses are delivered separately before students start a subject-specific academic course or whether they are integrated into the academic course, they are likely to focus primarily on the skills needed for academic study and the specialised vocabulary and topics related to particular occupations. The skills involved are mainly reading, writing, and listening, although speaking activities relevant to the academic may be included, such as presentations and job interviews. Students undertaking ESP courses are likely to have a wide range of English speaking proficiency, with individual needs to improve their pronunciation so that they can be confident when speaking with lecturers and colleagues. The present checklist has the potential to enable ESP teachers and lecturers to provide their students with additional support for their pronunciation, if they are able to provide

a series of links to online resources. As Feak (2014) stated, the speaker is responsible for making sure their speech is understandable to listeners (p. 43), and the checklist can assist them to take the responsibility. Whilst comprehensibility from the listener's perspective requires other elements of speech to be present, such as vocabulary, grammar and fluency, the first essential in communication is clear pronunciation.

5.6 Comparison of checklist designs

In the absence of pedagogical tools directly related to pronunciation self-assessment, three checklists were chosen for the comparison of designs. Two were intended for use by students, one related to advanced speaking skills which included a section on pronunciation (Jankowska & Zielińska, 2015), and one related to all four English language skills (University of Auckland, n.d.). The third was a Self-Assessment Tool (SAT) intended for teachers to identify their competence and professional development needs and which was evaluated by Borg and Edmett (2019). The designs of the checklists reflect their different purposes.

Jankowska and Zielińska (2015) based their self-assessment instrument on the CEFR descriptors for C1 which was the level students were expected to attain, the guidelines for the college's own English oral examination, and their own observations and experience (p. 258). Drawing inspiration from different sources including external standards and their own experience was similar to the approach taken by the researcher in the present study. They adopted an action research methodology (p. 255).

The checklist, reproduced in Figure 5-1, shows they used an introductory phrase to cover all the items, similar to the design decision in the present study. It is interesting they chose 'In my speech I managed to…' rather than 'I can…', allowing weaker students to feel more included in the process. Pronunciation was addressed in items 7 to 12, although they included fluency and specific segmentals known to cause difficulty for their Polish-

speaking students. Students were required to grade themselves on a 5-point scale with options such as 'very good' and 'quite good', and to support their grading with comments and examples. They were also invited to comment on the usefulness of the checklist itself in terms of their own view of whether it was "very useful", "useful", "not useful", "not useful at all" or "cannot say" or write a free response about their reactions to the checklist (p. 259).

Table 1 Self-assessment checklist

	In my speech I managed to…	5	4	3	2	1	Comments/ examples
1.	Clearly present complex ideas						
2.	Adequately describe experience						
3.	Follow a logical order of events						
4.	Emphasize important arguments						
5.	Give appropriate examples						
6.	Draw conclusions						
7.	Speak fluently						
8.	Use appropriate intonation						
9.	Pronounce "th" correctly						
10.	Pronounce final voiced consonants correctly						
11.	Pronounce vowels correctly						
12.	Use correct stress in words						
13.	Use advanced vocabulary						
14.	Vary sentence structures						
15.	Form grammatically correct sentences						
16.	Use correct verb forms						
17.	Use correct articles						

Figure 5-A Speaking self-assessment checklist Jankowska and Zielińska (2015)

Source: Jankowska and Zielińska (2015), p. 260

The benefits of this particular design include: a focus on aspects of speaking presentations known to be problematic, and brevity. One of the disadvantages is that reflection only takes place after the event, when it is too late to improve the speaking activity before delivery.

Data collected from interviews with 15 students after they had completed the checklist led the authors to identify a number of issues, some related directly to design and others related to the implementation of self-assessment. The design issues were modification

and clarification of some of the items. Whilst in action research it is conventional to pass through several cycles of tryout and review, the clarity of items could have benefited from being assured at an earlier stage, as in the present study. Jankowska and Zielińska (2015) also concluded that students should be involved in the design process in order for them to feel a sense of ownership, a point that was addressed in the present study within similar constraints of external standards (p. 264). Similarly, their point concerning students' choice of aspects to focus on the target language performance they want to focus on was addressed in the checklist design in the present study because students had an opportunity at the design stage to amend, remove or add items.

It is perhaps surprising that, in view of the history of difficulties the participants had with self-assessment, the checklist focused on giving an assessment of grade rather than an examination of how they self-assessed themselves and whether there were alternatives to transcription which some students disliked intensely because of the time needed to do it. The need to provide enough training in skills of self-assessment was highlighted as was the need to address the issue of differences between students' and teachers' gradings (p. 264.).

The University of Auckland English language self-assessment checklist was designed for use by international students who might wish to improve their language by signposting themselves to the English enrichment provision at the university. The authors based the checklist on Dunworth and Briguglio (2011); it is advisable to adapt an existing instrument if possible because doing so builds on previous work. The University of Auckland checklist contained seven sections on vocabulary, grammar, pronunciation, reading, writing, speaking, and listening. The items mainly used 'I can' statements but included some other items such as 'I feel confident about my grammar' and 'I usually realise when someone is joking…'. Two items of particular interest to the academic community were 'referencing' and 'plagiarism'. The pronunciation items were 'Most

people can understand my pronunciation' and 'I feel confident about my pronunciation'. A simple tick in a column headed 'yes' was the required response to each statement; a box without a tick acted as a prompt for the student to consider improvement in the skill concerned. Students were reminded to complete the checklist every six months or so while they were studying at the university.

The purpose of this checklist was to enable students to identify areas where they would welcome additional help in areas linked to their achievement on academic courses in which they were enrolled, so therefore each section concluded with a statement 'I would like to improve' which allowed students to respond freely. This checklist assumed that students would accurately identify their needs in order for them to access further appropriate sources of help. In this respect, it was actually more oriented toward students' perceived needs, which would depend on their individual confidence level, but as a mechanism for signposting many students to a source of further help fulfilled its purpose.

Borg and Edmett (2019) reported and examined the design of a Self-Assessment Tool (SAT) for English teachers to use in identifying their competence and professional development needs. It was based on the British Council's continuing professional development (CPD) framework which consisted of 139 individual elements grouped within 12 broad professional practice areas (p. 659). In designing the SAT, the first step was to reduce the items to a manageable number that could be responded to in a reasonable amount of time. Areas that lacked global relevance or were covered elsewhere were omitted, after which ELT experts were consulted to reduce the remaining number of items by retaining those which were deemed key elements (Borg & Edmett, 2019, p. 659). Items were mainly expressed as 'can do' statements. Two 5-point scales were selected; each included an option for respondents to say they did not understand a statement, so that they were not forced to rate their competence if they were uncertain of the meaning (p. 660). Several reviews of the SAT followed before piloting. Responses

to an open question inviting comments were analysed by one author then a sample checked by the other; in line with good practice, the procedure and percentage agreements were reported (p. 661). A number of items were reworded for clarity; for example, where two ideas were contained in a single statement, one was removed. Some respondents wanted additional or different options on the Likert-type scales, although the authors kept to five but kept this decision under review, recognising that 'not applicable' could be an appropriate option that would take account of very different teaching contexts. The issue of accuracy was discussed because teachers were making judgements about their level of competence. The authors justified their approach on the basis that the SAT was created to be "a formative tool rather than one designed to make summative judgements about teachers" (Borg & Edmett, 2019, p. 672). It is notable that their decision contrasted with that of Jankowska and Zielińska (2015) who intended to address discrepancies between students' and teachers' assessments. Some features of the SAT development process were similar to those in the present study, such as the involvement of expert judgement and the attention paid to number and clarity of items, although the piloting of the SAT was on a much larger scale.

5.7 Development research

There have been calls for increased involvement of learners in instructional design for many years, but many teachers and learners are not used to implementing this approach, and implementation can prove challenging (Richey & Klein, 2014, p. 1108). Development research offers a way of involving learners at an early stage, if a more constructivist approach is wanted. From a practical point of view, teachers' experience and insights into the classroom context are potentially valuable sources of knowledge that can inform the development of pedagogical tools, but they are too often excluded from the early stages of development (Durkee, 2020). Historically, teachers and learners have tended to be involved in the later stages of end user testing and approval, when changes

can still be made but increase the costs of development. In the present study, the teachers' and learners' involvement in the earliest stages was invaluable in improving the wording of the checklist, the positioning of the guidance, and recommendations for the layout of the checklist in Google Forms. Choosing a development research process made this possible, which in turn facilitated the trialling of the checklist and may have contributed to the positive responses in the evaluation of usefulness questionnaires.

However, development research is centred on the development of projects in a particular context, which means that the findings are unlikely to be capable of generalisation, and this may make development research less attractive to many academic research communities. On the other hand, research projects designed to meet a very specific need tend to generate a range of projects which can amount to a considerable body of new knowledge over time.

A more traditional approach would have been to separate expert validation as a distinct first stage before seeking end user views, and to follow this with statistical validation of the checklist before trialling the tool. However, this would have led to a different construct of pronunciation, a different set of criteria, and possibly a more wordy explanation of the technical terms and more wordy items. It was clear from end user perspectives that video clips and fewer words, rather than more words, were preferred. Thus, a more orthodox approach might have produced a result acceptable to experts but not to users. Moreover, since the pedagogical tool is not intended to generate results that can be compared to test results, statistical validation is less important in this particular case.

Another more traditional research design would have been to employ the Delphi method, which requires several rounds of questionnaires to be completed by experts in order to reach consensus. However, early approaches and repeated approaches to experts failed

to identify enough experts with a particular interest in self-assessment of pronunciation to produce a panel of experts, so a Delphi method was not feasible in the case of the present study.

There are other contemporary models of instructional design, which are mostly used for designing whole courses, especially online learning. Many are based on the ADDIE model, which is the longest established such mode. The ADDIE model of instructional systems design originated in the United States military in the mid-20th century as a rigorous approach to designing technical training, since when it has evolved and also given rise to other models of instructional design (Allen, 2006, p. 430-431). ADDIE stands for the five phases of Analysis, Design, Develop, Implement, and Evaluate (Shelton & Saltsman, 2006, p. 14). ADDIE is essentially linear, with detailed steps in each phase; the output from each phase typically leads into the next phase, reflecting its behaviourist origins (Allen, 2006, p. 432). In recent years it has increasingly incorporated the principles and practices of rapid prototyping in response to the ever-greater involvement of software designers and programmers in creating courses and materials, which means feedback from developers, team members and end users is reviewed and changes are made during development (e.g., Shakeel et al., 2023). Whilst ADDIE offers a complete process for the design and development of a whole course, the constraints of the context for the present study would have made it impossible to move beyond the development phase. For example, it remains conventional to complete the design of a module drawing on the expertise of instructors and experts, whilst taking into account user expectations, before piloting the module or instrument with students, with the emphasis on correct design. In the present study, the checklist would have been delayed and, more importantly, would have involved the addition of challenging written explanations of stress and intonation (presumably with stress and intonation markings on examples) before being piloted by students. In the present study, the involvement of users

in early consideration of the checklist meant that unnecessary effort and time was saved. Moreover, the analysis phase is essential for developers who are external to the end users, but where the researcher is a teacher and developer, as in the present study, it is possible for sufficient knowledge of the learning needs and context to be available without an analysis phase as prescribed in ADDIE.

On balance, due to the input from the end users' perspectives during the design and calibration stages, the development research approach saved time in this present study. Additionally, revised and update versions of the ADDIE model are expected to lead to final evaluation of the end product; the five phases are non-negotiable (Allen, 2006, p. 438). Depending on available resources in the research and learning environments to complete the process, the process is likely to take longer than a development research approach, which offers flexibility in terms of the boundary of the project and the number of phases needed to complete it.

5.8 Contribution to knowledge

Despite the recent increase in pronunciation-related studies, such as Strachan et al., (2019), Khonamri et al., (2021) and Brannen et al. (2022), there remains a lack of pedagogical tools that can stimulate students to think about their learning strategies in specific contexts. The availability of multi-media language courses and applications that give feedback on the accuracy of pronunciation does not appeal to every learner as something they can do in their own time. An opportunity to explore learning styles and consider alternatives enables students to consider and try other methods that might suit them better, and there was evidence from the interviews that the checklist prompted some students to become aware of their strategies and regulation of cognition. There was potential to explore this much further in follow-up activities such as peer discussions.

The involvement of students in the design and development stages enabled their contributions to be taken into account and thus give at least some of them a sense of ownership of the checklist. Moreover, their input to considerations of design for Google Forms associated with the use of a variety of mobile phones was invaluable. End user feedback in all stages, as proposed by Richey and Klein (2005), is quite widely accepted in some fields such as medicine and on-line learning programmes. However, the involvement of students and teachers in developing a pedagogical tool for classroom use remains underrepresented in research and the present study provides a clear example of the benefits of involving them in this way.

5.9 Limitations of the study

The checklist was developed for a specific target group. This may limit its applicability to other educational contexts where a different approach is taken to pronunciation, although its potential use in ESP/ESL courses has been highlighted, and the design principles may be more widely applicable. Practical limitations arose in the course of the study, notably time constraints and the availability of participants, that may have reduced the benefits of end user involvement in terms of exploring the next steps to further refine the self-assessment instrument and to conduct a further round of formative evaluation. The impact of off-site learning during COVID-19 reduced the available contact time with student participants and caused difficulties in scheduling some of the steps in the process.

A second limitation which resulted from these constraints is that the checklist was not implemented, thus conclusions cannot be reached about whether students would adjust their learning strategies or undertake more independent learning as a result of using the checklist more than once in a real-life classroom context.

5.10 Reflections on the research approach

A development research approach was chosen because of its capacity to put the end users at the heart of the process and the flexibility it offered in terms of dealing with the real-world complexities of educational settings. The flexibility enabled some of the complexities and challenges to be overcome, although the flexibility itself can present challenges for the researcher. A major challenge in this present study concerned the timing and sequences of steps in the research process. For example, expert validation comments and teacher and student comments came immediately after each other in the design phase, due to a combination of timetabling constraints and public holidays. Another example of problems with the timing was the difficulty and delay in obtaining sufficient expert validation until after the checklist had been trialled, due to a combination of some experts being extremely busy, others retired, and a mismatch between the checklist and several experts' research interests at the time. Combining consideration of expert validation with end user views is supported by the partnership approach to design, while eliciting expert validation at a later time is supported by the iterative process of development research (Richey & Klein, 2014, p. 1118).

The researcher faced some important challenges to establishing and maintaining the focus of design of the checklist in the face of comments from experts on two separate occasions that indicated that some aspects of the design were fundamentally unsuitable. The design aspects concerned were the selection of the ICAO scale descriptors in the design phase, and then, much later, the topic choice of a checklist for classroom use. A similar challenge arose with the evaluation of usefulness questionnaire, with expert validation comments leading the researcher to include detailed explanations of the model adopted, and of each aspect of the impact domain of the model. The researcher was new to development research and new to self-assessment, which led her to initially assume her design must be wrong, but then she found herself reflecting-in-action during the periods

of uncertainty, revisiting her understanding of the factors involved, seeking additional views and knowledge if needed, until she was able to make a decision. Reflection-in-action and allowing oneself to feel confused and uncertain are part of the development research experience (McKenney & Reeves, 2014; Tracey et al., 2014). Although the researcher found the experiences difficult at times, she also found they led to greater clarity in the next step of the process and greater clarity overall.

Whilst the overall flexibility offered by development research was valuable, the two-year time limit set for the present study did not permit progression to a further iteration of the checklist, which was disappointing as there was no opportunity to identify whether there would be a feedforward effect. The less usual choice of development research, which was based on involving teachers and students as partners in the study from an early stage, also revealed an aspect of what Rose (2019) described as a widening gap between the research community and teacher-practitioners in the TESOL community, a gap which is also reported in the applied linguistics field (Becker, 2023, p. 1). Different disciplines have different traditions of research and methodologies which can conflict with each other, and at times the researcher was aware of the tensions between them. For example, the literature review showed that the majority of subject-relevant studies employed research designs that could not be applied in a wide range of educational settings. Many of the studies involved participants who were English majors, students undertaking pronunciation courses, or international students on preparatory English courses. Their interests and expectations were too dissimilar to technical students in an ESL environment for the instruments and procedures to be directly transferable to the present study.

The nature of the development research process and the scope of product evaluation within it meant the researcher frequently returned to the literature to seek theory or applied research that supported or refuted suggestions for improvement, in agreement with the idea of designer or developer as a person who gives structure to a problem and

then solves it (Richey & Klein, 2014, p. 1118-1119). Adjustments are considered to be an unavoidable part of the development process, in contrast to the more linear approach of ADDIE, and the designer of an instructional tool needs to be able to respond to changes by reflecting on uncertainties as they occur (Tracey et al., 2014). The researcher's personal preference at the start of the study was for greater certainty about the steps to be taken and the time to take them, possibly because her teaching experience had enabled her to know how and when to be flexible within the constraints of a timetable and curriculum. In other words, there was less control in the research environment than in the teaching environment. However, through adopting a development research approach and reflecting on the unexpected difficulties that occurred, she was able to adjust more easily to such problems by the end of the study.

5.11 Reflections as a teacher-researcher

A personal challenge for the researcher was to balance her role as a teacher with her role as a researcher, in terms of time as well as positionality. Richey and Klein (2014, p.1116) mention that it is not uncommon to find researchers who are also study participants, although this is not ideal. Corroboration of data, where possible, is recommended, and transparency regarding data collection and analysis is essential. In the present study, existing relationships with other members of the teaching team and regular contact with students initially facilitated access to participants but the dates and times for trialling the checklist and involving a second coder in the data analysis stage needed considerable negotiation.

The teacher-researcher role is often seen as a problem, whereas it can be an enriching experience. The researcher in the present study has understood the importance of consulting the literature and previous studies to inform classroom practice when appropriate. Based on her teaching experience, she has understood the importance of

bringing other teachers' experience, as well as the learners' perspectives, into applied research in the classroom.

At the same time, she was aware of her positionality as an insider-outsider and of the imbalance of power in the teacher-student relationship. The methodology chapter (Chapter 3) referred to her reaction in some of the early interviews to rephrase a question without giving the student time to think, a habit formed from her experience in the classroom of rephrasing for clarity of understanding or a desire to help out a student who might be struggling to find the right words. As Rose (2020) mentions, interviews are not simply data collection techniques but involve the interviewer and interviewee in co-constructing knowledge (p. 114) and the researcher had to be careful not to impose her knowledge of students on their responses as interviewees. Too much rephrasing could easily have led to students giving a response they thought was expected when they noticed a change in the wording of a question. The researcher's new awareness also led her to reflect on the balance in the classroom between 'teacher talk' and 'elicitation from students', depending on the content of the session; the interviewing experience provided feedback into her role as a teacher. Moreover, she was aware of the potential impact of power imbalance between teacher and students, especially in the Malaysian context where teachers are still usually formally addressed by their students, and she encouraged them to say more. There was some evidence that the power imbalance had relatively little effect; one of the ten students interviewed responded to her final question 'Do you wish to add anything?' by advising her on how to conduct the study, even though he admitted he had no knowledge of such research. The fact that some students felt able to challenge the criteria by asserting that only native speakers could reach the highest level on the ICAO scale also suggests they were comfortable with the teacher-student relationship.

The methodology chapter also documented the researcher's struggles in the early stages of data analysis to avoid making assumptions and moving too quickly to interpretations

based on her knowledge and experience of teaching rather than on the data themselves. Once she became aware of this, she checked every time that she returned to the data analysis that she was not making assumptions based on her teaching experience. The experience of creating a codebook and assuring inter-coder and code reliability was an additional way of testing her own relationship with the data, and another step in what Palaganas et al. (2017) describe as a learning journey for a qualitative researcher (p. 436). The journey continues, as there is further research that would usefully follow from the present study.

5.12 Directions for future research

Development research is a research approach located within the pragmatic paradigm that is suitable for applied research in instructional design, including the design of pedagogical tools. Its suitability for addressing complex problems in real-world contexts such as classrooms, and in partnerships with end users, experts and potentially other practitioners, may offer an alternative to more traditional approaches, whether action research, which has traditionally been used in classroom-based research, or refinement and validation through statistical techniques.

The present study has only really captured the first stage of a longer research process. The guidance in the checklist will need to be updated to include new sources of help as they arise, such as the use of AI and chatbots in the process of learning how to self-assess their pronunciation. The teachers and a majority of students who returned valid completed questionnaires judged the checklist to be very useful, and there was some evidence of knowledge and regulation of cognition, as well as some students stating they had identified areas for improvement. However, the effect of the checklist on students' learning strategies and their pronunciation remains to be investigated through a longitudinal study covering two or three semesters. It would also be enlightening to

investigate how students give themselves feedback and/or whether they use feedforward in order to determine how to implement the checklist most effectively. In addition, the use of the checklist alone by one group of students could be compared with other groups using peer discussion or feedback from their teacher, as well as with a control group in order to investigate which type of feedback might be most productive and whether a specific approach would promote feedforward more than the others.

5.13 Conclusion

The study has designed and developed a pronunciation self-assessment checklist as a pedagogical tool for use by BAET students in a technical university in Malaysia. Expert validation as well as students' and teachers' comments on the clarity and feasibility of the checklist were used to improve the checklist. After trialling the second version of the checklist, students' and teachers' reactions to it were gathered through semi-structured interviews with a sample of students and both the teachers. Finally, the researcher collated and analysed responses to an evaluation of usefulness questionnaire, which covered the domains of reliability, validity, impact and practicality.

There was sufficient evidence that the checklist could encourage students to think about their thinking, and develop a metacognitive approach to reflection and self-assessment, to make implementation and continued research worthwhile. Students' comments in the evaluation of usefulness questionnaire clearly indicated that there remains much more to be done before they fully grasp the importance of independent learning, since they requested more teaching about pronunciation rather than seeking the information for themselves. Nonetheless, the checklist offered a promising beginning to the process.

The development research approach selected for the study proved challenging due to the lack of previous studies which could provide a model that could be more easily followed. However, its flexibility was helpful to completing the study and it broadened the scope

of the researcher's knowledge and experience. In particular it highlighted the contribution that students can make to the design of pedagogical tools.

REFERENCES

Accessibility Guidelines Working Group (2022).

Ahmad, S., Sultana, N., & Jamil, S. (2020). Behaviorism vs constructivism: A paradigm shift from traditional to alternative assessment techniques. *Journal of Applied Linguistics and Language Research, 7*, 19-33.

Al Bashir, M. M. (2016). The value and effectiveness of feedback in improving students' learning and professionalizing teaching in higher education. *Journal of Education and Practice, 7*(16). 38-41.

Alderson, J.C. (2009). Air safety, language assessment policy, and policy implementation: The case of Aviation English. *Annual Review of Applied Linguistics*, *29*, 168–187. doi:10.1017/S0267190509090138

Alderson, J. C. (2011). The politics of Aviation English testing. *Language Assessment Quarterly, 8*(4), 386-403. https://doi.org/10.1080/15434303.2011.622017

Aljohani, M. (2017). Principles of constructivism in foreign language teaching. *Journal of Literature and Art Studies, 7*, 97-107. https://doi.org/10.17265/2159-5836/2017.01.013

Allen, W. C. (2006). Overview and evolution of the ADDIE training system. *Advances in Developing Human Resources, 8*(4), 430-441. DOI: 10.1177/1523422306292942

Amerstofer, C. M. (2018) Past its expiry date? The SILL in modern mixed-methods strategy research. *Studies in Second Language Learning and Teaching, 7*(2), 497-523. https://doi.org/10.14746/ssllt.2018.8.2.14

Anderson, N. J. (2008). Metacognition and good language learners. In C. Griffiths (Ed.), *Lessons from Good Language Learners (Cambridge Language Teaching Library)* (pp. 99-109). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511497667.010

Anderson, N. (2012). Metacognition: Awareness of language learning. In S. Mercer et al. (Eds.), *Psychology for Language Learning.* Basingstoke, Palgrave Macmillan. https://doi.org/169-18710.1057/9781137032829_12

Andrade, H., (2013). Classroom assessment in the context of learning theory and research. In J. H. McMillan (Ed.). *Sage handbook of research on classroom assessment*. (pp. 17–34). Sage. https://doi.org/10.4135/9781452218649.n2

Andrade, H, (2019). A critical review of research on student self-assessment. *Front. Educ. 4*(87), 1-12. https://doi.org/10.3389/feduc.2019.0008

Andrade, H., & Du, Y. (2007) Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education, 32*(2), 159–181. https://doi.org/10.1080/02602930600801928

Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary

school students' writing. *Educ. Meas. 27*, 3–13. doi: 10.1111/j.1745-3992.2008.00118.x

Andrade, H., & Valtcheva, A. (2009) Promoting learning and achievement through self-assessment. *Theory Into Practice, 48*(1), 12-19. https://doi.org./10.1080/00405840802577544

Anh, K. H., Dong, M. H., & Trang, N. H. (2022). An investigation into English-majored students' self-assessment of their speaking performance. *International Journal of Instruction, 15*(3), 191-208. https://doi.org/10.29333/iji.2022.15311a

Anjomshoa, L., & Sadighi, F. (2015). The importance of motivation in second language acquisition. *International Journal on Studies in English Language and Literature (IJSELL), 3*(2), 126-137. ISSN 2347-3134 (Online). www.arcjournals.org

Ardnt, J. D. (2017). Self-directed learning for English language Learners. *The Center for ELF Journal, 3*, 38-58. https://www.tamagawa.ac.jp/celf/assets/research_pdf01/celf_journal_final3.pdf

Arndt, H., & Rose, H. (2023). Capturing life as it is truly lived? Improving diary data in educational research. *International Journal of Research & Method in Education, 46*(2), 175-186. https://doi.org/10.1080/1743727X.2022.2094360

Arnold, J. (2022). Prioritising students in assessment for learning: a scoping review of research on students' classroom experience. *BERA Review of Education, 10*(3), 36. https://doi.org/10.1002/rev3.3366

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267. https://doi.org/10.3102/0013189x07306523

Azevedo, R. (2020). Reflections on the field of metacognition: Issues, challenges, and opportunities. *Metacognition and Learning, 15*(2), 919–998.

Aziz, A. A., & Kashinathan, S. (2021). ESL learners' challenges in speaking English in Malaysian classroom. *International Journal of Academic Research in Progressive Education and Development, 10*(2), 983–991. https://api.semanticscholar.org/CorpusID:242861784

Bachman, L. F. (2000). Foreword.. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL*. (pp. ix-xii). New Jersey: Lawrence Erlbaum Associates, Inc.

Bachman, P., & Palmer, A.S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Baniabdelrahman, A. A. (2010). The effect of the use of self-assessment on EFL students' performance in reading comprehension in English. *Teaching English as a Second or Foreign Language*, 14(2).

Becker, A. (2023). Applied linguistics communities of practice: Improving the research-practice relationship. *Applied Linguistics*, amad010, https://doi.org/10.1093/applin/amad010

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7-74. doi:10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the Theory of Formative Assessment. *Educational Assessment, Evaluation and Accountability*, *21*, 5-31. https://doi.org/10.1007/s11092-008-9068-

Borg, S., & Edmett, A. (2019). Developing a self-assessment tool for English language teachers. *Language teaching research*, *23*(5), 655-679. https://doi.org/10.1177/1362168817752543

Boud, D. (1995). Developing a typology for learner self-assessment practices. *Research and Development in Higher Education, 18*, 130-135.

Boud, D. and Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment and Evaluation in Higher Education, 38*(6), 698-712. https://doi.org/10.1080/02602938.2012.691462

Bourke, B. (2014). Positionality: Reflecting on the research process. *The Qualitative Report, 19*(33), 1-9. https://doi.org/10.46743/2160-3715/2014.1026

Brannen, K., Rosales, E., Wouters, I., & John, P. (2022). The effects of self-assessment activities on accuracy and awareness in ESL pronunciation classes. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. https://doi.org/10.31274/psllt.13262

British Educational Research Association (2011). *Ethical Guidelines for Educational Research 2011.* British Educational Research Association,

Brown, H. D. (2007). *Principles of Language Learning and Teaching.* Pearson Longman.

Brown, G. T. L., & Harris, L. R. (2013). *Student Self-Assessment. In J. H. McMillan (Ed.), The SAGE Handbook of Research on Classroom Assessment* (pp. 367-393). SAGE.

Burns, A. (2003). Understanding pronunciation. In A. Burns., & Claire, S. (2003). *Clearly speaking: Pronunciation in action for teachers*. (pp. 5-8). Sydney, National Centre for English Language Teaching and Research.

Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity, 56*, 1391–1412. https://doi.org/10.1007/s11135-021-01182-y

Cañete, G., & Inostroza, M. J. (2022). Exploring the contribution of self-assessment checklists to improve oral presentations. *HOW,* 29(5). 57-80. https://doi.org/10.19183/how.29.2.702

Carless, D. Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4) 395-407. https://doi.org/10.1080/03075071003642449

Chang, J. (2001). Chinese speakers. In: M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems*. (2nd edition). (pp. 310-324). Cambridge University Press.

Cheng, L., Sun, Y., & Ma, J. (2015). State-of-the-art article Review of washback research literature within Kane's argument-based validation framework. *Language Teaching, 48*(4), 436-470. doi:10.1017/S0261444815000233.

Cheong, K. C., Hill, C., Fernandez-Chung, R., & Leong, Y. C. (2016). Employing the 'unemployable': Employer perceptions of Malaysian graduates. *Studies in Higher Education, 41*(12), 2253-2270. https://doi.org/10.1080/03075079.2015.1034260

Cheung, K. K. C., & Tai, K. W. H. (2021). The use of intercoder reliability in qualitative interview data analysis in science education. *Research in Science & Technology Education,* 1-21. https://doi.org/10.1080/02635143.2021.1993179

Cieślicka, E. G., & Rojczyk, A. (2017). Self-reported vs. self-rated pronunciation in a non-native language. T*heory and Practice of Second Language Acquisition*, 3(2), 69-86. http://polona.pl/item/83076809

Civil Aviation Authority of Malaysia. (2021). *List of CAAM Part 147 Approved Maintenance Training Organisations.* https://www.caam.gov.my/wp-content/uploads/2021/07/MTO-Listing-2021-V1-20210706.pdf

Clough, P., & Nutbrown, C. (2012) *A student's guide to methodology*. London, SAGE Publications Ltd.

Clynes, A., & Deterding, D. (2011). Standard Malay (Brunei). *Journal of the International Phonetic Association, 41*(2), 259–268. http://www.jstor.org/stable/44527038

Cojo Guatame, A. (2019). *Fostering segmental pronunciation through self-assessment*. Unpublished Master Report. Universidad de La Sabana.

Cookson, S. (2011). "Tell them we are in emergency": Linguistic factors contributing to the crash of Avianca Flight 052. *Studies in Language and Culture*, *2*, 17-33.

Coşkun, Y. (2018). A study on metacognitive thinking skills of university students. *Journal of Education and Training Studies, 6* (3), 38-46. ISSN 2324-805X E-ISSN 2324-8068. URL: http://jets.redfame.com

Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment (CEFR)*. Strasbourg, France: Council of Europe (Modern Languages Division)

Council of Europe (2018). *Common European framework of reference for languages: Learning, teaching, assessment: Companion Volume with New Descriptors*. Strasbourg, France: Council of Europe. Council_of_Europe_2018_Common_European_F.pdf

Couper, G. (2019). Teachers' cognitions of corrective feedback on pronunciation: Their beliefs, perceptions and practices. *ScienceDirect System, 84,* 41-52. https://doi.org/10.1016/j.system.2019.04.003

Creswell, J. W. A. (2014). *Concise introduction to mixed methods research.* (4th edition). SAGE Publications.

Croasmun, J. T., & Ostrom, L. (2011). Using Likert-type scales in the social sciences. *Journal of Adult Education, 40* (1), 19-22.

Cruttenden, A. (Ed.). (2014). *Gimson's Pronunciation of English*. (8th edition.) Routledge.

Darmi, R. (2013). *Learning English: Challenges for Malaysian language learners*. Conference: 11th Asia TEFL International Conference. At: Ateneo de Manila University

Dawadi, S. (2021). Factors affecting washback of a high-stakes English as a foreign language test. *TESL-EJ: The Electronic Journal for English as a Second Language*, 25(3). https://tesl-ej.org/pdf/ej99/a1.pdf

De Saint Léger, D. (2009). Self-assessment of speaking skills and participation in a foreign language class. *Foreign Language Annals, 42*(1), 158-178. https://doi.org/10.1111/j.1944-9720.2009.01013.x

DeCuir-Gunby, J. T., Marshall, P. L., & McCulloch, A. W. (2011). Developing and using a codebook for the analysis of interview data: An example from a professional development research project. *Field Methods, 23*(2), 136–155. https://doi.org/10.1177/1525822X10388468

Demir, Y., & Ertas, A. (2014). A suggested eclectic checklist for ELT coursebook evaluation. *The Reading Matrix, 14*(2), 243-254.

Department of Civil Aviation Malaysia (2014). English Language Proficiency (ELP) for aircraft maintenance licence (AML) personnel. Airworthiness Guidance Ag No. 1101, Issue 1. https://www.caam.gov.my/wp-content/uploads/2021/02/6.-AG1101ELP.pdf

Derakhshan, A., & Karimi, E. (2015). The interference of first language and second language acquisition. *Theory and Practice in Language Studies, 5*(10), 2112-2117. http://dx.doi.org/10.17507/tpls.0510.19

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*, 1–16. https://doi.org/10.1017/S0272263197001010

Derwing T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379–397. https://doi.org/10.2307/3588486

Derwing, T., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42*(4), 476-490. doi:10.1017/S026144480800551X

Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System, 30*, 155-166. http://dx.doi.org/10.1016/S0346-251X

Deterding, D. H. (2011). Measurements of the rhythm of Malay. Online proceedings of the 17th ICPhS XVII, Hong Kong, 17-21 August 2011. (pp. 576-579). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Deterding/Deterding.pdf.

Deterding, D. H. (2015). Segmentals. In: M. Reed & J. M. Levis (Eds.). *The Handbook of English Pronunciation.* (pp. 67-84). John Wiley & Sons, Inc.

Deterding, D. H., Gardiner, I. A., & Noorashid, N. (2022). *The phonetics of Malay*. Cambridge University Press

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Heath & Co

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*, 391–409. https://doi.org/10.1007/s10648-008-9083-6

Dlaska, A., & Krekeler, C. (2008). Self-assessment of pronunciation. *System: An International Journal of Educational Technology and Applied Linguistics, 36*(4), 506-516. https://doi.org/10.1016/j.system.2008.03.003

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education A review. *Studies in Higher Education, 24*, 331-350. doi:10.1080/0307507991233137993

Dörnyei, Z. (2003). *Questionnaires in Second Language Research: Construction, Administration, and Processing*. New Jersey. Lawrence Erlbaum Associates

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press

Dörnyei, Z., & Muir, C. (2019). Creating a Motivating Classroom Environment. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 719-736). Springer. https://doi.org/10.1007/978-3-030-02899-2_36

Douven, I. (2018). A Bayesian perspective on Likert scales and central tendency. *Psychonomic Bulletin & Review, 25*(3), 1203–1211. https://doi.org/10.3758/s13423-017-1344-

Dreyer, C. & Oxford, R., (1996). Learning strategies and other predictors of ESL proficiency among Afrikaans-speakers in South Africa. In R. Oxford (Ed.), *Language learning strategies around the world: Cross-cultural perspectives* (pp. 61-74). University of Hawaii.

Drury, C. G., & Ma, J. (2002). *Language error analysis: Report on literature of aviation language errors and analysis of error databases*. Federal Aviation Administration. University of Buffalo, State University of New York.

Drury, C. G., & Ma, J. (2003). *Language errors in aviation maintenance: Year 1 interim report*. Reports to William J. Hughes Technical Center, the Federal Aviation Administration under research grant #2002-G-025.

Drury, C. G., & Ma, J. (2004). Experiments on language errors in aviation maintenance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48*(1), 118–122. https://doi.org/10.1177/154193120404800126

Drury, C. G., Ma, J., & Marin, C. V. (2005). *Language error in aviation maintenance: Final report.* Prepared for: Federal Aviation Administration William J. Hughes Technical Center, the Federal Aviation Administration under research grant #2002-G-025

Dubiner, D. (2018). Second language learning and teaching: From theory to a practical checklist. *TESOL Journal, 10*(2), e00398. https://doi.org/10.1002/tesj.398

Dudley-Evans, T., & St John. M. J. (1998). *Developments in English for specific purposes*. Cambridge University Press.

Dunham, S., Lee, E. & Persky, A. M. (2020). The psychology of following instructions and its implications. *American Journal of Pharmaceutical Education, 84*(8), 1052-1056. https://doi.org/10.5688/ajpe7779

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. SAGE Publications

Dunworth, K., & Briguglio, C. (2011). *Teaching students who have English as an additional language: a handbook for academic staff in higher education.* Higher Education Research and Development Society of Australasia.

Durkee, E. K. (2020). *Instructional tool design and development using formative evaluation: Qualitative case study*. [Unpublished doctoral thesis]. Cappella University. ProQuest Number 27743537

Eckstein, G. T. (2007). A correlation of pronunciation learning strategies with spontaneous English pronunciation of adult ESL learners. [Unpublished master dissertation]. Brigham Young University.

Educational Testing Service (2019). TOEFL iBT® independent speaking rubric. https://www.ets.org/content/dam/ets-org/pdfs/toefl/toefl-ibt-speaking-rubrics.pdf

Ellis, A. K., Denton, D. W., & Bond, J. B. (2014). An analysis of research on metacognitive teaching strategies. *Procedia - Social and Behavioral Sciences, 116*, 4015-4024.

Enxhi, S.Y., Hoon, T.B., & Fung, Y.M. (2012). Speech disfluencies and mispronunciations in English oral communication among Malaysian undergraduates. *International Journal of Applied Linguistics and English Literature, 1*, 19-32. DOI:10.7575/IJALEL.V.1N.7P.19

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59*, 395-430. https://doi.org/10.2307/1170205

Farris, C., Trofimovich, N., Segalowitz, N., & Gatbonton, E. (2008). Air traffic communication in a second language: Implications of cognitive factors for training and assessment. *TESOL Quarterly, 42*(3), 397–410. https://doi.org/10.1002/j.1545-7249.2008.tb00138.x

Feak, C. B. (2014). ESP and speaking. In B. Paltridge, & S. Starfield (Eds.), *The handbook of English for specific purposes* (2nd edition). (pp. 35-54). John Wiley & Sons

Fisher, R. (1998). Thinking about thinking: Developing metacognition in children. *Early Child Development and Care, 141*(1), 1-15. https://doi.org/10.1080/0300443981410101

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence*. (pp. 231-235). Lawrence Erlbaum.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906-911.

Flowerdew, L. (2014). Needs analysis and curriculum development in ESP. In B. Paltridge, & S. Starfield (Eds.), *The handbook of English for specific purposes* (2nd edition). (pp. 325-346). John Wiley & Sons

Frank, R. A. (2000). Medical communication: non-native English speaking patients and native English speaking professionals. *English for Specific Purposes, 1*(19), 31-62. https://doi.org/10.1016/S0889-4906(98)00012-X

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: an advanced resource book*. Routledge.

Fulcher, G. & Davidson, F. (Eds.). (2012). *Routledge handbook of language testing*. Routledge.

Galaczi, E. D. (2016). Assessing second language pronunciation: Distinguishing features of rhythm in learner speech at different proficiency levels. In T. Isaacs, & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives*. (pp.157-182). Multilingual Matters

Giraldo, F. (2019). Designing language assessments in context: theoretical, technical, and institutional considerations. *HOW, 26*(2), 123-143. DOI:10.19183/how.26.2.512

Gralińska-Brawata, A. (2022). Polish learners' self-assessment and reflections on their pronunciation progress. *Językoznawstwo, 2*(17), 229-252. https://doi.org/10.25312/2391-5137.17/2022_17agb

Gu, Y. (2014). To code or not to code: Dilemmas in analysing think-aloud protocols in learning strategies research. *System, 43*, 74-81. https://doi.org/10.1016/j.system.2013.12.011

Gu, P. Y. (2021). An argument-based framework for validating formative assessment in the classroom, *Frontiers in Education, 26*, Article 605999. https://doi.org/10.3389/feduc.2021.605999

Hammersley, M. (2012). *Methodological paradigms in educational research.* British Educational Research Association on-line resource. Available on-line at https://www.bera.ac.uk/publication/methodological-paradigms-in-educational-research

Harder, B., Trottler, S., Vialle, W. & Ziegler, A. (2015). Diagnosing resources for effective learning via teacher and parent checklists. *Psychological Test and Assessment Modeling, 57*(2), 201-221.

Harding, L. (2017). What do raters need in a pronunciation scale?: The user's view. In T. Isaacs & P. Trofimovich (Eds.), *Second Language Pronunciation Assessment: Interdisciplinary Perspectives*, (Vol. 107, pp. 12–34). Multilingual Matters / Channel View Publications. http://www.jstor.org/stable/10.21832/j.ctt1xp3wcc.6

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education, 35*(2), 213 – 224. https://doi.org/10.1080/03057640500146880

Harlen, W. (2005). Teachers' summative practices and assessment for learning: Tensions and synergies. *Curriculum Journal, 16*(2), 207-223. DOI:10.1080/09585170500136093

Hashim, H. U., & Yunus, M. M. (2018). English as a second language (ESL) learning: setting the right environment for second language acquisition. *Tadris: Jurnal Keguruan dan Ilmu Tarbiyah 3*(2), 207-215 (2018) DOI: 10.24042/tadris.v3i2.2941

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Oxford, UK: Routledge. https://doi.org/10.4324/9780203887332

Hattie, J. A. C., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https:// https://doi.org/10.3102/003465430298487

Haukås, A., Bjørke, C., & Dypedahl, M. (2018). Introduction. *Metacognition in language learning and teaching*. (pp. 1-10). Routledge.

Hismanoglu, M., & Hismanoglu, S. (2011). Task-based language teaching: What every EFL teacher should do. *Procedia - Social and Behavioral Sciences, 15*, 46-52. https://doi.org/10.1016/j.sbspro.2011.03.049

Hodgetts, J. (2020). *Pronunciation instruction in English for academic purposes: An investigation of attitudes, beliefs and practices*. Springer

Holmes, A. G. D. (2020). Researcher positionality - a consideration of its influence and place in qualitative research - a new researcher guide. *Shanlax International Journal of Education* (*8*)4, 1-10. https://doi.org/10.34293/education.v8i4.3232

Hosseini, M., & Nimehchisalem, V. (2021). Self-assessment in English language teaching and learning in the current decade (2010-2020): A systematic review. *Open Journal of Modern Linguistics, 11*, 854-872. https://doi.org/10.4236/ojml.2021.116066

Howatt, A. P. R., & Widdowson, H. G. (2004). *A history of ELT* (2nd edition). Oxford University Press.

Huang, B. H., Alegre, A., & Eisenberg, A. R. (2016). A cross-linguistic investigation of the effect of raters' accent familiarity on speaking assessment. *Language Assessment Quarterly, 13*, 25-41. https://doi.org/10.1080/15434303.2015.1134540

Huang, Y. (n.d.). *Analysis on the Mandarin vowel system and the English vowel system*. [Paper presentation]. The International Conference on Interdisciplinary Humanities and Communication Studies. DOI: 10.54254/2753-7064/3/2022719

Hurd, S. (2008). Second language learning at a distance: metacognition, affect, learning strategies and learner support in relation to the development of autonomy. (Doctoral thesis, The Open University). https://doi.org/10.21954/ou.ro.0000550d

International Civil Aviation Organization (2010.). Publication of Doc 9835-AN/453. Order Number: 9835 ISBN 978-92-9231-549-8. Available at https://www4.icao.int/aelts/uploads/icao%20doc9835%202nd%20edition.pdf

International English Language Testing system (n.d.) IELTS Speaking band descriptors: Scoring criteria for academic and general training tests. https://takeielts.britishcouncil.org/sites/default/files/ielts_speaking_band_descriptors.pdf

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English speaking graduate students. *Canadian Modern Language Review, 64*(4), 555–580.

Isaacs, T. & Harding, L. (2017). Research timeline: pronunciation assessment. *Language Teaching, 50*(3), 347–366. https://doi.org/10.1017/S0261444817000118

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition, 34*(3), 475-505. https://doi.org/10.1017/S0272263112000150

Isaacs, T., & Trofimovich, P. (2017). Second language pronunciation assessment: Interdisciplinary perspectives. *Second Language Acquisition.* Multilingual Matters. https://doi.org/10.21832/ISAACS6848

Isaacs, T., Trofimovich, P., & Foote, J. A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing, 35*(2), 193–216. https://doi.org/10.1177/0265532217703433

Isbell, D. R., & Sakai, M. (2022). Pronunciation assessment in classroom contexts. In J. M. Levis, T. M. Derwing, & S. Sonsaat-Hegelheimer (Eds.), *Second language pronunciation: Bridging the gap between research and teaching*. (pp. 194-214). Wiley-Blackwell

Jamrus, M. H. M., & Razali, A. B. (2019). Using self-assessment as a tool for English language learning. *English Language Teaching, 12*(11), 94-111. https://doi.org/10.5539/elt.v12n11p64

Jankowska, A., & Zielińska, U. (2015). Designing a self-assessment instrument for developing the speaking skill at the advanced level. In: M. Pawlak, & E. Waniek-Klimczak (Eds.), *Issues in teaching, learning and testing speaking in a second language. Second language learning and teaching*. (pp. 251-265). Springer. https://doi.org/10.1007/978-3-642-38339-7_16

Jebb, A.T., Ng, V., & Tay, L. (2021). A review of key Likert scale development advances: 1995–2019. *Front. Psychol. 12*(637547), 1-14. https://doi.org/10.3389/fpsyg.2021.637547

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics, 23*(1), 83–103. https://doi.org/10.1093/applin/23.1.83

Jenkins, J. (2004). Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics, 24, 109-125*. https://doi.org/10.1017/S0267190504000054

Jenkins, J. (2006). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly, 40*(1), 157-181. https://doi.org/10.2307/40264515

Jenkins, J. (2009). Teaching pronunciation for English as a Lingua Franca: A sociopolitical perspective. In C. Gnutzmann & F. Intemann (Eds.), *In The Globalisation of English and the English Language Classroom*, (pp.145-158). Gunter Narr.

Jessner, U. (2018) Metacognition in multilingual learning: A DMM Perspective. In Å. Haukås, C. Bjørke, & M. Dypedahl (Eds.), *Metacognition in language learning and teaching*. (pp. 31-47). Routledge.

Johns, R. A. (2005). One size doesn't fit all: Selecting response scales for attitude items. *Journal of Elections, Public Opinion and Parties, 15*(2), 237-264. https://doi.org/10.1080/13689880500178849

Jones, T. S., & Richey, R. C. (2000). Rapid prototyping methodology in action: A developmental study. *Educational Technology Research and Development, 48*(2), 63-80. https://doi.org/10.1007/BF02313401

Józsa, K., & Morgan, G. A. (2017). Reversed items in Likert scales: Filtering out invalid responders. *Journal of Psychological and Educational Research, 25*(1):7-25. ISSN 2247-1537

Kamarudin, N. I., & Kamal, M. A. A. (2021). Mispronunciation of English monophthong and diphthong among Malay native speakers. *International Journal of Academic Research in Business and Social Sciences, 11*(10), 814–822. http://dx.doi.org/10.6007/IJARBSS/v11-i10/11451

Kaplan, M., Silver, N., Lavaque-Manty, D., & Meizlish, D. (Eds.). (2013). *Using reflection and metacognition to improve student learning.* Virginia: Stylus.

Kashinathan, S., & Abdul Aziz, A. (2022). TOP: Fostering ESL students' overall speaking performance using topical-based oral presentation. *International Journal of Academic Research in Progressive Education and Development, 11*(2), 1194 - 1215. DOI:10.6007/ijarped/v11-i2/13888

Kaufman, D. (2004). Constructivist issues in language learning and teaching. *Annual Review of Applied Linguistics*, 24, 303–319. https://doi.org/10.1017/S0267190504000121

Kember, H., Choi, J., Yu, J., & Cutler, A. (2021). The processing of linguistic prominence. *Lang Speech, 64*(2):413-436. doi: 10.1177/0023830919880217

Kepol, N. (2017). Quality Malaysian English language teachers: Examining a policy strategy. *Malaysian Journal of Learning and Instruction, 14*(1), 187-209.

Kerr, P. (2017). *Giving feedback on speaking.* Part of the Cambridge Papers in ELT series. Cambridge: Cambridge University Press. 2-14. https://www.cambridge.org/elt/blog/wp-content/uploads/2018/02/Cambridge-Press_Whitepaper_Feedback_Speaking_2018.pdf

Khaldi, A. (2017). Self-assessment in foreign language learning. *International Journal of Language Learning and Applied Linguistics World (IJLLALW), 15*(1), 106-110.

Khonamri, F., Králik, R., Vítečková, M., & Petrikovičová, L. (2021). Self-assessment and EFL literature students' oral reproduction of short stories. *European Journal of Contemporary Education, 10*(1), 77-88. https://doi.org/10.13187/ejced.2021.1.77

Kim, H., & Elder, C. (2009). Understanding Aviation English as a lingua franca: Perceptions of Korean aviation personnel. *Australian Review of Applied Linguistics, 32*(3), 23, 1-23.17. https://doi.org/10.2104/aral0923

Kim, H., & Elder, C. (2015). Interrogating the construct of aviation English: Feedback from test takers in Korea. *Language Testing, 32*(2), 129–149. https://doi.org/10.1177/0265532214544394

Kirkpatrick, A. (2011). English as an Asian lingua franca and the multilingual model of ELT. *Language Teaching, 44*(2), 212-224. https://doi.org/10.1017/S0261444810000145

Kirkpatrick, A. (2021). The Routledge handbook of world Englishes. Routlegde. https://doi.org/10.4324/9781003128755

Kolb, D. A. (1984). *Experiential learning experience as the source of learning and development*. Prentice Hall.

Konings, K. D., Seidel, T., & van Merrienboer, J. J. G. (2014). Participatory design of learning environments: integrating perspectives of students, teachers, and designers. *Instructional Science, 42*(1), 1-9. https://doi.org/10.1007/s11251-013-9305-2

Kuhn, D. (2000). Theory of mind, metacognition, and reasoning: A life-span perspective. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind*. (pp. 301–326). Psychology Press/Taylor & Francis UK

Kuhn, D., & Dean, D. (2004). A bridge between cognitive psychology and educational practice. *Theory into Practice, 43*(4), 268-273. https://doi.org/10.1207/s15430421tip4304_4

Kurasaki, K. (2000). Intercoder reliability for validating conclusions drawn from open-ended interview data. *Field Methods, 12*, 179-194. K http://fmx.sagepub.com/cgi/content/abstract/12/3/179

Lazaraton, A., & Taylor, L. (2007). Qualitative research methods in language test development and validation. In: J. Fox, D. Bayliss, & M. Wesche (Eds.). *Language Testing Reconsidered*. (pp. 113-129). University of Ottawa Press.

Lea, S. J., Stephenson, D., & Troy, J. (2003). Higher education students' attitudes to student-centred learning: Beyond 'educational bulimia'? *Studies in Higher Education, 28*(3), 321-334. https://doi.org/10.1080/03075070309293

Leahy, S., Lyon, C.J., Thompson, M., & Wiliam, D. (2005). Classroom assessment: minute by minute, day by day. *Educational Leadership, 63*(3), 18–24.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369-377. URL: http://www.jstor.org/stable/3588485

Levis, J. M. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics, 27*, 184–202. doi: 10.1017/S0267190508070098

Levis, J.M. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation, 6*(3), 310 – 328. DOI: https://doi.org/10.1075/jslp.20050.lev

Levis, J. M., & McCrocklin, S. (2018). Reflective and effective teaching of pronunciation. In: M. Zeraatpishe, A. Faravani, H. R. Kargozari, & M. Azarnoosh (Eds.), *Issues in applying SLA theories toward reflective and effective teaching.* (pp. 77–89). Brill. https://doi.org/10.1163/9789004380882_007

Levis, J. M., & Sonsaat, S. (2017). Pronunciation teaching in the early CLT era. In O. Kang, R. Thomson, & J. Murphy (Eds.), T*he Routledge handbook of English pronunciation.* (pp. 267–283). Routledge.

Lim, B. (2014). Creating knowledgeable students through critical thinking. *New Straits Times, 30*.

Ling, L. E. (2020). English in Singapore and Malaysia: Differences and similarities. In A. Kirkpatrick (Ed.), *The Routledge Handbook of World Englishes*. (pp. 229-246). Routledge

Little, D. (2005). The Common European Framework and the European Language Portfolio: Involving learners and their judgements in the assessment process. *Language Testing, 22*(3), 321–336. https://doi.org/10.1191/0265532205lt311oa

Little, D., & Perclová, R. (n.d.). *The European Language Portfolio: A guide for teachers and teacher trainers.* https://rm.coe.int/1680459fa6

Liu, S., Niu, Z., & Hao, Y. (2020). Development and construct validation of a diagnostic pronunciation rating scale by many-facet Rasch analysis. *International Journal of TESOL Studies. Vol. 2*(4), 97-109. https://doi.org/10.46451/ijts.2020.12.09

Lombard, K. (2017). Exploring a cohort of South African lecturers' views on the implementation of rubrics in a higher education environment. *Journal for New Generation Sciences, 15* (2), 48-65. doi/epdf/10.10520/EJC-f099cc789

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching, 46*(1), 1– 40. Doi: 10.1017/S0261444812000365

Mahmoodi-Shahrebabaki, M. (2014). Using self-assessment checklists to make English language learners self-directed. *International Journal for Research in Education, 3*(6). 9-20.

Mamoon-Al-Bashir, M., Kabir, M. R., & Rahman, I. (2016). The value and effectiveness of feedback in improving students' learning and professionalizing teaching in higher education. *Journal of Education and Practice, 7*(16), 38-41. ISSN 2222-288X (Online)www.iiste.org

Mathews, E. (2004). New provisions for English language proficiency are expected to improve aviation safety. *ICAO Journal, 59*(1), 4–6.

Mazloomi, S., & Khabiri, M. (2016). Diagnostic assessment of writing through dynamic self-assessment. *International Journal of English Linguistics;6*(6), 19-31. http://dx.doi.org/10.5539/ijel.v6n6p19

McKenney, S. (2013, January 31). *An introduction to educational design research*. Guest Lecture for the Hogeschool Zuyd. University of Twente. https://research.ou.nl/ws/portalfiles/portal/23769063/McKenneyZuid_2013.pdf

McKenney, S., & Reeves, T. (2014). Methods of evaluation and reflection in design research. *Zeitschrift für Berufs- und Wirtschaftspädagogiek*, *27*, 141-153.

McMillan, J. H. (2004). *Classroom assessment: Principles and practice for effective instruction*. Pearson Allyn Brown.

McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, *87*(1), 40–49.

Meddings, L., & Thornbury, S. (2009). *Teaching unplugged*. Delta Publishing

Medina-Díaz, M. del R., & Verdejo-Carrión, A. (2020). Validity and reliability in student learning evaluation throughout active methodologies. *Alteridad, 15*(2), 263-275. https://doi.org/10.17163/alt.v15n2.2020.10

Meihami, H. & Varmaghani, Z. (2013). The implementation of self-assessment in EFL writing classroom: An experimental study. *International Letters of Social and Humanistic Sciences. 9*. 39-48.

Miekley, J. (2005). ESL textbook evaluation checklist. *The Reading Matrix, 5*(2), 9-17.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Sage Publications, Inc.

Ming, T.S. (2009). Investigating autonomy of Malaysian ESL learners: A comparison between public and private universities. *3L: The Southeast Asian Journal of English Language Studies, 15*, 97-124. http://www.ukm.my/smthang/pdf/2009_4.pdf

Ministry of Education Malaysia (2013). Malaysia EDUCATION BLUEPRINT 2013 – 2025 (Preschool to Post-Secondary Education).

https://www.moe.gov.my/en/muat-turun/penerbitan-dan-jurnal/pppm-2013-2025-pendidikan-prasekolah-hingga-lepas-menengah/1207-malaysia-education-blueprint-2013-2025/file

Ministry of Higher Education, M. (2007). *The National Higher Education strategic plan: Laying the foundation beyond 2020*. Putrajaya: Ministry of Higher Education Malaysia

Misbah, N. H., Mohamad, M., Md Yunus, M., & Ya'acob, A. (2017). Identifying the factors contributing to students' difficulties in the English language learning. *Creative Education, 8*, 1999-2008. https://doi.org/10.4236/ce.2017.813136

Miskam, N. N., & Saidalvi, A. (2019). Investigating English language speaking anxiety among Malaysian undergraduate learners. *Asian Social Science, 15*(1). 1-7. https://doi.org/10.5539/ass.v15n1p1

Mizumoto, A., & Takeuchi, O. (2009). Examining the effectiveness of explicit instruction of vocabulary learning strategies with Japanese EFL university students. *Language Teaching Research, 13*(4), 425-449. https://doi.org/10.1177/1362168809341511

Moder, C. L. (2014). Aviation English. In B. Paltridge, & S. Starfield (Eds.), *The handbook of English for specific purposes*. (pp. 227-242). John Wiley & Sons. Pp. 227-242

Moder, C. L., & Halleck, G. (2009). Planes, politics and oral proficiency: Testing international air traffic controllers. *Australian Review of Applied Linguistics, 32*(3), 25.1-25.16. DOI: 10.2104/aral0925

Morse, J. M. and Niehaus, L. (2016). *Mixed method design, principles and procedures*. Routledge.

Moss, G., Goldstein, H., Hayes, S., Chereau, B. M., Sammons, P., Sinnott, G., & Stobart, G. (2021). *High standards, not high stakes: An alternative to SATs that will transform England's testing & school accountability system in primary education & beyond*. British Educational Research Association. https://www.bera.ac.uk/publication/high-standards-not-high-stakesan-alternative-to-sats

Muhamad, A. J., Shah, M. I. A., Ibrahim, E. H. I., Sarudin, I., Malik, F. A., & Ghani, R. A. (2013). Oral presentation errors of Malaysian students in an English for Academic Purposes (EAP) course. *World Applied Sciences Journal 21 (Special Issue of Studies in Language Teaching and Learning).* 19-27. https://doi.org/10.5829/idosi.wasj.2013.21.sltl.2133

Muijs, D., & Bokhove, C. (2020). *Metacognition and self-regulation: Evidence review.* London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/metacognition-and-self-regulation

Mukundan, J., Mohammadi, R., & Nimehchisalem, V. (2011). Developing An English Language Textbook Evaluation Checklist. *Contemporary Issues in Education Research. 4*(6). 21-27. http://dx.doi.org/10.19030/cier.v4i6.4383

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73-97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Murphy, J. M., & Baker, A. A. (2015). History of ESL Pronunciation Teaching. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation*. (pp. 36-65). John Wiley & Sons, Inc.

Nadarajah, J. (2021). Measuring the gap in employability skills among Malaysian graduates. *International Journal of Modern Trends in Social Sciences, 4*(15). 81-87. http://doi 10.35631/ijmtss.415007

Nair, G. K. S., Rahim, R. A., Setia, R., Husin, N., Sabapathy, E., Abd Jalil, N. A., Razlan, R. M., Mohamad, R., So'od, S. M. M., & Yusoff, N. I. M. (2012). Malaysian graduates English adequacy in the job sector. *Asian Social Science, 8*(4), 143-147. http://dx.doi.org/10.5539/ass.v8n4p14

Nakatsuhara, N., May, L., Lam, D., & Galaczi, E. (2018). *Learning oriented feedback in the development and assessment of interactional competence. Research Notes, 70,* 4-67. UCLES/Cambridge University Press.

Narasimhan, S. (2001). Speakers of Dravidian languages: Tamil, Malayalam, Kannada, Telugu. In: M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems*. (2nd edition). (pp. 244 -250). Cambridge University Press.

Navaie, L. A. (2018). The effect of learning-oriented assessment on learning pronunciation among Iranian EFL learners. *International Journal of Education & Literacy Studies, 6*(2), 63-68. http://dx.doi.org/10.7575/aiac.ijels.v.6n.2p.63

Nawai, R., & Said, N. E. M. (2020). Implementation challenges of Common European Framework Reference (CEFR) in a Malaysian setting: Insights on English teachers' attitude. *International Journal of Academic Research in Business and Social Sciences, 10*(7), 28–41. DOI:10.6007/ijarbss/v10-i7/7394

Nelson, T. O, & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In: G. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory*. (pp. 125-173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Neo, M., Neo, T. & Tan, G. X. (2007). A constructivist approach to learning an interactive multimedia course: Malaysian students' perspectives. *Australasian Journal of Educational Technology, 23*(4), 470-489. https://doi.org/10.14742/AJET.1247

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education, 14*, 149–170. https://doi.org/10.1080/09695940701478321

Nieveen, N. (1999). Prototyping to reach product quality. In J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen, & T. Plomp. (Eds.), *Design approaches and tools in education and training*. (pp. 125-137). Kluwer Academic Publishers.

Nimehchisalem, V., Kalajahi, R. S. A., Hussin, M. N., Rafik-Galea, S., Abdullah, A., Rashid, M. S., & Yong, M. (2018). Developing a self-assessment guide for undergraduates' report writing. *Opcion*, *34* (14). 594-634.

Nimehchisalem, V., & Mukundan, J. (2013). Usefulness of the English language teaching textbook evaluation checklist. *Pertanika Journal of Social Sciences and Humanities (JSSH), 21*(2), 797–816.

Nimehchisalem, V., & Mukundan, J. (2015). Refinement of the English language teaching textbook evaluation checklist. *Pertanika Journal of Social Science and Humanities, 23*(4). 761-780.

Nimehchisalem, V., Mukundan, J., Rafik-Galea, S., & Samad, A. A. (2021). Assessment of the analytic scale of argumentative writing (ASAW). *Pertanika J. Soc. Sci. & Hum., 29*(S3), 1-25. dOI: https://doi.org/10.47836/pjssh.29.S3.01

Nimehchisalem, V., Yoong, D., Singh, S., Zainuddin, S. Z., Norouzi, S., & Khalid A. S. (2014). A self-assessment checklist for undergraduate students' argumentative writing. *Advances in Language and Literary Studies. 5*(1*).* 65-80. http://dx.doi.org/10.7575/aiac.alls.v.5n.1p.65

Nissan, S., & Schedl, M. (2012). Prototyping new item types. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing*. (pp. 281-294). Routledge.

Nixon, E. K., & Lee, D. (2001). Rapid prototyping in the instructional design process. *Performance Improvement Quarterly, 14*(3), 95-116. https://doi.org/10.1111/j.1937-8327.2001.tb00220.x

Occupational English Test (2018). *Speaking Assessment Criteria and Level Descriptors (from September 2018) (public version)*. https://prod-wp-content.occupationalenglishtest.org/resources/uploads/2018/08/22102547/speaking-assessment-criteria-updated-2018.pdf

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: debates and practical guidelines. *International Journal of Qualitative Methods, 19*, 1-13. 10.1177/1609406919899220.

O'Donovan, B. M., Price, M., & Rust, C. (2008). Developing student understanding of assessment standards: a nested hierarchy of approaches. *Teaching in Higher Education, 13*(2), 205-217. https://doi.org/10.1080/13562510801923344

Olson, K. (2010). An examination of questionnaire evaluation by expert reviewers. *Field Methods*, 22(4), 295–318. https://doi.org/10.1177/1525822X10379795

O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge University Press.

Osburne, A. G. (2003). Pronunciation strategies of advanced ESOL learners. *International Review of Applied Linguistics in Language, 41*, 131-143. https://doi.org/10.1515/iral.2003.005

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Newbury House Publishers.

Oxford, R. L. (2017). *Teaching and researching language learning strategies: Self-regulation in context*. Routledge.

Oxford, R. L., & Burry-Stock, J. A. (1995). Assessing the use of language learning strategies worldwide with the ESL/EFL version of the Strategy Inventory for Language Learning (SILL). *System, 23*(1), 1–23. doi:10.1016/0346-251x (94)00047-a

Palaganas, E. C., Sanchez, M. C., Molintas, M. P., & Caricativo, R. D. (2017). Reflexivity in qualitative research: A journey of learning. *The Qualitative Report, 22*(2), 426-438. https://doi.org/10.46743/2160-3715/2017.2552

Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual differences, 22*(6), 806-813. https://doi.org/10.1016/j.lindif.2012.04.007

Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology, 11*(2), 551–576. https:// https://doi.org/10.14204/ejrep.30.12200

Panadero, E., Brown, G. T. L., & Strijbos, J. W. (2016a). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review, 28*(4), 803-830. doi:10.1007/s10648-015-9350- 2

Panadero, E., Jonsson, A., & Strijbos, J. W. (2016b). Scaffolding self-regulated learning through self-assessment and peer assessment: guidelines for classroom implementation. In D. Laveault, & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation*. (pp. 311-326). Springer.

Panadero, E., Lipnevich, A. A., & Broadbent, J. (2019). Turning self-assessment into self-feedback. In D. Boud, M. D. Henderson, R. Ajjawi, & E. Molloy (Eds.), *The Impact of Feedback in Higher Education: Improving Assessment Outcomes for Learners*. (pp. 147-164). Springer.

Papadopoulou, I., Kantaridou, Z., Platsidou, M., & Gavriilidou, Z. (2018). The SILL revisited in light of the S2R model of language learning. *The Language Learning Journal, 46*(5), 544–556. https://doi.org/10.1080/09571736.2018.1502739

Pawlak, M. (2010). Designing and piloting a tool for the measurement of the e use of pronunciation learning strategies. *Research in Language, 8*, 189-202.

Pawlak, M., & Szyszka, M. (2018). Researching pronunciation learning strategies: An overview and a critical look. *Studies in Second Language Learning and Teaching, 8*(2), 293-323.

Pennington, M.C., & Rogerson-Revell, P. (2019). *English pronunciation teaching and research: Contemporary perspectives*. Palgrave Macmillan.

Perkins, D. (2008). *Smart schools: From training memories to educating minds*. The Free Press.

Peterson, S. S. (2000). *Pronunciation learning strategies: A first look*. Research report. ERIC.

Phakiti, A. (2008). Construct validation of Bachman and Palmer's (1996) strategic competence model over time in EFL reading tests. *Language Testing. 25*(2). 237–272. https://doi.org/10.1177/0265532207086783

Piccardo, E. (2016). *Common European Framework of Reference for languages: Learning, teaching, assessment*. Phonological scale revision process. Report. Council of Europe

Pillai, S. (2017). Local features of English pronunciation: To embrace or ignore in the ELT classroom?. *J-ELLiT (Journal of English Language, Literature, and Teaching), 1*(1), 1-8. http://dx.doi.org/10.17977/um046v1i1p1-7

Pillai, S., & Ong, L. T. (2018). English(es) in Malaysia. *Asian Englishes*, *20*(2), 147-157. https://doi.org/10.1080/13488678.2018.1459073

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp, & N. Nieveen (Eds.), *Educational Design Research: Part A: An introduction.* (pp. 10-51). SLO.

Pysarchyk, O. L., & Nypadymka, A. S. (2019). Students` pronunciation skills improvement: Peer and self-assessment of pronunciation. *Young Scientist, 5*, 158-161.

Raman, J., Leveson, N., Samost, A. L., Dobrilovic, N., Oldham, M., Dekker, S., & Finkelstein, S. (2016). When a checklist is not enough: How to improve them and what else is needed. *J Thorac Cardiovasc Surg., 152*(2),585-92. doi: 10.1016/j.jtcvs.2016.01.022

Raoofi, S., Chan, S. H., Mukundan, J., & Rashi, S. M. (2014). Metacognition and second/foreign language learning. E*nglish Language Teaching, 7*(1), 36-49. DOI:10.5539/elt.v7n1p36

Rashid, R. A., Abdul Rahman, S. B., & Yunus, K. (2017). Reforms in the policy of English language teaching in Malaysia. *Policy Futures in Education, 15*(1), 100–112. https://doi.org/10.1177/1478210316679069

Ravindran, L., Ridzuan, I., & Wong, B. E. (2022). The impact of social media on the teaching and learning of EFL speaking skills during the COVID-19 pandemic. *MDPI Proceedings, 82*, 38. https://doi.org/10.3390/proceedings2022082038

Reinders, H. & Lewis, M. (2006). An evaluative checklist for self-access materials. *ELT Journal, 60*, 272-278. https://doi.org/10.1093/elt/ccl007

Richey, R. C., Klein, J. D., & Nelson, W. A. (2004). Developmental research: Studies of instructional design and development. In D. H. Jonassen (Ed.), *Handbook of*

research on educational communications and technology. (pp. 1099–1130). Lawrence Erlbaum Associates Publishers.

Richey, R. C., and Klein, J. D. (2005). Developmental research methods: Creating knowledge from instructional design and development practice. *Journal of Computing in Higher Education*, *16*, 23-38. https://dx.doi.org/10.1007/BF02961473

Rickey, D., & Stacy, A. M. (2000). The role of metacognition in learning chemistry. *Journal of Chemical Education, 77*, 915-920. doi:10.1021/ed077p915

Rivas S. F., Saiz, C., & Ossa, C. (2022). Metacognitive strategies and development of critical thinking in higher education. *Front. Psychol. 13*, 1-13. https://doi.org/10.3389/fpsyg.2022.913219

Rose, H. (2019). Dismantling the ivory tower in TESOL: A renewed call for teaching-informed research. *TESOL Quarterly, 53*(3), 895–905. https://doi.org/10.1002/tesq.517

Rose, H., McKinley, J., & Briggs Baffoe-Djan, J. (2020). *Data Collection Research Methods in Applied Linguistics*. Bloomsbury Academic

Ross, J. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment, Research, and Evaluation, 11*, Article 10. DOI: https://doi.org/10.7275/9wph-vv65

Rowlands, K. D. (2007). Check it out! Using checklists to support student learning. *English Journal, 96*(6), 61-66.

Rueda, M. R., Moyano, S., & Rico-Picó, J. (2023). Attention: The grounds of self-regulated cognition. *Wiley interdisciplinary reviews. Cognitive science*, *14*(1), e1582. https://doi.org/10.1002/wcs.1582

Sadler, I., Reimann, N., & Sambell, K. (2023). Feedforward practices: a systematic review of the literature. *Assessment & Evaluation in Higher Education, 48*(3), 305-320. https://doi.org/10.1080/02602938.2022.2073434

Sahin, S. (2020). Developing a checklist for English language teaching course book analysis. *International Journal of Education and Research Vol. 8* (1). 107-120.

Saito, K., Trofimovich, P., Isaacs, T., & Webb, S. (2016). Re-examining phonological and lexical correlates of second language comprehensibility: The role of rater experience. In T. Isaacs, & P. Trofimovich (Eds.), *Second Language Pronunciation Assessment: Interdisciplinary Perspectives.* (pp. 141-156). Multilingual Matters.

Saldaña, J. (2013). *The coding manual for qualitative researchers*. (2nd ed.). SAGE

Salehi, M., & Daryabar, B. (2014). Self-and peer assessment of oral presentations: Investigating correlations and attitudes. *English for Specific Purposes World 15*(42), 1-12. ISSN 1682-3257. Available at http://www.philologician.com/Articles_42/Documents/Salehi.pdf

Sarudin, I., Noor, Z. M., Zubairi, A. M., Ahmad, T. B. T., & Nordin, M. S. (2013). Needs assessment of workplace English and Malaysian graduates' English

language competency. *World Applied Sciences Journal, 21*, 88-94. doi:10.5539/ass.v8n4p143

Schaetzel, K., & Low, E. L. (July, 2009). *Teaching pronunciation to adult English language learners*. CAELA network brief. The Center for Adult English Language Acquisition. https://www.cal.org/caelanetwork/resources/briefs.html

Schmıdt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T.Suthiwan, & I. Walker (Eds.), *Proceedings of ClaSIC* (721-737). https://nflrc.hawaii.edu/PDFs/SCHMIDT%20Attention,%20awareness,%20and%20individual%20differences.pdf

Schmeiser, C. B., & Welch, C. J. (2006). Test development. In R. L. Brennan (Ed.), *Educational measurement* (4th edition). (pp. 307-353). Greenwood.

Schmidt, N., & Wehmeyer, H. (2016). Self-assessment training in the ESL classroom: A crucial step in developing learner autonomy. *Arizona Working Papers in SLA & Teaching*, *23,* 74-97.

Schön, D. A. (1991). *The reflective turn: Case studies in and on educational practice*. Teachers Press, Columbia University.

Schraw, G., & Dennison, R.S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460-475. https://psycnet.apa.org/doi/10.1006/ceps.1994.1033

Seels, B. B., & Richey, R. C. (1994). *Instructional technology: the definition and domains of the field*. Washington: Association for Educational Communications and Technology.

Seidlhofer, B. (2005). English as a lingua franca. *ELT Journal, 59*(4), 339-341. https://doi.org/10.1093/elt/cci064

Şentürk, B. (2017). The effect of different self-assessment tools on students' attitudes towards learning English. [Unpublished doctoral thesis]. Hacettepe University.

Setter, J., & Jenkins, J. (2005). State-of-the-art review article. *Language Teaching, 38*(1), 1-17. doi:10.1017/S026144480500251X

Shackle, C. (2001). Speakers of South Asian languages. In: M. Swan & B. Smith (Eds.). *Learner English: A teacher's guide to interference and other problems*. (2nd edition.) (pp. 227-243). Cambridge University Press

Shak, P., Lee, C. S., & Stephen, J. (2016). Pronunciation problems: A case study on English pronunciation errors of low proficient students. *International Journal of Language Education and Applied Linguistics, 4*, 25-35. https://doi.org/10.15282/ijleal.v4.483

Shakeel, S.I., Al Mamun, M. & Haolader, M. (2023). Instructional design with ADDIE and rapid prototyping for blended learning: validation and its acceptance in the context of TVET Bangladesh. *Education and Information Technologies, 28*, 7601–7630. https://doi.org/10.1007/s10639-022-11471-0

Shelton, K., & Saltsman, g, (2006). Using the Addie model for teaching online. *International journal of information and communication technology education: an official publication of the Information Resources Management Association, 2*, 14-26. DOI:10.4018/jicte.2006070102

Shih, C. M. (2007). A new washback model of students learning. *The Canadian Language Review*, 64(1), 135-161. http://dx.doi.org/10.1080/15434301003664196

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. http://dx.doi.org/10.3102/0034654307313795

Siegesmund, A. (2016). Increasing student metacognition and learning through classroom-based learning communities and self-assessment. *Journal of Microbiology and Biology Education, 17*, 204-214. https://doi.org/10.1128/jmbe.v17i2.954

Sijtsma, K. (2009). Reliability beyond theory and into practice. *Psychometrika, 74*(1), 169-173

Sijtsma, K. (2009b). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika,74*(1), 107–120. https:// doi.org.10.1007/s11336-008-9101-0.

Sikes, P., & Potts, A. (Eds.). (2008). *Researching education from the inside: Investigations from within*. Routledge.

Silitonga, D. H., Sinambela, E., & Pasaribu, A. N. (2021). An analysis of Batu Bara Malay students' pronunciation in English consonant cluster and word stress. *IDEAS Journal of Language Teaching and Learning, Linguistics and Literature, 9*(1), 384-395. https://doi.org/10.24256/ideas.v9i1.1843

Singh, G. K. G., & Singh, S. K. G. (2008). Malaysian graduates' employability skills. *UNITAR E-JOURNAL, 4*(1), 15-45

Sinha, A., Banejee, N., Sinha, A., & Shastri, R. (2009). Interference of first language in the acquisition of second language. *Journal of Psychology and Counseling, 1*(7), 117-122. http://www.academicjournals.org/JPC

Sterling, S., Plonsky, L., Larsson, T., Kytö, M., & Yaw, K. (2023). Introducing and illustrating the Delphi method for applied linguistics research. *Research Methods in Applied Linguistics, 2*(1), 100040. https://doi.org/10.1016/j.rmal.2022.100040

Strachan, L., Kennedy, S., & Trofimovich, P. (2019). Second language speakers' awareness of their own comprehensibility: Examining task repetition and self-assessment. *Journal of Second Language Pronunciation, 5*, 347–373. https://doi.org/10.1075/jslp.18008.st

Sulistyowati, T., Mujianto, J., Rukmini, D., & Hartono, R. (2022). The influence of socio-affective learning and metacognition levels on EFL listening and speaking skills in online learning. *International Journal of Learning, Teaching and Educational Research, 21*(10). https://doi.org/10.26803/ijlter.21.10.24

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: enabling students to make decisions about the quality of work., *Higher Education, 76,* 467–481. https://doi.org/10.1007/s10734-017-0220-3

Tajima, A. (2004). Fatal miscommunication: English in aviation safety. *World Englishes, 23*(3), 451-470. https://doi.org/10.1111/j.0883-2919.2004.00368.x

Tanner, K. D. (2012). Promoting student metacognition. *CBE life sciences education, 11*(2), 113–120. https://doi.org/10.1187/cbe.12-03-0033

Taras, M. (2008) Summative and formative assessment: Perceptions and realities. *Active Learning in Higher Education, 9*(2), 172-192. DOI:10.1177/1469787408091655

Teh, J. L., & Pilus, Z. (2019). International students' perspectives of Malaysian English teachers' spoken English. Indonesian Journal of Applied Linguistics, 8, 554-566. doi: 10.17509/ijal.v8i3.15255

Tessmer, M. (1993). *Planning and conducting formative evaluations.* Routledge.

The University of Auckland: ELE 02/02/12. This checklist is based on: Dunworth, K., & Briguglio, C. (2011). *Teaching students who have English as an additional language: A handbook for academic staff in higher education.* (Appendix B). HERDSA.

Tiew, C. C. & Abdullah, M. N. L. Y. (2021). Heutagogy approach in 21st century teaching and learning: Practices and challenges in Malaysian Higher Education: *AJTLHE, 13*(1), 20-40.

Ting, S-H., Marzuki, E., Chuah, K-M., Misieng, J., & Jerome, C. (2017). Employers' views on the importance of English proficiency and communication skill for employability in Malaysia. *Indonesian Journal of Applied Linguistics, 7*(2), 315-327. doi: dx.doi.org/10.17509/ijal.v7i2.8132

Touchie, H. Y. (1986). Second language learning errors: Their types, causes, and treatment. *JALT Journal, Volume 8*(1), 75-80. https://jalt-publications.org/sites/default/files/pdf-article/art5_8.pdf

Tracey, M. W., Hutchinson, A., & Grzebyk, T. Q. (2014). Instructional designers as reflective practitioners: Developing professional identity through reflection. *Educational Technology Research & Development, 62*(3), 315-334. DOI 10.1007/s11423-014-9334-9

Tracey, M.W., Richey, R.C. ID model construction and validation: a multiple intelligences case. *Education Tech Research Dev 55*, 369–390 (2007). https://doi.org/10.1007/s11423-006-9015-4

Tran, N. T. N. (2021). The relationship between language learning strategies and gender in learning English as a second or foreign language. *Journal of English Language Teaching and Applied Linguistics, 3*(6), 120–126. https://doi.org/10.32996/jeltal.2021.3.6.17

Trofimovich, P, Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: investigating self- and other-perception of second language speech.

Bilingualism: Language and Cognition, 19(01), 122-140. ISSN 1366-7289. Downloaded from: https://eprints.bbk.ac.uk/id/eprint/13310/

Tsagari, D., & Cheng, L. (2017). Washback, impact, and consequences revisited. In E. Shohamy, I. Or, & S. May. (Eds.), *Language testing and assessment: Encyclopedia of language and education*. (pp. 359–372). Springer. https://doi.org/10.1007/978-3-319-02261-1_24

Tsushima, R. (2015). Methodological diversity in language assessment research: The role of mixed methods in classroom-based language assessment studies. *The International Journal of Qualitative Methods, 14*(2), 104-121. https://doi.org/10.1177/160940691501400202

Umida, K., Dilora, A., & Umar, E. (2020). Constructivism in teaching and learning process. *European Journal of Research and Reflection in Educational Sciences 8* (3), 134-137.

Uthayakumaran, A., & Kassim, H. (2018). Student perceptions on the effectiveness of using tell me more for pronunciation learning. *International Journal of Language Education and Applied Linguistics, 8*(2), 83–91. https://doi.org/10.15282/ijleal.v8.1243

Van Den Akker, J. (1999). Principles and methods of development research. In: J. Van Den Akker, N. Nieveen, R.M. Branch, K.L. Gustafson, and T. Plomp (Eds.), *Design approaches and tools in education and training.* (pp. 1-14). Kluwer Academic Publishers.

Vandergrift, L., Goh, C. C. M.; Mareschal, C, J., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning, 56*(3), 431-462. https://doi.org/10.1111/j.1467-9922.2006.00373.x

Vasu, K. (2020). *The effects of self-assessment and teacher feedback on undergraduate students' argumentative writing performance and self-regulated learning. (*Doctoral thesis, Universiti Putra Malaysia). UPM Institutional Repository. http://psasir.upm.edu.my/id/eprint/85434/

Visscher-Voerman, I., Gustafson, K. & Plomp, T. (1999). Educational design and development: An overview of paradigms. In J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen, N., & T. Plomp. (Eds.), *Design approaches and tools in education and training*. (pp. 15-28). Kluwer Academic Publishers.

Vitanova, G., & Miller, A. (2002). Reflective practice in pronunciation learning. *The Internet TESOL Journal, 8*(1). http://iteslj.org/Articles/Vitanova-Pronunciation.html

Vold, E. T. (2018). Using machine-translated texts to generate L3 learners' metalinguistic talk. In Å. Haukås, C. Bjørke, & M. Dypedahl (Eds.), *Metacognition in language learning and teaching*. 9pp. 67-97). Routledge

Wang, X. (2020). Segmental versus suprasegmental: Which one is more important to teach? *RELC Journal, 53*(1), 1-9. https://doi.org/10.1177/0033688220925926

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research, 49*(5), 737-747. doi: 10.1509/jmr.11.0368

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing, 27*(3), 236-247, https://doi.org/10.1016/j.ijresmar.2010.02.004

Wellington, J. (2015). *Educational research: Contemporary issues and practical approaches.* Bloomsbury.

White, C. (2008). Language learning strategies in independent language learning: an overview. In: S. Hurd & T. Lewis (Eds.). *Language learning strategies in independent settings*. (pp. 3-24). Multilingual Matters, 2008. https://doi.org/10.21832/9781847690999

White, C.W., Kroes, M., & Watson, J. (2000). *Aviation maintenance technician training: Training requirements for the 21st century*. Available at https://www.faa.gov/about/initiatives/maintenance_hf/library/documents/media/human_factors_maintenance/aviation_maintenance_technician_training_training_requirements_for_the_21st_century.pdf

White, K. A. (2011). *The development and validation of a tool to measure self-confidence and anxiety in nursing students while making clinical decisions*. UNLV Theses, Dissertations, Professional Papers, and Capstones. 1384. University of Nevada. http://dx.doi.org/10.34917/3276068

Wiliam, D. (2006). Formative assessment: getting the focus right. *Educational Assessment, 11*(3-4), 283-291. https://doi.org/10.1080/10627197.2006.9652993

Winke, P., Gass, S., and Myford, C. (2013). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing, 30*, 231–252. https://doi.org/10.1177/0265532212456968

Wyatt, R. C., & Meyers, L. S. (1987). Psychometric properties of four 5-point Likert type response scales. *Educational and Psychological Measurement, 47,* 27–35.

Xi, X., Higgins, D., Zechner, K., & Williamson, D. M. (2008). Automated scoring of spontaneous speech using SpeechRater? v1.0. Research Report. ETS RR-08-62. *ETS Research Report Series*. http://dx.doi.org/10.1002/j.2333-8504.2008.tb02148.x

Yan, Z. (2016). The self-assessment practices of Hong Kong secondary students: Findings with a new instrument. *Journal of Applied Measurement, 17*(3), 335–353. PMID: 28027056

Yan, Z. (2018). The self-assessment practice scale (SAPS) for students: Development and psychometric studies. *The Asia-Pacific Education Researcher, 27*, 123–135. https://doi.org/10.1007/s40299-018-0371-8

Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education, 45*(2), 224-238. https://doi.org/10.1080/02602938.2019.1629390

Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education, 42*(8), 1247–1262.

Yan, Z. & Carless, D. (2022). Self-assessment is about more than self: the enabling role of feedback literacy. *Assessment & Evaluation in Higher Education. 47* (7), 1116-1128. https://doi.org/10.1080/02602938.2021.2001431

Yates, L. (2017). Learning how to speak: Pronunciation, pragmatics and practicalities in the classroom and beyond. *Language Teaching, 50*(2), 227-246. doi:10.1017/S0261444814000238

Yazid, N. S. M., & Zaiyadi, Z. A. (2017). Pronunciation problems among KUIS students. National Pre University Seminar 2017 (NpreUS2017). RHR Hotel, 23 Ogos 2017. E-ISBN: 978-967-2122-11-1. http://conference.kuis.edu.my/npreus/2017/document/10-Nur%20Syahida%20Mohd%20Yazid.pdf

Yong, J. Y. (2001). Malay/Indonesian speakers. In: M. Swan & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems*. (2nd edition). (pp. 279-295). Cambridge University Press.

Yoshida, M. T. (2018). Choosing technology tools to meet pronunciation teaching and learning goals. *The CATESOL Journal*, *30* (1), 195-212.

Yoshimura, F. (2009). Effects of connecting reading and writing and a checklist to guide the reading process on EFL learners' learning about English writing. *Procedia Social and Behavioral Sciences, 1,* 1871–1883. DOI:10.1016/j.sbspro.2009.01.330

Zakaria, A. H., & Shah, P. M. (2019). Communicative Language Teaching (CLT): Its implementation in teaching English to Malaysian ESL primary learners. *International Journal of Scientific & Engineering Research, 10*(12), 785-799.

Zhang, W., Zhao, M., & Zhu, Y. (2022). Understanding individual differences in metacognitive strategy use, task demand, and performance in integrated L2 speaking assessment tasks. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.876208

Zhao, M. R., Mu, B. L., & Lu, C. P. (2016). Teaching to the test: approaches to teaching in senior secondary schools in the context of curriculum reform in China. *Creative Education*, 7, 32-43. http://dx.doi.org/10.4236/ce.2016.71004

Zhong, W. (2019). Pronunciation rating scale in second language pronunciation assessment: A review *Journal of Language Teaching and Research, 10*, No. (1), 141-149. DOI: http://dx.doi.org/10.17507/jltr.1001.16

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education.* (pp. 311–328). Routledge

APPENDIX A Pronunciation Self-Assessment Checklist v1.0

Dear Student,

This pronunciation self-assessment checklist has been developed to help Bachelor of Aircraft Engineering Technology students assess their own pronunciation in English for speaking activities. The objective of this checklist is for students to prepare and evaluate their pronunciation *before, during* and *after* a speaking activity.

Please answer ALL the questions.

1. Class: _____
2. Gender: Male ☐ Female ☐
3. Latest English examination and result: _____

In the following section, there is a list of criteria that is used to develop a pronunciation self-assessment checklist for a speaking activity. Please circle (1-5) to indicate the frequency level of each criterion according to the key below:

1 = Almost never
2 = Rarely
3 = Sometimes
4 = Often
5 = Almost always

If you think there is a criterion missing, you may add it to the end of the list. Besides that, if there is a term which is difficult to understand, you may add your suggested improvement to this term in the column, *Reword*. If you have any comments about any of the criteria, you can write them in the *Comment* column.

SECTION A: Before the speaking activity

Evaluative criteria	Frequency level	Reword	Comment
1. I practise my pronunciation before a speaking activity.	1 2 3 4 5		
2. When preparing for a speaking activity, I choose words which I can pronounce easily.	1 2 3 4 5		
3. I check on the pronunciation of difficult words before a speaking activity.	1 2 3 4 5		
4. When I am preparing, I pronounce the words clearly in English.	1 2 3 4 5		
5. As part of my preparation, I stress the words accurately in English.	1 2 3 4 5		
6. During preparation, I speak English with a regular rhythm.	1 2 3 4 5		
7. Before a speaking activity, I practise speaking English with a natural intonation.	1 2 3 4 5		
8. I refer to the ICAO Language Proficiency Rating Scale (LPRS) as guidance for my pronunciation for my speaking activities.	1 2 3 4 5		

SECTION B: During the speaking activity

Evaluative criteria	Frequency level	Reword	Comment
1. On the speaking activity day, I am careful when pronouncing the words in English.	1 2 3 4 5		
2. I notice my pronunciation mistakes when I am speaking.	1 2 3 4 5		
3. I self-correct my pronunciation during the presentation.	1 2 3 4 5		
4. I pronounce the words clearly in English.	1 2 3 4 5		
5. I stress the words accurately in English.	1 2 3 4 5		
6. I pay attention to speaking English with a regular rhythm	1 2 3 4 5		
7. I speak English with a natural intonation.	1 2 3 4 5		
	1 2 3 4 5		

SECTION C: After the speaking activity

Evaluative criteria	Frequency level	Reword	Comment
1. I reviewed a recording of my speaking activity for self-improvement.	1 2 3 4 5		
2. I listed down the words I mispronounced.	1 2 3 4 5		
3. I took note of the words that I stressed inaccurately in English.	1 2 3 4 5		
4. I took note of where I spoke with the wrong rhythm.	1 2 3 4 5		
5. I took note of where my intonation caused problems for my listeners.	1 2 3 4 5		
6. I listened to correct examples in English.	1 2 3 4 5		
7. I practised speaking correctly after listening to examples.	1 2 3 4 5		

ICAO Language Proficiency Rating Scale

LEVEL	PRONUNCIATION *Assumes a dialect and/or accent intelligible to the aeronautical community.*
Expert 6	Pronunciation, stress, rhythm, and intonation, though possibly influenced by the first language or regional variation, almost never interfere with ease of understanding.
Extended 5	Pronunciation, stress, rhythm, and intonation, though influenced by the first language or regional variation, rarely interfere with ease of understanding.
Operational 4	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding.
Pre-operational 3	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation and frequently interfere with ease of understanding.
Elementary 2	Pronunciation, stress, rhythm, and intonation are heavily influenced by the first language or regional variation and usually interfere with ease of understanding.
Pre-elementary 1	Performs at a level below the Elementary level.

Source: Manual on the Implementation of ICAO Language Proficiency Requirements, International Civil Aviation Organization (2010)

Extended Guide for Using Pronunciation Self-Assessment Checklist for

Aircraft Engineering Technology Students

This extended guide provides more detailed descriptions and examples of the items in the Pronunciation Self-Assessment Checklist for Aircraft Engineering Technology students. You may find the information helpful before you use the checklist. The explanation covers mostly the possible actions before and after a speaking activity.

1. Description of the measurement scale:

Frequency	Description
1 = Almost never	You are not careful when pronouncing words during a speaking activity. You can make people understand you anyway, maybe by changing words, or gestures or repeating what you say.
2 = Rarely	You attempt to be careful when pronouncing words during a speaking activity, but not for all the words. It could be only 5 words in a 15-minute speaking activity. The word choice would be based on your preference (e.g., difficulty of pronunciation).
3 = Sometimes	You attempt to be careful when pronouncing words during a speaking activity, but not for all the words. It could be only 30 words in a 15-minute speaking activity. The word choice would be based on your preference (e.g., difficulty of pronunciation).
4 = Often	You are careful when pronouncing as many words as possible throughout a speaking activity, regardless of how difficult they are to pronounce, although you still make mistakes.
5 = Almost always	You are careful when pronouncing almost all the words throughout a speaking activity regardless of how difficult they are to pronounce.

2. Definitions

1. Syllable is defined as "a word or part of a word usually containing a vowel sound. For example, 'cheese' has one syllable, 'but-ter' two and 'mar-ga-rine' three. (Cambridge University Press, 2022). https://youtube.com/clip/UgkxQJW7Xj1mU3d4G_H8kuxJJG4h2RwAqQ0R

2. Word stress is when one (or more than one) syllable in a word will be higher in pitch, longer in duration, and generally a little louder than unstressed syllables. https://youtube.com/clip/Ugkxi6aQB9_AAIpNI1a54WJ4qSUntCqSfdYP

3. Rhythm is the sense of movement in speech, marked by the stress, timing, and quantity of a syllable - word or a part of word that only has one vowel sound. https://youtube.com/clip/Ugkx8GEXE7Ygl76sAJpErJhCx-QSzxOaRVjQ

4. Intonation is the way the pitch of your voice goes up and down as you talk. For example, when you are surprised, we can detect your surprised intonation in your voice. https://youtube.com/clip/Ugkx_iJ3nD0pu-4JZ-jTUBAXh6OoK4Co_gYt

Before the speaking activity

Item	Why & How
1. I practise my pronunciation before a speaking activity.	Why? You will have more confidence during a speaking activity. Your listeners will understand you better. How? Online pronunciation website and YouTube Self-recording
2. When preparing for a speaking activity, I choose the words which I can pronounce easily.	Why? This is to ensure your listeners understand you during a speaking activity. This is to give you confidence when explaining during a speaking activity. How? Check vocabulary and pronunciation options online. List the words and synonyms (if necessary) and choose the easiest for you to pronounce, be it new words or words you already know.
3. I check on the pronunciation of difficult words before a speaking activity.	Why? This is to ensure you will explain clearly and confidently during a speaking activity. How? Online pronunciation website and YouTube Online dictionary Refer to your lecturers Practice with your friends
4. When I am preparing, I pronounce the words clearly in English.	Why? This is to ensure you articulate clearly during a speaking activity. You will be able to self-correct yourself during a speaking activity. How? Online pronunciation website and YouTube Online dictionary Refer to your lecturers Practise with your friends
5. As part of my preparation, I stress the words accurately in English	
6. During preparation, I speak English with a regular rhythm.	
7. Before a speaking activity, I speak English with a natural intonation.	
8. I refer to the ICAO Language Proficiency Rating Scale (LPRS) as guidance for my pronunciation for my speaking activities.	Why? The ICAO scale is the standard for the aviation industry where you will work when you finish your degree. How Refer to the rating scale

After the speaking activity

<table>
<tr><th>Item</th><th>Why & How</th></tr>
<tr><td>1. I reviewed a recording of my speaking activity for self-improvement.</td><td>Why?
Reviewing a recording helps you to notice your mistakes and correct them.
How?
Online recording.
Offline speaking activity, record yourself on your handphone or, if necessary, ask a friend to assist you in recording your speaking activity.</td></tr>
<tr><td>2. I listed down the words I mispronounced.
3. I took note of the words that I stressed inaccurately in English.
4. I took note of the words that I spoke with the wrong rhythm.
5. I took note of where my intonation caused problems for my listeners.</td><td>Why?
This to ensure better pronunciation in the future.
This is to avoid repeating the same mistake.
How?
As soon as you have finished your speaking activity, list down as many mistakes as you could remember.
List mistakes when reviewing your speaking activity recordings.</td></tr>
<tr><td>6. I listened to correct examples in English
7. I practised speaking correctly after listening to examples</td><td>Why?
This to ensure you know how to make your pronunciation clearer.
This gives you extra practice before your next speaking activity.
How?
Repeat your speaking activity in your own time and try to notice and self-correct any mistakes.</td></tr>
</table>

-THANK YOU ☺-

APPENDIX B Interview Schedule (Teacher): Design phase

Self-assessment has two main purposes. One is for students to grade themselves to assess what they have already learned, in other words to give themselves a mark. The other is for students to think about where they are now but also about their strengths and weaknesses and what the next steps are for them to improve. This study is all about the second kind of assessment. The idea behind this type of assessment is that, although teachers can teach students, students themselves have a responsibility for their own learning.

Area and question number	Question
Self-assessment	
Q1	Let me start by showing you three statements about self-assessment. Please tell me which one is closest to your own views. 1. Self-assessment helps students to become independent learners; this is a useful skill that can help them in their careers in the future. 2. Self-assessment may or may not be helpful or necessary, depending on how it is used and if there is enough time to do it. 3. Self-assessment is not necessary; teachers provide all the assessment that students need.
Q2	I have a copy of the checklist that I emailed you. Take a few minutes to look at the checklist if you need to, then please tell me your first impressions of it.
How clear are the items listed in the checklist?	
Q3	The checklist looks at pronunciation of whole words, stress, rhythm, and intonation. Do you think this covers all the aspects of pronunciation that it needs to?
Q4	You can see there are three sections (if they haven't already commented on this) – plus the 'assess your proficiency' using the ICAO scale. What is your initial reaction to having these different sections?
Q5	What are your views on having the first stage separate from the other two?
	I'd like to look at each section now to see if everything is clear and complete, or if changes need to be made
'Before' section of checklist	
How appropriate is the language for undergraduate students?	
Please look at the 'before' section and…	
Q6	What do you understand by first item in the 'before' group….? Then the same for all the other items in this stage
Q7	In your view, is anything missing here?
Q8	In your view, is there anything here that does not need to be here?
'During' section	
Q9	Now please have a look at the 'during' section and go through it as you did for the 'before' section

Q10	What do you think about the use of the word ‘during’?
‘After section’	
Q11	Now please do the same for the ‘after’ section
**** How clear is the layout of the checklist?***	
Q12	Are there any other comments you would like to make about this section?
Extended guide	The extended guide is intended to help students complete the checklist, so I would value your opinions on the guide as well.
**** How helpful is the extended guide?***	
Q13	In your view, how helpful are the examples explaining ‘never’, ‘sometimes’ etc?
Q14	How helpful are the explanations of ‘syllable’, ‘rhythm’ and ‘intonation’?
Q15	How helpful is the section on ‘why and how’ do it?
Q16	Any comments on the order?
Q17	Any comments on the layout?
Summary	I’ll recap what you’ve told me so that you can make sure I have got it right
Q18	Any other comments you would like to make?

* Indicates key exploratory question

APPENDIX C Structured Group Interview Questions (Student): Design Phase

INTRODUCTION

I'd like to thank you for being here. You've been asked to take part in this group discussion to help with the design of a self-assessment checklist and the guidance that goes with it.

PURPOSE OF THE DISCUSSION

Our aim today is to be sure the checklist and guide are clear and understandable.

Let me explain what I mean by self-assessment. Self-assessment has two main purposes. One is for students to grade themselves to assess what they have already learned, in other words to give themselves a mark. The other is for students to think about where they are now but also about their strengths and weaknesses and what the next steps are for them to improve. This study is all about the second kind of assessment. The idea behind this type of assessment is that, although teachers can teach students, students have a responsibility for their own learning. We can teach you, but only you can learn.

Topic & Question number	**Question**
Intro	Let me start by showing you three statements about self-assessment. 1. Self-assessment helps students to become independent learners; this is a useful skill that can help them in their careers in the future. 2. Self-assessment may or may not be helpful or necessary, depending on how it is used and if there is enough time to do it. 3. Self-assessment is not necessary; teachers provide all the assessment that students need. Please tell me which one is closest to your own views. I expect you to hear a variety of opinions so let me hear from everyone please.
How do you feel about assessing yourself?	
Checklist items	
What are your first reactions to this checklist?	
Q1	What are your views on the items on the checklist?
Q2	What do you think about having the three stages of 'before' 'during' and 'after'?
Q3	What do you understand by first item in the 'before' group…. Then the same for all the other items in this stage
Q4	In your view, is anything missing here?
Q5	In your view, is there anything here that does not need to be here?
Q6	What do you understand by first item in the 'during' group…. Then the same for all the other items in this stage

Q7	What do you think about the use of the word ‘during’?
Q8	In your view, is anything missing here?
Q9	In your view, is there anything here that does not need to be here?
Q10	What do you understand by first item in the ‘after’ group…. Then the same for all the other items in this stage
Q11	In your view, is anything missing here?
Q12	In your view, is there anything here that does not need to be here?
Q13	Is the number of items in each stage manageable?
Q14	Any suggestions for changing the order?
Q15	Recap students’ suggestions and promote discussion among students to reach agreed list of proposed changes Then ‘Are you agreed that these are the changes you wish to make?’ – and list them
Q16	Any comments on the layout?
Extended guide	
**** How helpful do you think the extended guide will be for students completing the checklist?***	
Q17	How helpful are the examples explaining ‘never’, ‘sometimes’ etc?
Q18	How helpful are the explanations of ‘syllable’, ‘rhythm’ and ‘intonation’? Ask one at a time
Q19	How helpful is the section on ‘why and how’ do it?
Q20	Any comments on the layout?
Q21	Recap students’ suggestions for extended guideline and promote discussion among students to reach agreed list of proposed changes Then ‘Are you agreed that these are the changes you wish to make?’ – and list them
Q22	Any other comments you would like to make?

* Indicates key exploratory question

CLOSING COMMENTS

Thank you very much for your time and contributions today. The information you have given us will be used to revise the checklist and extended guide before they are tried out. Thank you again for your help.

APPENDIX D Pronunciation Self-Assessment Checklist v2.0

Dear Student

This checklist is to help you assess your own pronunciation in English *before, during* and *after* a speaking activity based on the ICAO Language Proficiency Rating Scale.

Along the way, you will find guidance to help you complete the checklist.

You may find it useful to read the guidance for each section before completing the section.

Please answer ALL the questions.

1. Name: _____
2. Class: _____
3. Gender: Male ☐ Female ☐
4. Latest English examination and result: _____

Guidance 1: Read this to help you understand better the ICAO Language Proficiency Rating Scale

LEVEL	PRONUNCIATION *Assumes a dialect and/or accent intelligible to the aeronautical community.*	
Expert 6	Pronunciation, stress, rhythm, and intonation, though possibly influenced by the first language or regional variation, almost never interfere with ease of understanding.	..\..\..\..\Video Clips\Finalised\Level 6.mp4
Extended 5	Pronunciation, stress, rhythm, and intonation, though influenced by the first language or regional variation, rarely interfere with ease of understanding.	..\..\..\..\Video Clips\Finalised\Level 5.mp4
Operational 4	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding.	..\..\..\..\Video Clips\Finalised\Level 4.mp4
Pre-operational 3	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation and frequently interfere with ease of understanding.	..\..\..\..\Video Clips\Finalised\Level 3.mp4
Elementary 2	Pronunciation, stress, rhythm, and intonation are heavily influenced by the first language or regional variation and usually interfere with ease of understanding.	..\..\..\..\Video Clips\Finalised\Level 2.mp4
Pre-elementary 1	Performs at a level below the Elementary level.	..\..\..\..\Video Clips\Finalised\Level 1.mp4

Source: Manual on the Implementation of ICAO Language Proficiency Requirements, International Civil Aviation Organization (2004)

Guidance 2: Word Definitions of the important terms from Guidance 1

1. Syllable is defined as "a word or part of a word usually containing a vowel sound. For example, 'cheese' has one syllable, 'but-ter' two and 'mar-ga-rine' three. (Cambridge University Press, 2022). https://youtube.com/clip/UgkxQJW7Xj1mU3d4G_H8kuxJJG4h2RwAqQ0R

2. Word stress is when one (or more than one) syllable in a word will be higher in pitch, longer in duration, and generally a little louder than unstressed syllables. https://youtube.com/clip/Ugkxi6aQB9_AAIpNI1a54WJ4qSUntCqSfdYP

3. Rhythm is the sense of movement in speech, marked by the stress, timing, and quantity of a syllable - word or a part of word that only has one vowel sound. https://youtube.com/clip/Ugkx8GEXE7Ygl76sAJpErJhCx-QSzxOaRVjQ

4. Intonation is the way the pitch of your voice goes up and down as you talk. For example, when you are surprised, we can detect your surprised intonation in your voice. https://youtube.com/clip/Ugkx_iJ3nD0pu-4JZ-jTUBAXh6OoK4Co_gYt

Guidance 3: Read this to help you better understand Section A: Before the Speaking Activity

Item	Why & How
Before the speaking activity:	
1. I practise my pronunciation.	Why? • You will have more confidence during a speaking activity. How? • Practise with friends or record yourself.
2. I choose words which I can pronounce easily.	Why? • This is to ensure your listeners understand you during a speaking activity. How? • Check vocabulary and pronunciation options online and choose the easiest for you to pronounce.
3. I check on the pronunciation of difficult words.	Why? • This is to ensure you will explain clearly and confidently during a speaking activity. How? • Online dictionary, pronunciation website and YouTube.
4. I pronounce the words clearly in English.	Why? • You will be able to correct yourself during a speaking activity. How? • Online dictionary, pronunciation website and YouTube • Practise with your friends, or record yourself
5. I stress the words accurately in English.	
6. I speak English with a regular rhythm.	
7. I practise speaking English with a natural intonation.	
8. I refer to the ICAO Language Proficiency Rating Scale (LPRS) as guidance for my pronunciation for my speaking activities.	Why? • The ICAO scale is the standard for the aviation industry where you will work when you finish your degree. How • Refer to the rating scale and video clips.

SECTION A: Before the Speaking Activity

Please circle (1-5) to indicate the frequency level of each criterion according to the key given:

1 = Almost never
2 = Rarely
3 = Sometimes
4 = Often
5 = Almost always

Evaluative criteria	Frequency level
Before the speaking activity:	
1. I practise my pronunciation.	1 2 3 4 5
2. I choose words which I can pronounce easily.	1 2 3 4 5
3. I check on the pronunciation of difficult words.	1 2 3 4 5
4. I pronounce the words clearly in English.	1 2 3 4 5
5. I stress the words accurately in English.	1 2 3 4 5
6. I speak English with a regular rhythm.	1 2 3 4 5
7. I practise speaking English with a natural intonation.	1 2 3 4 5
8. I refer to the ICAO Language Proficiency Rating Scale (LPRS) as guidance for my pronunciation for my speaking activities.	1 2 3 4 5

Guidance 4: Use these descriptions for the measurement scale in Section B: During the speaking activity

Frequency	Description
1 = Almost never	You are not careful when pronouncing words during a speaking activity. You can make people understand you anyway, maybe by changing words, or gestures or repeating what you say.
2 = Rarely	You attempt to be careful when pronouncing words during a speaking activity, but not for all the words. It could be only 5 words in a 15-minute speaking activity. The word choice would be based on your preference (e.g., difficulty of pronunciation).
3 = Sometimes	You attempt to be careful when pronouncing words during a speaking activity, but not for all the words. It could be only 30 words in a 15-minute speaking activity. The word choice would be based on your preference (e.g., difficulty of pronunciation).
4 = Often	You are careful when pronouncing as many words as possible throughout a speaking activity, regardless of how difficult they are to pronounce, although you still make mistakes.
5 = Almost always	You are careful when pronouncing almost all the words throughout a speaking activity regardless of how difficult they are to pronounce.

SECTION B: During the speaking activity

Please circle (1-5) to indicate the frequency level of each criterion according to the key given.

1 = Almost never
2 = Rarely
3 = Sometimes
4 = Often
5 = Almost always

Evaluative criteria	Frequency level
During the speaking activity:	
9. I am careful when pronouncing the words in English.	1 2 3 4 5
10. I notice my pronunciation mistakes when I am speaking.	1 2 3 4 5
11. I self-correct my pronunciation whenever I mispronounce.	1 2 3 4 5
12. I try to pronounce each and every word clearly in English.	1 2 3 4 5
13. I stress the words accurately in English.	1 2 3 4 5
14. I pay attention to speaking English with a regular rhythm	1 2 3 4 5
15. I speak English with a natural intonation.	1 2 3 4 5

Guidance 5: Read this to help you better understand Section C: After the Speaking Activity

Item	Why & How
After the speaking activity:	
1. I reviewed a recording of my speaking activity for self-improvement.	Why? • Reviewing a recording helps you to notice your mistakes and correct them. How? • Online recording, handphone, tablet or laptop.
2. I listed down the words I mispronounced.	Why?

3. I took note of the words that I stressed inaccurately in English. 4. I took note of the words that I spoke with the wrong rhythm. 5. I took note of where my intonation caused problems for my listeners.	• This is to avoid repeating the same mistake. How? • List down as many mistakes as you can remember as soon as you have finished your speaking activity, or when reviewing your speaking activity recording.
6. I listened to correct examples of pronunciation in English. 7. I practised speaking correctly after listening to examples of pronunciation in English.	Why? • This to ensure you know how the words should be pronounced and give you extra practice before your next speaking activity. How? • Repeat your speaking activity in your own time and try to notice and self-correct any mistakes.

SECTION C: After the speaking activity

Please circle (1-5) to indicate the frequency level of each criterion according to the key given.

1 = Almost never
2 = Rarely
3 = Sometimes
4 = Often
5 = Almost always

Evaluative criteria	Frequency level
After the speaking activity:	
16. I reviewed the recording of my speaking activity for self-improvement.	1 2 3 4 5
17. I listed down the words I mispronounced.	1 2 3 4 5
18. I took note of the words that I stressed inaccurately in English.	1 2 3 4 5
19. I took note of where I spoke with the wrong rhythm.	1 2 3 4 5
20. I took note of where my intonation caused problems for my listeners.	1 2 3 4 5
21. I listened to correct examples of pronunciation in English.	1 2 3 4 5
22. I practised speaking correctly after listening to examples of pronunciation in English.	1 2 3 4 5

-THANK YOU ☺-

APPENDIX E Interview Guide (Students): Calibration Phase

Topic and question number	Question
Introduction	Let me start by explaining the aim of this interview. I just want to find out about your experience of doing the self-assessment process (the pronunciation checklist). I am interested in what you thought about doing both parts of the checklist (before and after your recent presentation). I am also interested in what you thought about the guide to completing the checklist.
Self-assessment process?	
Q1	Please can you tell me how you feel about the self-assessment process now that you have done it?
Q2	What do you think could be changed to make the explanation better at the start of the process?
Q3	If I ask other classes to try out the checklist, do you think they would need the explanation and support from their teacher?
Instruments	
Q4	Were some parts of the checklist easier to use than others?
Q5	How clear was the guidance given in the checklist?
Q6	How helpful were the videos clips in illustrating the different levels 1,2,3,4,5 and 6 of the ICAO scale?
Classroom learning tool	
Q7	Is the checklist a suitable way for you to self-assess your pronunciation?
Q8	In your opinion, is using google form a suitable way of attempting the checklist in class?
Q9	Should the links be given separately during the class?
Q10	How long did it take you?
Q11	Do you think that's a reasonable amount of time?
Q12	Any other comments you would like to make?

APPENDIX F Interview Guide (Teachers): Calibration Phase

Topic and question number	Question
Introduction	Let me start by explaining the aim of this interview. I just want to find out about your students' experience of doing the self-assessment process (using the pronunciation checklist). I am interested in what you thought about doing both parts of the checklist (before and after your students' presentation). I am also interested in what you thought about the guidance in the checklist.
Self-assessment process?	I'd like to talk about the process first and then go on to talk about the checklist, okay?
Q1	Thinking about the whole process… Statement/screenshare 1 Please choose ONE of the statements which is the closest to your own feelings: 1: I found the self-assessment process easy to do. 2: I found the self-assessment process difficult to do. 3: I found some of the self-assessment process easy to do, and some of the process difficult to do. Please explain your choice. Follow up: 'You said […] . Tell me more about that'. Use 'go on' if they stop halfway and look at me
Q2	Next, I'd like you to think about the briefing session for students at the start of the process. Statement/screenshare 2 Thinking about the briefing session. Choose ONE of the statements which is the closest to your own feelings: A The briefing session was very unhelpful. B The briefing session was unhelpful. C The briefing session was helpful. D The briefing session was very helpful. Please explain your choice. Follow up: If they suggest changes are needed, ask 'What changes would you like to see?'

Topic and question number	Question
	Then 'why do you think that change is important?' If more than one change, ask about each one
Q3	Okay, thank you. Now let's talk about your role. What was your role in helping students through the process of self-assessment? Follow up: If you were given a chance, what would you want to do differently during the process? What are your reasons for wanting to do that differently?
The checklist	Right, now I'd like to move on to talking about the sections of the checklist and the guidance they contain.
Q4	Okay, so think about the various sections of the checklist, the questions and the guidance for each section. By guidance, I mean the definitions of stress, rhythm and intonation at the beginning, the text explaining 'why and how' to do the 'before' and 'after' sections (that's sections A and C), the examples of what 'how often' could mean in the 'during' section (Section B), and the ICAO descriptors and links to the video clips. Have another look at the checklist to remind yourself of the different sections (Screen share Google Forms) Statement/screenshare 3. For the statement below, please say which option-best describes your opinion. Some sections of the checklist were more helpful than others during the self-assessment process. 1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Please explain your choice. Follow up: 'You said […] . Tell me more about that' 'Go on' If more than one section is mentioned, follow them all up.
Q5	(To use if they haven't talked about easy or difficult in the previous question.) Statement/screenshare 4

Topic and question number	Question
	Please choose ONE of the statements which is the closest to your own feelings: Teacher 1: I found all parts of the checklist easy to use with my students. Teacher 2: I found all parts of the checklist difficult to use with my students. Teacher 3: I found some parts of the checklist easy to use, and some parts of the checklist difficult to use with my students. Follow up: What made you choose that one? 'Go on'...'can you say a bit more about that?
Video clips	Let's move on and talk a bit more about the video clips.
Q6	Statement/screenshare 5 For the statement below, please say which option describes your opinion. The video clips clearly illustrated the differences between the six levels of the ICAO scale. 1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Please explain your choice Follow up: 'Go on'. You said […] – tell me more about that.
Future use	Thank you. Now I'd like to ask you some questions about if and how the checklist could be used in future.
Q7	Statement/screenshare 7 Please choose ONE of the statements which is the closest to your own opinion: Teacher 1: The checklist will work best when it is used in the classroom. Teacher 2: The checklist will work best when students use it in their own time. Teacher 3: The checklist will work equally well whether it is used in the classroom or by individual students in their own time.

Topic and question number	Question
	Please explain your choice. Follow up: When you talked about your role in helping students through the self-assessment process, you said […]. How is that affected by what you've just said about when the checklist will work best?
Q8	Statement/screenshare 8 For the statement below, please say which option best describes your opinion. Google Forms is an efficient way of administering the checklist in class. 1 Strongly agree 2 Agree 3 Disagree 4 Strongly disagree Please tell me why you chose that one. Follow up: [if disagrees] What do you think could work better?
Q9	How much classroom time did it take for your students to complete the checklist?
Q10	Do you think that is a reasonable amount of time?
Q11	How could this self-assessment be embedded in the curriculum?
Q12	Any other comments you would like to make?

APPENDIX G Checklist Administration Instructions

BEFORE	
Introduction	Good evening class, we will start our class with Madam Ain's research which she told us about last week. The pronunciation self-assessment checklist will be divided into 2 parts. One will be done before your forum presentation and the other one after your presentation. Additionally, the video clips for the ICAO rating scale that Madam Ain explained to you last week are available on our One Drive. Before we start, I would like to remind you to submit the ethics form via One Drive. If you have submitted, we would like to thank you for your cooperation.
Link	I have given you the link for the first part of the checklist. Please relate your forum presentation task with this research. I would like to remind everyone that using your laptop would be helpful in completing the checklist. Remember there is guidance with links of information to assist you in completing the checklist.
Start	Before we start, are there any questions? Feel free to ask questions as we go along (and teachers will ask as we go along).
Done	Are we done? Madam Ain will check in the system; do we have all the responses we needed? Thank you everyone.
FORUM	We shall now start our Forum presentation.
AFTER	
Link	We have come to the end of our presentation and we shall continue with the pronunciation self-assessment checklist which you will be working on 2 sections. I have given you the link for the second part of the checklist. Like I mentioned just now, please relate your forum presentation task with this research. I would like to remind everyone that using your laptop would be helpful in completing the checklist. Remember there is guidance with links of information to assist you in completing the checklist.
Start	Before we start, are there any questions? Feel free to ask questions as we go along (and teachers will ask as we go along).
End	Are we done?

	Madam Ain will check in the system; do we have all the responses we needed? Thank you everyone.

APPENDIX H Evaluation Questionnaires

EVALUATION OF USEFULNESS QUESTIONNAIRE
FOR PRONUNCIATION SELF-ASSESSMENT CHECKLIST (Students' version)

Dear Student,

This questionnaire is intended to evaluate the usefulness of the Pronunciation Self-Assessment Checklist for Aircraft Engineering Technology Students based on your judgement of its quality.

Background information:

- Class: _____
- Latest English examination and result: _____

For each of the statements below, please circle the level of agreement for that best describes your evaluation of the checklist.

1 Strongly Disagree (SD) 2 Disagree (D) 3 Neutral (N) 4 Agree (A) 5 Strongly Agree (SA)

Item	SD	D	N	A	SA
1. I find it easy to work with the checklist.	1	2	3	4	5
2. I will use this checklist when I prepare for speaking activities.	1	2	3	4	5
3. Using the checklist, I paid more attention to my pronunciation.	1	2	3	4	5
4. The checklist fully covers the aspects of pronunciation I need for speaking activities.	1	2	3	4	5
5. I will encourage other students to use this checklist for a speaking activity.	1	2	3	4	5
6. Because of the checklist, I will have more interest in using the ICAO Language Proficiency Rating Scale as guidance.	1	2	3	4	5
7. All the terms in the checklist are clear and easy to understand.	1	2	3	4	5
8. I found this checklist difficult to work with.	1	2	3	4	5
9. The checklist follows a logical order.	1	2	3	4	5
10. The checklist has made me think about different ways to improve my pronunciation.	1	2	3	4	5
11. This checklist has made me more likely to assess my pronunciation from time to time in accordance with the ICAO Language Proficiency Rating Scale.	1	2	3	4	5
12. I found some of the items in the checklist difficult to understand.	1	2	3	4	5
13. I will use this checklist to review speaking activities.	1	2	3	4	5
14. Pronunciation is clearly defined in this checklist.	1	2	3	4	5
15. There are important elements of pronunciation missing from the checklist.	1	2	3	4	5
16. This checklist can be used with speaking activities in my technical engineering classes.	1	2	3	4	5
17. The checklist is too complicated for me to use.	1	2	3	4	5

Item	SD	D	N	A	SA
18. Overall, the checklist will help me understand the ICAO/aircraft industry pronunciation standard.	1	2	3	4	5
19. The checklist helps me to assess all aspects of my pronunciation (whole word, stress, rhythm, intonation).	1	2	3	4	5
20. Using this checklist makes me worry too much about my pronunciation.	1	2	3	4	5
21. The checklist can be completed quickly.	1	2	3	4	5
22. The checklist is easily accessible through Google Forms.	1	2	3	4	5
23. What changes do you think can be made to the checklist to improve it?					

-THANK YOU ☺-

EVALUATION OF USEFULNESS QUESTIONNAIRE
FOR PRONUNCIATION SELF-ASSESSMENT CHECKLIST (Teachers' version)

Dear Teacher,

This questionnaire is intended to evaluate the usefulness of the Pronunciation Self-Assessment Checklist for Aircraft Engineering Technology Students based on your judgement of its quality.

Background information:

- Male _____ Female _____
- Level of education: BA MA PhD
- Major: _____
- Teaching experience: ____ years

For each of the statements below, please circle the level of agreement that best describes your evaluation of the checklist.

1 Strongly Disagree (SD) 2 Disagree (D) 3 Neutral (N) 4 Agree (A) 5 Strongly Agree (SA)

Item	SD	D	N	A	SA
1. The checklist is easy to work with.	1	2	3	4	5
2. I will use this checklist with my own students.	1	2	3	4	5
3. Using the checklist, students will pay more attention to their pronunciation.	1	2	3	4	5
4. The checklist fully covers the aspects of pronunciation needed for speaking activities.	1	2	3	4	5
5. I will encourage my students to use this checklist.	1	2	3	4	5
6. Because of the checklist, students will have more interest in using the ICAO Language Proficiency Rating Scale as guidance.	1	2	3	4	5
7. All the terms in the checklist are clear and easy to understand.	1	2	3	4	5
8. Students will find this checklist difficult to work with.	1	2	3	4	5
9. The checklist follows a logical order.	1	2	3	4	5
10. The checklist has made me think about different ways my students can improve their pronunciation.	1	2	3	4	5
11. I can fit the time the checklist takes to use into the class time I have available.	1	2	3	4	5
12. I will recommend using this checklist to my students.	1	2	3	4	5
13. There are some important elements of pronunciation missing from the checklist.	1	2	3	4	5
14. Pronunciation is clearly defined in this checklist.	1	2	3	4	5
15. I found some of the items in the checklist difficult to understand.	1	2	3	4	5
16. The checklist can be used with other speaking activities and courses in this university.	1	2	3	4	5
17. The checklist is too challenging for my students to use.	1	2	3	4	5
18. Overall, the checklist will help students to understand the ICAO/aircraft industry pronunciation standard.	1	2	3	4	5
19. The checklist will help students to assess all aspects of my pronunciation (whole word, stress, rhythm, intonation).	1	2	3	4	5
20. The checklist is easily accessible through Google Forms.	1	2	3	4	5
21. What changes do you think can be made to the checklist to improve it?					

-THANK YOU ☺-

APPENDIX I Expert Reviewer Rating Form

Student instructions: In the following sections, students are asked to click on the frequency level of each criterion as follows: 1 = Almost never; 2 = Rarely; 3 = Sometimes; 4 = Often; 5 = Almost always

Expert Reviewer Instructions: For each item please indicate:
A. How relevant the item is to the processes of self-assessment of pronunciation. Please place a checkmark in the appropriate box.
1 = not at all relevant 2 = slightly relevant 3 = moderately relevant 4 = highly relevant
B. Please indicate items which should be removed or reworded by placing a checkmark in the appropriate box. If you wish to suggest rewording, and have time to do so, please use the space in the 'reword' column to do so.

Note: Frequency levels are greyed out to limit visual distraction.

SECTION A: Before the speaking activity

Relevance				Evaluative criteria	Frequency level	Remove	Reword
Before the speaking activity:							
1	2	3	4	1. I practise my pronunciation.	1 2 3 4 5		
1	2	3	4	2. I choose words which I can pronounce easily.	1 2 3 4 5		
1	2	3	4	3. I check on the pronunciation of difficult words.	1 2 3 4 5		
1	2	3	4	4. I pronounce the words clearly in English.	1 2 3 4 5		
1	2	3	4	5. I stress the words accurately in English.	1 2 3 4 5		
1	2	3	4	6. I speak English with a regular rhythm.	1 2 3 4 5		
1	2	3	4	7. I practise speaking English with a natural intonation.	1 2 3 4 5		
1	2	3	4	8. I refer to the ICAO Language Proficiency Rating Scale (LPRS) as guidance for my pronunciation for my speaking activites.	1 2 3 4 5		

SECTION B: During the speaking activity

Relevance				Evaluative criteria	Frequency level	Remove	Reword
During the speaking activity:							
1	2	3	4	9. I am careful when pronouncing the words in English.	1 2 3 4 5		
1	2	3	4	10. I notice my pronunciation mistakes when I am speaking.	1 2 3 4 5		
1	2	3	4	11. I self-correct my pronunciation whenever I mispronounce.	1 2 3 4 5		
1	2	3	4	12. I try to pronounce each and every word clearly in English.	1 2 3 4 5		
1	2	3	4	13. I stress the words accurately in English.	1 2 3 4 5		
1	2	3	4	14. I pay attention to speaking English with a regular rhythm.	1 2 3 4 5		
1	2	3	4	15. I speak English with a natural intonation.	1 2 3 4 5		

SECTION C: After the speaking activity

Relevance				Evaluative criteria	Frequency level	Remove	Reword
After the speaking activity:							
1	2	3	4	16. I reviewed the recording of my speaking activity for self-improvement.	1 2 3 4 5		
1	2	3	4	17. I listed down the words I mispronounced.	1 2 3 4 5		
1	2	3	4	18. I took note of the words that I stressed inaccurately in English.	1 2 3 4 5		
1	2	3	4	19. I took note of where I spoke with the wrong rhythm.	1 2 3 4 5		
1	2	3	4	20. I took note of where my intonation caused problems for my listeners.	1 2 3 4 5		
1	2	3	4	21. I listened to correct examples of pronunciation in English.	1 2 3 4 5		

1	2	3	4	22. I practised speaking correctly after listening to examples of pronunciation in English.	1 2 3 4 5		
1	2	3	4	24. I practised speaking correctly after listening to examples.	1 2 3 4 5		

-THANK YOU ☺-

APPENDIX J Three Selected Transcripts

Student Identifier	GROUP INTERVIEW
	Text
S3	I think the No. 2. I think it depends on the students also. It really depends on students. If they want to do themselves, they can. Some of the student, they do not need self-assessment aaa to… for them.
S1	My name is S1. I think I choose No. 2 because as a student, our job is to learn, and we need guidance to learn and that's where the lecturer's role comes in, you know... because students doing self-assessment may not be accurate because sometimes we don't know where we did wrong. With the help of lecturers, we can know where we did wrong and we can improve ourselves. As of doing the self-assessment alone and not knowing which part we did wrong, so I agree with no. 2 also.
S2	Hello Good Morning, my name is S2. I agree on aaa statement No. 1 where it says it helps students to become independent learner. To me, the self-assessment will not only helps students to become independent learner, it also helps us to further pinpoint which part of our… to determine exactly do we have problems in learning the language, and help us pinpoint which part we need to improve more… later on in our studies
	so I think self-assessment is very helpful and useful, very, very necessary for the students to do the self-assessment, whether or not it is provided by the teachers or they do the SA themselves using other sources.
	Personally, it does help myself a lot to make sure which part should I put my focus on more in order to improve English proficiency. It does help me a lot, helps aaa… in improving skills with the more focus to look on which part needs to be actually improved. I think that's all.
S1	As I said we are students right, our job is to study and we have to learn new stuff, so we might not be correct all the time, so we need the lecturer's help to correct us when something like we are not understanding something so that's where the lecturer's role comes in laa…
S5	Hai. My name is S5. I think the first statement described me the best. Self-assessment helps students to be independent learners.

Student Identifier	GROUP INTERVIEW
	Text
	I think personally, I know that students have their own flaws and it is nice to have their teachers or lecturers to help them doing the self-assessment, but at the same time you know yourself best and you know your weaknesses and you know how to strengthen yourself, so I think by doing self-assessment with myself I made it also useful skills. I get to know what my flaws are, what my strengths are. Yeah I think it is the best.
	Maybe students can do it among friends as well where they do the self-assessment together. I think it is very helpful. Personally, I think it is helpful for myself.
S6	The best statement to describe me is No 2. This is because sometimes self-assessment not working because of some students not understand the topic, some yes, he understand so the poor for the not understand, it may helpful or not helpful. Sometimes some students understand and some students not understand because it hard. That's all.
S7	My name is S7. The thing I am agree with is the statement No. 1. Self-assessment is what students really need to help them boost their skill. Not just using it in the classroom, but also in the future later because without the self-assessment, they wouldn't know how to do the things, how to applicable the things in the outside world because it's really different, in the classroom and with the outside world. The situation is really different so with the self-assessment that they receive in the classroom the can use in the future to help them to improve their skills towards the other human being outside there. Not only to our teachers, but to others person.
S8	Hai, Good Morning. For me, personally I will choose No. 1. Self-assessment will help student to be independent learners.
	This is because when I self-assessed myself, I know that I have like lack of knowledge, I have something that is I don't have. So when I know that I am not capable of doing it or learning it, so with this self-assessment, I can improve myself and seek help from lecturers and friends. This useful skill can help me in the future, for my career.
S2	My initial reaction when I read, I actually read the whole thing. I knew it was made by someone. It was greatly done. It was neatly, emm… organised, and it was a very well done assessment sheet.
	I think about the content is well-chosen, very focused on few specific things about self-assessment. I think

Student Identifier	GROUP INTERVIEW
	Text
S1	I think the self-assessment checklist is really good and contents all the information that students to fill it up. It contains all the information that we need know before we fill the checklist. It is a complete checklist for me.
S8	Hai, S8 here. When I saw the checklist, I'm surprised a bit and went through the checklist, and did my own self-assessment to see where am I at before, during and after, but I don't during and after because I don't know myself too, so I do like self-assess.
	When I see the checklist I do one by one, and see which part am I lack of and I think back about myself and I want to improve more.
S6	OK. Alright. My first impression is I need to improve my pronunciation because in English lot of words same pronunciation but in different meaning. I need to improve my pronunciation in terms of my career soon because it is so dangerous if I say this word but they think I say this word, so ya.
S7	For me, the pronunciation self-assessment is what students really need because not everyone can speak in English very well so their pronunciation self-assessment, they can improve their skills of speaking using right pronunciation. Because as what S6 said, English has the same meaning, same word, same spelling, but different pronunciation with different meaning. With the self-assessment pronunciation, they can help them to speak English very well using the right pronunciation with the right way to use the word.
S2	My name is S2. I think the items in the checklist is enough, but at the same time I think you could add just a little bit more, maybe fill up to 10 in Section A. Section A only stop at no. 8. I mean it won't be a problem if you add up to 10 for Section A, but overall, it is enough. Like I said earlier, it really focus on specific things eee specific things and good enough if it is only this much, but adding won't be a problem I think.
	I do have one suggestion it could like how often do I use a complex word in my speaking, during my presentation, how often do I engage with complex words in my speaking.
S1	Hello. My name is S1. The items is actually ok. You like too many questions will lead people answering the questionnaire is gonna get a little bit boring. The question is really straightforward so is easy for people to understand and answer according to what they believes and I think the No. of questions in every section is adequate enough

Student Identifier	GROUP INTERVIEW
	Text
	and the pronunciation proficiency level is also OK, is also straightforward and not like complex, like people have trouble in understanding it.
	The three sections really does do it's job.
	I think this three sections already represents someone's decision in answering the questions. They already present themselves in these four questions, three sections.
S3	I don't think I have. I am trying to find, but the checklist is good.
	The Section B, during the activity, the question number 15. The question says, not question, but it says I stress the words accurately in English. First, I didn't understand what were stress, but when I scroll down, there's definition. Maybe you can put the stress definition, the syllable definition up top. So maybe the student will know what is the stress, what is intonation maybe. Other than that, it's all good.
S5	Sure. I think everything is good. The sections they all divided really good.
	The amount of questions in each section is perfect, not too much, it's just the right amount. Just like what S1 mentioned earlier, when the people who are self-assessing themselves with this checklist, is because the questions are all very straightforward, and simple, compact and just perfect. I think they won't get bored answering and they won't feel it's too much to answer. I think they are at a good amount of questions and sections.
S7	My point of view, my reaction of aaa… when I see this checklist, I think everything is just complete. Like the question, the use everything is just a simple question, not really like we need a long question because if we use long question, some people wouldn't be interested to answer, to use the items that are prepared. People really need like a simple things so that they can easily answer and they can easily understand the questions that really that are, they received that they need to use. They can use wisely before, during and after session. They can read themselves using each section. Everything is there. Nothing need nothing be adding more or putting out.
S8	Hai, S8 here. For me, the checklist… all the sections is all okay

Student Identifier	GROUP INTERVIEW
	Text
	maybe for the before section, we can add something like: I often speak English. I always speak English at home or in classroom. So that when I self-assess, I know how often I speak English with my friends, my family and when I know that I talk less, I speak English less and I should improve more, for example, if my scale is 1, when during and after, I can see if there is any improvement there.
S2	I think it's well-chosen, and very specific.
S8	I think it's very good that we have scale for our English so that we are able to know which level are we, for example like the expert, moderate.
	We know how we used our grammar and vocabulary, so I think it's very thoughtful to have that scale. When working in industry, we have to speak in English in our daily basis so we have to learn and at least we are able to strengthen vocabulary and grammar to work with people.
S7	I don't have any point on view regarding that.
S2	My name is S2. What I think is when you translate to Google form, the format will be different. I don't think the difference in the format will be affect the arrangement of the checklist, but then… you are transferring the checklist in the Google form, but not attaching, right? Ok. So what I think as for this amount, the Google form will be somewhat lengthy but not too much because the format in Google form, when we try to squeeze everything in 1 page, so it won't look good, it would be lengthy but this much won't be a problem.
	At first I thought you will be going to attach the file and then after doing the checklist, then the participants will have to upload again the file. If you are transferring it directly, then it won't be a problem, but it's going to be lengthy because there's a lot of work.
	I wouldn't say it would be better because not all the participants would agree spending their time and extra effort for that. Sticking to the Google Forms is good enough. And then I want to add about the aaa… when I mentioned about adding another statement to the checklist, I think adding the statement to the checklist won't make it, wouldn't make it as obvious as it is, longer than it already has. If you think of adding extra statement to the checklist, I think that is good as well, so overall I think it's a great idea translating to Google form.

Student Identifier	GROUP INTERVIEW
	Text
	The idea is good enough for this moment.
S6	For me, if you make it in a Google Drive, I think you need to compress the questions because some people may get annoyed because there so many questions and just tick, tick, tick and you will not get accurate information,
	so I think what you can do is you collect the people for the survey and you make a explanation like in our class right now and then you give the google form to them to fill it. If you just spread the link, I think it won't work perfectly.
S3	Ok, this is S3. Does in the google form, does it have the reword and comment? Just the frequency level, right? Then, it's good.
S1	To be honest, I think that is pretty good also.
S5	I think everything is good. I have no other views.
S7	Everything is good. Nothing.
S8	So far so good. No comment.
S3	It's S3. It think it is very helpful if included in the checklist for the Google form. I think the extended guide first. No.
S1	Hai, my name is S1. I agree with S3, put the extended guide first because students will know what to expect if answering. No anything else. We are good.
S6	Yes, for me, just first you put the checklist and then you put the extended. It would be nice if you do it like that. Ya, so after people do the checklist and then they know what they need to improve themselves based on the extended, ya.
S7	Yes.
	I think the extended guidance can either be before or after because for me I doesn't mind either it comes first or it comes later. Either way, we need to answer all the questions provided there. So

Student Identifier	GROUP INTERVIEW
	Text
	ya. I think it should be before, so that students can see that first before proceed on the questions.
S5	I think it should before, they know maybe like where they are at now and then proceed with the next one.
S8	Hai, S8 here. For me, you should put the extended guide first so that students can see the bigger picture and they can see what is that idea of doing this assessment and they can understand
	and they will answer the assessment, not doing it playfully because when they see, especially students when they see a lot of questions, assessment like this, usually students will follow their friends. When we put extended guide first, they will understand and they will answer honestly.
S6	For me, it's already clear. It's good already.
S2	I am S2. I think the content is good enough.
S1	I believe it's true what they said.
S3	So, do you put the before the why and how also?
	In the extended guide?
	OK. I don't see the during, it's just before and after.
S8	Hai, S8 here. I support what they say and I don't have any comment for this. S7?
S7	I don't have any comment about this too. Everything is just okay.
S5	Same goes with me also. I think everything is perfect.
S1	I think it's all already good.
S3	Me too also. Everything is good.
S6	Everything is perfect and good. And All the best for your survey.
S2	I don't have anything particular to add from we have discussed just now, but then I do want to make sure one thing, this checklist is only for MIAT students, right? BAET students, OK. Nothing to add, everything is perfect.
S5	Same. Nothing to add.

Student Identifier	GROUP INTERVIEW
	Text
S7	As what they said, everything is perfect.

T2 DESIGN PHASE
Text
I will go for the first one. Especially when in their tertiary education in the university, they are expected to be independent, so if they are not able to be independent, not be able to live on their own, understanding their mistakes, and how to progress and make changes so that they can improve, then that's going to be an issue for them, er, especially when they, er, when they work in the future.
No I haven't done but we do have like um sort of discussion well you know well when they are listening to my lecture and so on sometimes we do talk about that. , .
Well you know they are able to know your own mistakes or your own messes or whatnot and how they can use that to help them to improvise.
Especially when you are learning a language, right? It's not something you can do for just six months. You have to do it every day so if you don't understand how it works during an English class then I'm not sure if they can - you know - understand the same concepts that need to be applied to different subjects that maybe use English, if you don't understand how it works during an English class. I'm not sure if they can - you know - understand the same concepts that need to be applied to different subjects that maybe use English.
My first impression would be I think they're OK because there are three sections here. You have the first one before, and then the second one for during, and the last one for after so I think it's the good - no not good – yeah, I think it's the best way to actually see and analyze how a student should -uh - see themself and how they can improve - la - by seeing what they do before and what they do during and what they'd do after... I mean if they do whatever they need to do after to see if they can improvise on anything.
I think all of them are quite clear and easy to understand at student level especially the words are simple and straightforward and when there are certain…
or like some specific terms that you need them and everything, there are like explanation to what it actually means to help the student to understand and to be really able to answer the survey properly.

T2 DESIGN PHASE
Text
Except for, I think, the 1,2,3,4,5 there is a list of criteria that is used to develop the almost never, rarely, sometimes, often, always. I think I…this is my opinion...I think like 'almost never' and 'rarely' are kind of the same. And 'often' and 'almost always' are like er...almost the same I think. Like I don't really see the difference between the two like 'almost never', 'rarely'. Yes, the rest is OK, 'often' and 'sometimes' you know. Maybe instead of 'almost never' like 'never' without the 'almost'.
Yes I believe so ..because when we speak we would usually focus on the intonation and then all of the criteria that you have mentioned in the questionnaire or in the checklist - oops - the checklist. I think er not, from what I see it's fine, we do not need like extra questions or extra uh criteria. I believe these are enough uh enough to see on the students' speaking ability - la - through the ability when seeing the criteria.
Order…I think no problem, I think OK. I think it's logical. Yes.
When...you're meaning they have to look at the ICAO proficiency rating scale and then have to answer this according to the rating scale. I think it is correct because - this is my opinion - I think it is important for the students to understand that sometimes what they think they're… I mean when they think there's enough in terms of speaking pronunciation in terms of intonation and all of the other criteria that we have discussed earlier, sometimes they do feel like oh this is OK, right? and then when actually sometimes it is not what is needed in the aviation industry particularly in the aviation industry - la - because you need to be able to achieve a certain level of proficiency to be able to work in the aviation industry.
So when they have the guidance, I mean not just the guidance…when they are able to see the rating and whatnot they are able to see 'oh this is how I can improve, I…so I'm not …I...there are things that I have to improve in order for me to be qualified for the aviation industry' because like I said earlier sometimes they feel like it's OK whatever they have now, as long as people understand me it's OK. But then again sometimes what they feel is OK, it's not OK in the professional industry - la - but I feel…
Because this is just additional, it sometimes…we do mmm…When we do assignment right they don't really – not to say care – pay attention on the importance of pronunciation, because I believe that in Malaysia and also in other countries that speak English they don't really focus on [on on on] the right way of doing things but is it's always like as long as people understand me… but sometimes when people don't understand, you might cause problem in the future,

T2 DESIGN PHASE
Text
so that's why I think it's important for them to know you are not going to speak because you can speak, you're going to speak and if you can speak, you have to be able to understand the accuracy, the right intonation, so that you can deliver your message properly so that there will be like no miscommunication whatsoever.
mmm…I think...I think everything is fine for me. From the way I see it, it's quite simple to understand.
No...mmm…I did have a few questions like well when I'm going through it. Like how do the students…how can the students actually identify if they are speaking with the wrong rhythm or the right rhythm, or how are they going to be able to see that, because if they feel like…like…like the example I gave you just now, they feel like they're OK but to our ears…like…something is wrong, but how do they see that? how do they like analyze that on their own?
Mmm…no. I think all of this is OK. Yeah, I believe so. I think it's OK. For me, everything is quite OK - la - quite OK for me…er…everything is OK for me. I don't see any problems in the layout and what not.
Yes, very helpful, especially for those of the students who might need extra help in comprehending of the criterias and how to decide what's accurate for them.
Mmm…By giving examples…more examples, like for example I will give an example of the checklist and then explain to them if I write 'almost never' and then if I write, if I choose 'almost never', 'rarely' or 'sometimes', what would the end [end end] result be.
mmm…yes, very much, especially there are technical students, there are terms that are, that might not be, that might be alien to them because this is the first time I hear this word. So when you have this, then it will make it easier for them to understand the whole reason of why they are doing this…uh…self-assessment.
I am not sure if it's necessary since you have the app - sorry - when you have explained the meaning of every word or terms. I don't think uh for me I don't think it is necessary because while reading it uh the definition is already clear enough.
mmm…mmmm…yes, in a way, and so like I said they will have a better understanding. If they are confused with their own thoughts like if this is right or whatnot at least they will have like a guidance to help them to decide what's the most accurate to choose. Yes, and also to give like the more accurate result to the self-assessment.

T2 DESIGN PHASE
Text
Yes, I believe so, it's very suitable for undergraduate students. It's not too difficult, not middle, not difficult, it's straightforward, easy to understand.
Yes. I believe, so it is in according to the order, like you start with the frequency 1 to 5 and then definitions, of course we need to know like the definition first before you let them…
but then again, can I know why the extended guide is at the end not at the beginning? I think like before they answer the questions so that...I'm with you - I'm sorry - because if you have to read the question and then you have to look at the extended guide at the bottom you'll be like …some students might not notice there is a guidance at the end or even if they do, they might not - you know - might not look at the extended guide, I think.
I never use Google Forms so I'm not sure but [but but] in my opinion I think it's better for the extended guidance to be put first before the [the the] self-assessment questions, yeah.
I think I don't have anything else to say.

S3 CALIBRATION STAGE
Text
It's OK, I feel very good after I have done it because when I take a look at the questions of before, during and after speaking, the speaking activity, I can reassess and re-evaluate my speaking, intonation, how I speak in rhythm, so I can improve myself in the future when I speak…er…during the speaking activity.
No, everything is clear, 'cos when you click …er…you click the first time before the speaking activity, before we speak we speak in [unclear - our activity?] you give us the…er…er…what you call it? the Section A, so before, we have to answer before [unclear] and then after we speak, you give us the Section B and the section C, so it's very…er…very helpful and very easy to follow, for me at least.
Er, no. For me there's nothing confusing.
You see, I think you can. Yes, you can. Yes, before during and after, I could connect both links, the checklist, the guidance, with my presentation.
Er…I dunno...I think what you give to us, the question is already good, Madam.
So for me…er…there is no need to improve it.
Yeah, I'm happy.
Er…what do you mean by checklist?

S3 CALIBRATION STAGE
Text
I like the before, during and after speaking activities.
Because I can ask about the…what you call it…the students how…like it's a common thing for so many people to do, like before speaking we speak…speak...speak between ourself, like if intonation…so we could…so we can…and then if we ever have word we don't understand we can search it, we can also ask the question about that.
Er…I think you do it inside class with the students. Like…yeah like section A until section C they can look at independently.
For me, no…I need a professional guide.
Yeah, for me I can see the video clips very clear.
I think yes cos it feels…I think everybody has an account from Google, so when the teachers give the link in class, when they click on it they will…it will connect to their account immediately, so they will, they don't have to sign in again, like when I did [indistinct] I always have to sign in...before I get in... before I get in…
Yeah I think it's better for the link to be given separately. If it's a panel discussion it's better but if it's a normal class then just combine.
Yes because before the panel discussion, like we haven't started speaking, so we can evaluate before, what we will be attempting to say, to speak, then after we can look back to ourself, and see if we have done it, like from the checklist, if we have followed how we speak, how is the intonation, the rhythm, like that.
Er…combine…Combine, yeah. Like if we don't have a panel discussion, like if we just have er regular…er regular study, like regular topic, because I don't think you need a speaking activity for you to give this checklist cos er like if we. if students like from like students that study overseas like they will speak to foreign people in English so like during speaking activities if you give it to them they won't need it like they will always speak in English so they can assess it right without the help of a panel discussion or like speaking activity.
Around 4 to 7 minutes for each link.
Er yeah. I think section A is quite fast maybe 3 to 5 minutes.
Yeah, it's more than enough, yeah.
Yeah. It's not that long.

S3 CALIBRATION STAGE
Text
No er I think the checklist is all good you know.

APPENDIX K Instructions to Second Coder

Dear Coder,

Thank you for being part of my PhD research journey.

Description

There are two different sets of documents used in this process which are three transcripts of interviews and a codebook. The three transcripts, which are selected from a total of fifteen, are from one interview with a teacher (labelled as T2), a group interview with eight students (GI), and one interviews with individual students (S3). The transcripts have already been divided for you into chunks of data and are contained in an Excel spreadsheet. Your role as the second coder is to give a code to each of the chunks of data.

The idea of a code is explained by Saldaña (2009:

> "A code in qualitative inquiry is most often a word or short phrase that…assigns a…salient, essence-capturing…attribute for a portion of language-based or visual data. […] Just as a title represents and captures a book or film or poem's primary content and essence, so does a code represent and capture…primary content and essence" (p. 3).

The codes for you to use are shown in the codebook, which contains three code tables. Table 1 shows codes related to specific interview questions that were asked about the design and use of a checklist (structural codes). Table 2 shows the codes based on theories of self-assessment and metacognition that underpin the design of the checklist. Table 3 gives the codes for data which concerns participants' experience of using the checklist but does not fit easily into the other codes.

Process

1. As a second coder, you will use the codes in the codebook to give the most suitable code to each chunk of data. You may find a few chunks of data which are not directly relevant to the research topic, for example a mention of a problem with the internet connection. You do not need to give a code to chunks of data which are not directly relevant.
2. You will code the data for one transcript and then we will meet again to in order to identify any obvious problems with code definitions or interpretations, and to allow for any necessary refinement before you code the other transcripts.
3. Finally, you will then code the other six transcripts without any further contact with me because it is important that your coding is independent.
4. When you have finished, email your spreadsheet complete with codes to namdzulkifly1@sheffield.ac.uk.

APPENDIX L Codebook

Checklist review codes (structural codes)			
THEME	**CODE**	**DEFINITION**	**SAMPLE DATA**
Checklist review	**CR PRO** Checklist review pronunciation	Remarks about whether the concept of pronunciation is complete or incomplete in the checklist: about whether, for instance, the definitions of word stress, rhythm and intonation are clear and follow a logical order or not.	"…when we speak we would usually focus on the intonation and then all of the criteria that you have mentioned in the questionnaire or in the checklist - oops - the checklist… we do not need like extra questions or extra uh criteria. I believe these are enough uh enough to see on the students' speaking ability" (T2)
	CR SEC Checklist review sections	The respondent refers to the number and/or names and/or length and/or order of sections.	"It is easy to follow in other words, because it is divided into sections and we know that the first one is supposed to be for the before, during and after so easier actually to follow" (T1).
	CR ITEM Checklist review items	Remarks about checklist items: whether, for example, there are neither too few nor too many items, item length is not too long or too short, items should be added or removed, the actual wording of items is clear and easy to understand or not, and the order of items is logical or not.	"The items is actually ok. You like too many questions will lead people answering the questionnaire is gonna get a little bit boring. The question is really straightforward so is easy for people to understand" (S1). "It's like about up to [counting] first one, 8… And then 7... I think it is acceptable…because even if you like say there's too much, the students will not read the sentence carefully, but then if there's too little, let's say we just give them 5, it will not reach…it does not cover the whole purpose of this checklist, so I think maybe lower then 10 would be better. For each section. So I mean like more than 5 and less than 10, in other words" (T1).
	CR LAYOUT Checklist review layout	Comments about the layout: about whether, for example, the font and/or spacing of	"What I think is when you translate to Google Form, the format will be different. I don't think the difference in the format will be affect the

		the layout are easy to read or not, and/or about whether there is too much or too little information on each page or not.	arrangement of the checklist, but then… you are transferring the checklist in the Google Form, but not attaching, right? So what I think as for this amount, the Google Form will be somewhat lengthy but not too much because the format in Google Form, when we try to squeeze everything in 1 page, so it won't look good, it would be lengthy but this much won't be a problem" (S2).
	CR GUID Checklist review guidance	Remarks about the guidance: about, for example, the content of the guidance, the external content (videos), the wording of guidance items and instructions, and/or the position of the guidance in relation to the checklist sections.	"I am think about putting the reference before the actual checklist, so that they have an idea on what they should be really looking at instead of seeing the checklist first and then wondering how do I answer this and then after they have answered the checklist and then they go to the next page and by then they will only see the references and they have to go back and do it again according to the reference that they saw…" (S2).
	CR ADMIN Checklist review administration	The interviewee comments on the number of links to sections of the checklist, and how and when to give them to students.	"Yes, I think it [the links] should be given separately to avoid confusion between like the people who will assess themselves, so like when you separate them, they know which one to answer first. So and also they know which one to be submitted first" (S5). "Yeah, I think the two links can be done after the class. I mean after the panel discussion" (S12).
	CR FREQ Frequency scale	The interviewee expresses a view about the number and/or labels of categories on the frequency scales and/or whether the frequency scales are	"I think I…this is my opinion...I think like 'almost never' and 'rarely' are kind of the same. And 'often' and 'almost always' are like er...almost the same I think" (T2).

		easy or difficult to understand and use.	

Codes from theory (metacognition and self-assessment)			
THEME	**CODE**	**DEFINITION**	**SAMPLE DATA**
Metacognitive awareness	**KC** Knowledge of cognition	Comments regarding knowledge about the task and/or its aim, knowledge about himself or herself as a learner and/or what factors influence their performance, and/or knowledge about learning strategies and/or other procedures, and/or knowledge of why and when to use a particular learning strategy.	"…to know that part of myself and I know this assessment is done and I actually get something from it. I know which level I'm at and I know what is my proficiency level and I think it's very good…" (S5). "You know yourself best and you know your weaknesses and you know how to strengthen yourself" (S5). "So when I know that I am not capable of doing it or learning it, so with this self-assessment, I can improve myself and seek help from lecturers and friends" (S8). "I have to take some time, trying to think back, like what do I do if I get some words I don't know how to pronounce or I don't understand words, do I even look it up, maybe, I'm not sure" (S10).
	RC Regulation of cognition	Remarks about planning and/or monitoring and/or evaluating and/or reflecting on what they do or have done, as well as comments on the processes involved in any or all of these.	"When I take a look at the questions of before, during and after speaking, the speaking activity, I can reassess and re-evaluate my speaking, intonation, how I speak in rhythm" (S3). "Before you start the presentation, you do the assessment and then you answer it like this and that but then after you do the presentation you start to look again and then you verify it for yourself. Oh

Codes from theory (metacognition and self-assessment)			
THEME	**CODE**	**DEFINITION**	**SAMPLE DATA**
			I'm not at this level actually" (S11). "Because I know sometimes I don't pronounce the words properly so that when… when I see my proficiency level and for… for the pronunciation. I can see that myself, me myself doing the mistake so that next time whenever I want to… to pronounce the word I can remember the mistake and I'll fix it" (S8). "I think I'm around level 4 to level 5. I wouldn't say 6 because there are some things that I still need to. I still need to improve the mine, proficiency" (S2)
	ME Metacognitive experiences	Remarks about metacognitive feelings, for example a feeling of confidence, and/or about judgements, for example a judgement of their own learning, that relate to their thought processes when working with the checklist, for example realising that other people do not understand them very well when they speak English.	"This statement takes back when I was literally speaking with lecturers and classmate or even with my family whether my accent affects the clarity of my words and sentences" (S2).
Self-assessment	**SAPUR** Purpose of SA	Comments about the purpose of self-assessment.	"Well you know they are able to know your own mistakes or your own messes or whatnot and how they can use that to help them to improvise" (T2).

Codes from theory (metacognition and self-assessment)			
THEME	**CODE**	**DEFINITION**	**SAMPLE DATA**
			"I agree on aaa statement No. 1 where it says it helps students to become independent learner" (S1). "With the self-assessment pronunciation, they can help them to speak English very well using the right pronunciation with the right way to use the word" (S7).
	SAST Standards	Respondents talk about the standards or criteria used by themselves or others to assess pronunciation or speaking, including but not limited to the descriptors and levels of the ICAO scale.	"The pronunciation proficiency level is also OK, is also straightforward and not like complex, like people have trouble in understanding the questions" (S1). "I don't think the students understand the meaning of stress. I know the stress is to be like this for example but the student might think that the stress is different … I'm not sure whether the students' perception, they will understand unless we have exposed them to the rating scale and they understand the meaning of each word written in the ICAO language proficiency rating scale" (T1)

Data-driven codes			
THEME	**CODE**	**DEFINITION**	**SAMPLE DATA**
Roles	**TR** Teacher role	Respondents talk about the actual or desirable roles teachers play, such as instructor, facilitator, assessor, resource, and coach: whether, for	"If let's say they have difficulty in order to assess themselves, the teacher should help them with…in this situation, so in other words when they are

Data-driven codes			
THEME	**CODE**	**DEFINITION**	**SAMPLE DATA**
		example, they provide feedback, make suggestions on how to proceed, or how they use the ICAO scale in their teaching.	assessing themselves, if, let's say, there is a part in which they might need the teacher's help, there are parts in which they can do themselves" (T1).
	TU Teachers' use of SA	Teachers talk about their previous or potential use or non-use of student self-assessment in the classroom.	"That's the reason why I haven't done any self-assessment with students before… No I haven't done but we do have like um sort of discussion well you know well when they are listening to my lecture and so on sometimes we do talk about that" (T1).
	SR Student role	Respondents talk about students' roles as learners: whether in the past, present or future, and whether, for example, they may be relatively passive or active as learners, as indicated by their views and knowledge of self-assessment, as well as their understanding of industry expectations, and their skills in assessing themselves.	"Actually, I didn't even think of asking her (lecturer) for help, really because I felt the form was personal, you have to rate yourself on how often you do things or like that… maybe like if I was still in school or high school, maybe I would have asked my teacher for help, but like right now, I didn't even think of asking her for help, maybe it would have been a good thing to bring up" (S10)
Student understanding	**SUSA** Students' understanding of purpose and process of self-assessment	Students talk about the what, why and how of self-assessment, and why and how self-assessment	"Sometimes self-assessment not working because of some students not understand the topic… Sometimes

Data-driven codes			
THEME	**CODE**	**DEFINITION**	**SAMPLE DATA**
		can work successfully or fail to work.	some students understand and some students not understand because it hard" (S6). "For me, for my opinion they need to be explained again and again because student, when they received the question from the teachers sometimes they just like blindly do it without reading, so they just do it for the sake of I'm doing it, yeah" (S8).

APPENDIX M Teachers Evaluation Questionnaire Tables

Components of test usefulness covered in the teacher questionnaire

Question no.	Item	Component
1	The checklist is easy to work with.	Practicality
2	I will use this checklist with my own students.	Impact
3	Using the checklist, my students will pay more attention to their pronunciation.	Impact
4	The checklist fully covers the aspects of pronunciation needed for speaking activities.	Validity
5	I will encourage my students to use this checklist.	Impact
6	Because of the checklist, students will have more interest in using the ICAO Language Proficiency Rating Scale as guidance.	Impact
7	All the terms in the checklist are clear and easy to understand.	Reliability
8	Students will find this checklist difficult to work with.	Practicality
9	The checklist follows a logical order.	Reliability
10	The checklist has made me think about different ways my students can improve their pronunciation.	Impact
11	I can fit the time the checklist takes to use into the class time I have available.	Practicality
12	I will recommend using this checklist to my students.	Impact
13	There are some important elements of pronunciation missing from the checklist.	Validity
14	Pronunciation is clearly defined in this checklist.	Validity
15	I found some of the items in the checklist difficult to understand.	Reliability
16	The checklist can be used with other speaking activities and courses in this university.	Reliability
17	The checklist is too challenging for my students to use.	Impact
18	Overall, the checklist will help students to understand the ICAO/aircraft industry pronunciation standard.	Impact
19	The checklist will help students to assess all aspects of their pronunciation (whole word, stress, rhythm, intonation).	Validity
20	The checklist is easily accessible through Google Forms.	Practicality

21	What changes do you think can be made to the checklist to improve it?	All

Aspects of impact covered in the teacher questionnaire

Question no.	Item	Aspect of impact
2	I will use this checklist with my own students.	Perceived benefit
3	Using the checklist, my students will pay more attention to their pronunciation.	Perceived benefit
5	I will encourage my students to use this checklist.	End user satisfaction (Affective response)
6	Because of the checklist, students will have more interest in using the ICAO Language Proficiency Rating Scale as guidance.	Perceived benefit
10	The checklist has made me think about different ways my students can improve their pronunciation.	Knowledge of cognition
12	I will recommend using this checklist to my students.	End user satisfaction (Affective response)
17	The checklist is too challenging for my students to use.	End user satisfaction (Affective response)
18	Overall, the checklist will help students to understand the ICAO/aircraft industry pronunciation standard.	Knowledge of cognition

Summary of items by domain showing reversed items (teacher questionnaire)

Domain	**Items (R indicates reversed item, e.g., R of 1 = reversed item of item 1)**
Reliability	7, 9, 15 (R of 7), 16
Validity	4, 13 (R of 4), 14, 19
Impact	2, 3, 5, 6, 10, 12, 17 (R of 5), 18
Practicality	1, 8 (R of 1), 11, 20

APPENDIX N Participant Information Sheet and Consent Form

Participant Information

1. Research Project Title:

Design and Development of Pronunciation Self-Assessment Checklist for Bachelor of Aviation Engineering Technology students

2. Invitation paragraph

You are being invited to take part in a research project for PhD completion by the researcher. Before you decide whether or not to participate, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask the researcher if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

3. What is the project's purpose?

Aviation English lecturers who had previously been involved in research grading students using the International Civil Aviation Organization (ICAO) scale had stated that students could gain from a greater focus on pronunciation. A number of students had similarly expressed a need to focus more on their pronunciation. This project aims to develop a pronunciation self-assessment checklist that Bachelor of Aircraft Engineering Technology (BAET) students can use in an ESL Higher Education context. The self-assessment is purely for students to develop their own ability to assess themselves. Students' self-assessments undertaken as part of the project will not contribute in any way to their final grade on the course.

4. Why have I been chosen?

You have been chosen because you can contribute to the production of a useful pronunciation self-assessment checklist. In total, up to 200 teachers and students will take part in this project. They will be a mixture of Aviation English teachers, other ESL teachers and BAET students.

5. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep. You will also be asked to sign a consent form. Please note that you can withdraw from the project at any time without any negative consequences. For student respondents, if you decide to withdraw, it will not affect your grades. You do not have to give a reason. You should, however, note that if you withdraw after your data has been anonymised and included in data to be analysed, your data cannot be removed, although you can still withdraw. If you wish to withdraw from the research, please contact Madam Nurul Ain Binti Md.Zulkifly at +6012-3112920 or namdzulkifly1@sheffield.ac.uk

6. What will happen to me if I take part? What do I have to do?

By agreeing to take part in the project, you would be agreeing to take part in one or more of:

- completing a questionnaire
- being interviewed for approximately one hour being recorded (audio and / or video)
- participating in a focus group discussion for about two hours per session
- a session to learn about how to use the checklist and completing the checklist itself (around two hours in total)

Please tick the boxes to show which activities you agree to take part in using the checklist:

☐ ☐ ☐ ☐

questionnaire interview focus group discussion Others: _____

There will not be any travel expenses. Each stage of the data collection is relevant to achieving the research project objective. The questionnaire will give the researcher the opportunity to reach a larger group of teachers and students who can evaluate the usefulness of the checklist. The interview sessions and focus group will give the participants involved the opportunity to discuss all aspects of the checklist in depth as well as the opportunity to improve it. Interviews will be semi-structured and interviewees will be given an advance copy of the questions as well as the checklist. Participants in the focus group discussion will also be given a copy of the questions and checklist in advance.

7. What are the possible disadvantages and risks of taking part?

The only commitment is one of time. There are no reasonably foreseeable discomforts, disadvantages or risks attached to taking part in this research. Students' grades and teachers' careers will not be affected by taking part. Any unexpected problems that occur during the research will be brought to your attention immediately. If you experience such a problem, please make contact immediately with Madam Nurul Ain Binti Md.Zulkifly at +6012-3112920 or namdzulkifly1@sheffield.ac.uk.

8. What are the possible benefits of taking part?

Whilst there may not be any immediate benefits for those people participating in the project, it is intended that this work will benefit BAET students and their teachers by encouraging self-assessment of pronunciation in speaking activities. This will help students to develop self-assessment skills and may promote better pronunciation proficiency for classroom purposes. Teachers may benefit from having an additional instructional technique to use with their students.

9. Will my taking part in this project be kept confidential?

All the information that we collect about you during the course of the research will be kept strictly confidential and will only be accessible to members of the research team. You will not be able to be identified in any reports or publications unless you have given your explicit consent for this. If you agree to us sharing the information you provide with other researchers (e.g. by

making it available in a data archive) then your personal details will not be included unless you explicitly request this.

10. What is the legal basis for processing my personal data?

According to data protection legislation, we are required to inform you that the legal basis we are applying in order to process your personal data is that 'processing is necessary for the performance of a task carried out in the public interest' (Article 6(1)(e)). Further information can be found in the University's Privacy Notice https://www.sheffield.ac.uk/govern/data-protection/privacy/general.

11. What will happen to the data collected, and the results of the research project?

The data collected will deposited in ORDA (Online Research Data - the University of Sheffield's data repository) so that it can be used for future research and learning. Due to the nature of this research, it is very likely that other researchers may find the data collected to be useful in answering future research questions. We will ask for your explicit consent for your data to be shared in this way.

12. Who is organising and funding the research?

The research is not funded by any organisation other than the University of Sheffield and Universiti Putra Malaysia who jointly provide supervision and support to the researcher.

13. Who is the Data Controller?

The University of Sheffield will act as the Data Controller for this study. This means that the University is responsible for looking after your information and using it properly.

14. Who has ethically reviewed the project?

This project has been ethically approved via the University of Sheffield's Ethics Review Procedure, as administered by the School of English.

15. What if something goes wrong and I wish to complain about the research or report a concern or incident?

If you are dissatisfied with any aspect of the research and wish to make a complaint, please contact my Supervisor, [Professor Nigel Harwood; n.harwood@sheffield.ac.uk], in the first instance. If you feel your complaint has not been handled in a satisfactory way you can contact the Head of the School of English [Professor Jane Hodson; j.hodson@sheffield.ac.uk]. If the complaint relates to how your personal data has been handled, you can find information about how to raise a complaint in the University's Privacy Notice: https://www.sheffield.ac.uk/govern/data-protection/privacy/general.

and/or

If you wish to make a report of a concern or incident relating to potential exploitation, abuse or harm resulting from your involvement in this project, please contact the project's Designated Safeguarding Contact Supervisor, [Professor Nigel Harwood; n.harwood@sheffield.ac.uk]. If the concern or incident relates to the Designated Safeguarding Contact, or if you feel a report you have made to this Contact has not been handled in a satisfactory way, please contact the Head of the School of English [Professor Jane Hodson; j.hodson@sheffield.ac.uk] and/or the University's Research Ethics & Integrity Manager (Lindsay Unwin; l.v.unwin@sheffield.ac.uk).

16. Will I be recorded, and how will the recorded media be used?

The audio and/or video recordings of your activities made during this research will be used only for analysis of the data collected in this research. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings.

17. Contact for further information

Project contact details for further information:

Nurul Ain Binti Md.Zulkifly	Professor Nigel Harwood
PhD candidate	Supervisor
namdzulkifly1@sheffield.ac.uk	n.harwood@sheffield.ac.uk

You will be given a copy of the information sheet and a signed consent form to keep if you decide to participate in this research.

Thank you for considering participating in this project.

Participant Consent Form

Design and Development of Pronunciation Self-Assessment Checklist for Bachelor of Aviation Engineering Technology Students Consent Form

Please tick the appropriate boxes	Yes	No
Taking Part in the Project		
I have read and understood the project information sheet dated 19/04/2022 or the project has been fully explained to me. (If you answer 'No' to this question, please do not proceed with this consent form until you are fully aware of what your participation in the project will mean.)	☐	☐
I have been given the opportunity to ask questions about the project.	☐	☐
I agree to take part in the project. I understand that taking part in the project will include one or more of: completing a questionnaire, being interviewed, being recorded (audio and / or video, participating in a focus group discussion).	☐	☐
I understand that by choosing to participate as a volunteer in this research, this does not create a legally binding agreement nor is it intended to create an employment relationship with the University of Sheffield.	☐	☐
I understand that my taking part is voluntary and that I can withdraw from the study at any time during the project. I do not have to give any reasons for why I no longer want to take part and there will be no adverse consequences if I choose to withdraw.	☐	☐
How my information will be used during and after the project		
I understand my personal details such as name, phone number, address and email address etc. will not be revealed to people outside the project.	☐	☐
I understand and agree that my words may be quoted in publications, reports, web pages, and other research outputs. I understand that I will not be named in these outputs unless I specifically request this.	☐	☐
I understand and agree that other authorised researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.	☐	☐
I understand and agree that other authorised researchers may use my data in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.	☐	☐
I give permission for the questionnaire, interview, and focus group data (as applicable) that I provide to be deposited in ORDA (Online Research Data - the University of Sheffield's data repository) so that it can be used for future research and learning.	☐	☐
So that the information you provide can be used legally by the researchers		
I agree to assign the copyright I hold in any materials generated as part of this project to The University of Sheffield.	☐	☐

Name of participant [printed]	Signature	Date

Name of Researcher [printed]	Signature	Date
NURUL AIN BINTI MD.ZULKIFLY		

Project contact details for further information:

NURUL AIN BINTI MD.ZULKIFLY	Professor Nigel Harwood	Professor Jane Hodson
PhD candidate	Supervisor	Head of the School of English
namdzulkifly1@sheffield.ac.uk	n.harwood@sheffield.ac.uk	j.hodson@sheffield.ac.uk

APPENDIX O Ethics Approval

Downloaded: 21/05/2022
Approved: 18/05/2022
Nurul Ain Md Zulkifly
Registration number: 130197403
School of English
Programme: PhD Research

Dear Nurul Ain
PROJECT TITLE: Design and Development of Pronunciation Self-Assessment Checklist for Bachelor of Aviation Engineering Technology students
APPLICATION: Reference Number 046390
On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 18/05/2022 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:
University research ethics application form 046390 (form submission date: 17/05/2022); (expected project end date: 26/06/2022).
Participant information sheet 1105086 version 3 (17/05/2022).
Participant consent form 1105087 version 2 (17/05/2022).
If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.
Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely
Michelle Wegrzynska
Ethics Administrator
School of English
Please note the following responsibilities of the researcher in delivering the research project:
The project must abide by the University's Research Ethics Policy:
https://www.sheffield.ac.uk/rs/ethicsandintegrity/ethicspolicy/approval-procedure
The project must abide by the University's Good Research & Innovation Practices Policy:
https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements

APPENDIX P Pronunciation Self-Assessment Checklist v3.0

Dear Student

This checklist is to help you assess your own pronunciation in English *before, during* and *after* a speaking activity based on the the ICAO Language Proficiency Rating Scale.

Along the way, you will find guidance to help you complete the checklist.

You may find it useful to read the guidance for each section before completing the section.

Please answer ALL the questions.

1. Name: _____
2. Class: _____
3. Gender: Male ☐ Female ☐
4. Latest English examination and result: _____

Guidance 1: Read this to help you understand better the ICAO Language Proficiency Rating Scale

LEVEL	PRONUNCIATION *Assumes a dialect and/or accent intelligible to the aeronautical community.*	
Expert 6	Pronunciation, stress, rhythm, and intonation, though possibly influenced by the first language or regional variation, almost never interfere with ease of understanding.	..\..\..\..\Video Clips\Finalised\Level 6.mp4
Extended 5	Pronunciation, stress, rhythm, and intonation, though influenced by the first language or regional variation, rarely interfere with ease of understanding.	..\..\..\..\Video Clips\Finalised\Level 5.mp4
Operational 4	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding.	..\..\..\..\Video Clips\Finalised\Level 4.mp4
Pre-operational 3	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation and frequently interfere with ease of understanding.	\..\..\..\..\Video Clips\Finalised\Level 3.mp4
Elementary 2	Pronunciation, stress, rhythm, and intonation are heavily influenced by the first language or regional variation and usually interfere with ease of understanding.	..\..\..\..\Video Clips\Finalised\Level 2.mp4
Pre-elementary 1	Performs at a level below the Elementary level.	..\..\..\..\Video Clips\Finalised\Level 1.mp4

Source: Manual on the Implementation of ICAO Language Proficiency Requirements, International Civil Aviation Organization (2004)

Guidance 2: Word Definitions of the important terms from Guidance 1

1. Syllable is defined as "a word or part of a word usually containing a vowel sound. For example, 'cheese' has one syllable, 'but-ter' two and 'mar-ga-rine' three. (Cambridge University Press, 2022). https://youtube.com/clip/UgkxQJW7Xj1mU3d4G_H8kuxJJG4h2RwAqQ0R

2. Word stress is when one (or more than one) syllable in a word will be higher in pitch, longer in duration, and generally a little louder than unstressed syllables. https://youtube.com/clip/Ugkxi6aQB9_AAIpNI1a54WJ4qSUntCqSfdYP

3. Rhythm is the sense of movement in speech, marked by the stress, timing, and quantity of a syllable - word or a part of word that only has one vowel sound. https://youtube.com/clip/Ugkx8GEXE7Ygl76sAJpErJhCx-QSzxOaRVjQ

4. Intonation is the way the pitch of your voice goes up and down as you talk. For example, when you are surprised, we can detect your surprised intonation in your voice. https://youtube.com/clip/Ugkx_iJ3nD0pu-4JZ-jTUBAXh6OoK4Co_gYt

Guidance 3: Read this to help you better understand Section A: Before the Speaking Activity

Item	Why & How
Before the speaking activity:	
1. I practise my pronunciation.	Why? • To develop confidence during a speaking activity. How? By practising with friends or listen to my record speech.
2. I choose words which I can pronounce easily.	Why? • To ensure your listeners understand you during a speaking activity. How? • By checking vocabulary and pronunciation options online and choosing the easiest for you to pronounce.
3. I check on the pronunciation of difficult words.	Why? • To ensure you will explain clearly and confidently during a speaking activity. How? • By referring to online dictionaries, pronunciation websites and YouTube.
4. I pronounce the words clearly in English.	Why? • To correct yourself during a speaking activity. How?
5. I stress the words accurately in English.	

Item	Why & How
6. I speak English with a regular rhythm.	• By referring to online dictionaries, pronunciation websites and YouTube. • By practicing with your friends, or record yourself
7. I practise speaking English with a natural intonation.	
8. I refer to the ICAO Language Proficiency Rating Scale (LPRS) as guidance for my pronunciation for my speaking activities.	Why? • To use the language standard set by the ICAO for the aviation industry where you will work when you graduate. How • By referring to the rating scale and video clips.

SECTION A: Before the Speaking Activity

Please circle (1-5) to indicate the frequency level of each criterion according to the key given:

1 = Almost never
2 = Rarely
3 = Sometimes
4 = Often
5 = Almost always

Evaluative criteria	Frequency level
Before the speaking activity:	
1. I practise my pronunciation.	1 2 3 4 5
2. I choose words which I can pronounce easily.	1 2 3 4 5
3. I check on the pronunciation of difficult words.	1 2 3 4 5
4. I pronounce the words clearly in English.	1 2 3 4 5
5. I stress the words accurately in English.	1 2 3 4 5
6. I speak English with a regular rhythm.	1 2 3 4 5
7. I practise speaking English with a natural intonation.	1 2 3 4 5
8. I refer to the ICAO Language Proficiency Rating Scale (LPRS) as guidance for my pronunciation for my speaking activities.	1 2 3 4 5

Guidance 4: Use these descriptions for the measurement scale in Section B: During the speaking activity

Frequency	Description
1 = Almost never	You are not careful with your pronunciation during a speaking activity. You attempt to make people understand you, maybe by changing words, or gestures or repeating what you say.
2 = Rarely	You are hardly careful with your pronunciation during a speaking activity. The word choice would be based on your preference (e.g., easiness of pronunciation).
3 = Sometimes	You attempt to be careful with your pronunciation during a speaking activity. The word choice would be based on your preference (e.g., easiness of pronunciation).
4 = Often	You are careful when pronouncing as many words as possible throughout a speaking activity, regardless of how difficult they are to pronounce, although you still make mistakes.
5 = Almost always	You are careful when pronouncing almost all the words throughout a speaking activity regardless of how difficult they are to pronounce.

SECTION B: During the speaking activity

Please circle (1-5) to indicate the frequency level of each criterion according to the key given.

1 = Almost never
2 = Rarely
3 = Sometimes
4 = Often
5 = Almost always

Evaluative criteria	Frequency level
During the speaking activity:	
23. I am careful when pronouncing the words in English.	1 2 3 4 5
24. I notice my pronunciation mistakes when I am speaking.	1 2 3 4 5
25. I self-correct my pronunciation whenever I mispronounce.	1 2 3 4 5
26. I try to pronounce each and every word clearly in English.	1 2 3 4 5
27. I stress the words accurately in English.	1 2 3 4 5
28. I pay attention to speaking English with a regular rhythm	1 2 3 4 5
29. I speak English with a natural intonation.	1 2 3 4 5

Guidance 4: Read this to help you better understand Section C: After the Speaking Activity

Item	Why & How
After the speaking activity:	
1. I reviewed a recording of my speaking activity for self-improvement.	Why? • To reviewing the recording helps you to notice your mistakes and correct them. How? • By referring to online recording, handphone, tablet or laptop.
2. I listed down the words I mispronounced. 3. I took note of the words that I stressed inaccurately in English. 4. I took note of the words that I spoke with the wrong rhythm. 5. I took note of where my intonation caused problems for my listeners.	Why? • To avoid repeating the same mistake. How? • By listing down as many mistakes as you can remember as soon as you have finished your speaking activity, or when reviewing your speaking activity recording.
6. I listened to correct examples of pronunciation in English. 7. I practised speaking correctly after listening to examples of pronunciation in English.	Why? • To ensure you know how the words should be pronounced and give you extra practice before your next speaking activity. How? • By repeating your speaking activity in your own time and trying to notice and self-correct any mistakes.

SECTION C: After the speaking activity

Please circle (1-5) to indicate the frequency level of each criterion according to the key given.

1 = Almost never
2 = Rarely
3 = Sometimes
4 = Often
5 = Almost always

Evaluative criteria	Frequency level
After the speaking activity:	
30. I reviewed the recording of my speaking activity for self-improvement.	1 2 3 4 5
31. I listed down the words I mispronounced.	1 2 3 4 5
32. I took note of the words that I stressed inaccurately in English.	1 2 3 4 5
33. I took note of where I spoke with the wrong rhythm.	1 2 3 4 5
34. I took note of where my intonation caused problems for my listeners.	1 2 3 4 5
35. I listened to correct examples of pronunciation in English.	1 2 3 4 5
36. I practised speaking correctly after listening to examples of pronunciation in English.	1 2 3 4 5

-THANK YOU ☺-

APPENDIX Q Sample Transcript

Calibration Phase: Semi-structured Interview (S3)

Interviewer: Let me start by explaining the aim of this interview. I just want to find out about your experience of doing the self-assessment process, the pronunciation checklist, yeah that we did yesterday in class. I gave you a link before the panel discussion and another link after the panel discussion. All right um I'm interested in what you thought about doing both parts and also the guide um you know the guidance that I gave for you to read and refer before completing the checklist. All right, yeah?

Respondent: Okay

Interviewer: So now I'm going to start with self-assessment process. Please can you tell me how you feel about the self-assessment process now that you have done it.

Respondent: It's okay, I feel very good after I have done it because when I take a look at the questions of before, during and after speaking, the speaking activity, I can reassess and re-evaluate my speaking, intonation, how I speak in rhythm, so I can improve myself in the future when I speak…er…during the speaking activity.

Interviewer: Was there anything that wasn't clear?

Respondent: No, everything is clear, 'cos when you click …er…you click the first time before the speaking activity, before we speak we speak in our [indistinct] you give us the…er…er…what you call it? the Section A, so before, we have to answer before [unclear] and then after we speak, you give us the Section B and the section C, so it's very…er…very helpful and very easy to follow, for me at least.

Interviewer: Was there anything confusing?

Respondent: Er, no. For me there's nothing.

Interviewer: Could you connect both links with your presentation?

Respondent: You see, I think you can

Interviewer: Sorry, you're breaking

Respondent: Yes, you can. Yes, before during and after, you can just online.

Interviewer: All right, so you could connect both links, the checklist, the guidance, with your presentation.

Respondent: Yeah.

Interviewer: All right? All right. What do you think could be changed to make the explanation better at the start of the process?

Respondent: [thinking] Er…I dunno...I think what you give to us, the question is already good, Madam.

Interviewer: Okay.

Respondent: So for me…er…there is no need to improve it.

Interviewer: All right, you're happy.

Respondent: Yeah, I'm happy.

Interviewer: I always ask that question, if you're happy that's good.

Respondent: Er…what do you mean by checklist?

Interviewer: All right, in the checklist there are three sections.

Respondent: All right, yeah.

Interviewer: So which one is better than the others? Or which part do you like the best?

Respondent: I like the before during and after speaking activities.

Interviewer: Ah you like the before, during and after. Why?

Respondent: Because I can ask about the…what you call it…the students how…like it's a common thing for so many people to do, like before speaking we speak…speak...speak between ourself, like if intonation…so we could…so we can…and then if we ever have word we don't understand we can search it we can also ask the question about that.

Interviewer: All right…so how do I improve? Do you have any suggestions?

Respondent: Er…I think you do it inside class with the students. Like…

Interviewer: mmm…

Respondent: yeah like section A until section C they can look at independently

Interviewer: Okay that's what I've understood. Okay, I understand now. Okay, interesting. Were the video clips clear?

Respondent: Yeah, for me I can see it very clear.

Interviewer: Okay. In your opinion, is using Google Forms a suitable way of doing the checklist in class?

Respondent: I think yes cos it feels…I think everybody has an account from google, so when the teachers give the link in class, when they click on it they will…it will connect to their account immediately, so they will, they don't have to sign in again, like when I did [indistinct] I always have to sign in.

Interviewer: mmm…

Respondent: before I get in... before I get in…

Interviewer All righ.t

Respondent: Yeah

Interviewer: Should the links be given separately? Before and then a discussion and then after.

Respondent: You mean…

Interviewer: Like yesterday, I gave you before then you presented and then after. Should it be done that way in the future processes?

Respondent: Yeah I think it's better. If it's a panel discussion it's better but if it's a normal class then just combine.

Interviewer: Ah...

Respondent: Yes because before the panel discussion, like we haven't started speaking, so we can evaluate before, what we will be attempting to say, to speak, then after we can look back to ourself, and see if we have done it, like from the checklist, if we have followed how we speak, how is the intonation, the rhythm, like that.

Interviewer: All right. You said just now 'in a normal class, we should give one time'. What do you mean by that?

Respondent: Er…combine…

Interviewer: Yes. That's the word, I'm sorry.

Respondent: Combine, yeah.

Interviewer: Combine…if it's a normal class. What do you mean by normal class?

Respondent: Like if we don't have a panel discussion, like if we just have er regular…er regular study, like regular topic.

Interviewer: So you mean when there's no presentation?

Respondent: Yeah.

Interviewer: Okay, so when there's no presentation, just give one link with all the questions and guidance.

Respondent: Yeah, yeah.

Interviewer: So can you do it without having a speaking activity?

Respondent: I think…

Interviewer: So do you understand?

Respondent: Yeah, I think, I think that you can cos...

Interviewer: How? Because, because, sorry

Respondent: „„because I don't think you need a speaking activity for you to give this checklist cos er like if we if students like from like students that study overseas like they will speak to foreign people in English so like during speaking activities if you give it to them they won't need it like they will always speak in English so they can assess it right without the help of a panel discussion or like speaking activity.

Interviewer: so that means any activity so long as you speak in English...

Respondent: yeah.

Interviewer: you can use the checklist

Respondent: Yeah

Interviewer: so for you, like what we had in class, it's okay. Can you connect the panel discussion with both of the links?

Respondent: Yeah

Interviewer: and also for you if there's no speaking activity, you still can do it…

Respondent: yeah.

Interviewer: All the three sections?

Respondent: Yes, A to C

Interviewer: All right. How long did it take for you to complete each link?

Respondent: For each or like combine it?

Interviewer: How long to finish up the first link? Then how long to finish up the second link?

Respondent: Around 4 to 7 minutes.

Interviewer: 4 to 7 minutes for each link.

Respondent: Yeah.

Interviewer: So 4 to 7 minutes. then another 4 to 7 minutes.

Respondent: Er yeah. I think section A is quite fast maybe 3 to 5 minutes.

Interviewer: Do you think that's a reasonable amount of time?

Respondent: Yeah, it's more than enough, yeah.

Interviewer: So it's okay to use that amount of time in class?

Respondent: Yeah. It's not that long.

Interviewer: All right. Lovely. All right, last question. Any other comments you would like to make?

Respondent: No er I think the checklist is all good you know.

Interviewer: All right. Thank you, S3, for helping me.

APPENDIX R ICAO Language Proficiency Rating Scale

Pages A-7 and A-8, *Manual on the Implementation of ICAO Language Proficiency Requirements* (ICAO, 2010).

1.1 Expert, extended and operational levels

LEVEL	*PRONUNCIATION Assumes a dialect and/or accent intelligible to the aeronautical community.*	*STRUCTURE Relevant grammatical structures and sentence patterns are determined by language functions appropriate to the task.*	*VOCABULARY*	*FLUENCY*	*COMPREHENSION*	*INTERACTIONS*
Expert 6	Pronunciation, stress, rhythm, and intonation, though possibly influenced by the first language or regional variation, almost never interfere with ease of understanding.	Both basic and complex grammatical structures and sentence patterns are consistently well controlled.	Vocabulary range and accuracy are sufficient to communicate effectively on a wide variety of familiar and unfamiliar topics. Vocabulary is idiomatic, nuanced, and sensitive to register.	Able to speak at length with a natural, effortless flow. Varies speech flow for stylistic effect, e.g. to emphasize a point. Uses appropriate discourse markers and connectors spontaneously.	Comprehension is consistently accurate in nearly all contexts and includes comprehension of linguistic and cultural subtleties.	Interacts with ease in nearly all situations. Is sensitive to verbal and non-verbal cues and responds to them appropriately.
Extended 5	Pronunciation, stress, rhythm, and intonation, though influenced by the first language or regional variation, rarely interfere with ease of understanding.	Basic grammatical structures and sentence patterns are consistently well controlled. Complex structures are attempted but with errors which sometimes interfere with meaning.	Vocabulary range and accuracy are sufficient to communicate effectively on common, concrete, and work-related topics. Paraphrases consistently and successfully. Vocabulary is sometimes idiomatic.	Able to speak at length with relative ease on familiar topics but may not vary speech flow as a stylistic device. Can make use of appropriate discourse markers or connectors.	Comprehension is accurate on common, concrete, and work-related topics and mostly accurate when the speaker is confronted with a linguistic or situational complication or an unexpected turn of events. Is able to comprehend a range of speech varieties (dialect and/or accent) or registers.	Responses are immediate, appropriate, and informative. Manages the speaker/ listener relationship effectively.
Operational 4	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding.	Basic grammatical structures and sentence patterns are used creatively and are usually well controlled. Errors may occur, particularly in unusual or unexpected circumstances, but rarely interfere with meaning.	Vocabulary range and accuracy are usually sufficient to communicate effectively on common, concrete, and work-related topics. Can often paraphrase successfully when lacking vocabulary in unusual or unexpected circumstances.	Produces stretches of language at an appropriate tempo. There may be occasional loss of fluency on transition from rehearsed or formulaic speech to spontaneous interaction, but this does not prevent effective communication. Can make limited use of discourse markers or connectors. Fillers are not distracting.	Comprehension is mostly accurate on common, concrete, and work- related topics when the accent or variety used is sufficiently intelligible for an international community of users. When the speaker is confronted with a linguistic or situational complication or an unexpected turn of events, comprehension may be slower or require clarification strategies.	Responses are usually immediate, appropriate, and informative. Initiates and maintains exchanges even when dealing with an unexpected turn of events. Deals adequately with apparent misunderstandings by checking, confirming, or clarifying.
Levels 1, 2 and 3 are on subsequent page.						

1.2 Pre-operational, elementary and pre-elementary levels

LEVEL	*PRONUNCIATION Assumes a dialect and/or accent intelligible to the aeronautical community.*	*STRUCTURE Relevant grammatical structures and sentence patterns are determined by language functions appropriate to the task.*	*VOCABULARY*	*FLUENCY*	*COMPREHENSION*	*INTERACTIONS*
Levels 4, 5 and 6 are on preceding page.						
Pre-operational 3	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation and frequently interfere with ease of understanding.	Basic grammatical structures and sentence patterns associated with predictable situations are not always well controlled. Errors frequently interfere with meaning.	Vocabulary range and accuracy are often sufficient to communicate on common, concrete, or work-related topics, but range is limited and the word choice often inappropriate. Is often unable to paraphrase successfully when lacking vocabulary.	Produces stretches of language, but phrasing and pausing are often inappropriate. Hesitations or slowness in language processing may prevent effective communication. Fillers are sometimes distracting.	Comprehension is often accurate on common, concrete, and work- related topics when the accent or variety used is sufficiently intelligible for an international community of users. May fail to understand a linguistic or situational complication or an unexpected turn of events.	Responses are sometimes immediate, appropriate, and informative. Can initiate and maintain exchanges with reasonable ease on familiar topics and in predictable situations. Generally inadequate when dealing with an unexpected turn of events.
Elementary 2	Pronunciation, stress, rhythm, and intonation are heavily influenced by the first language or regional variation and usually interfere with ease of understanding.	Shows only limited control of a few simple memorized grammatical structures and sentence patterns.	Limited vocabulary range consisting only of isolated words and memorized phrases.	Can produce very short, isolated, memorized utterances with frequent pausing and a distracting use of fillers to search for expressions and to articulate less familiar words.	Comprehension is limited to isolated, memorized phrases when they are carefully and slowly articulated.	Response time is slow and often inappropriate. Interaction is limited to simple routine exchanges.
Pre-elementary 1	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.

Note.— The Operational Level (Level 4) is the minimum required proficiency level for radiotelephony communication. Levels 1 through 3 describe Pre-elementary, Elementary, and Preoperational levels of language proficiency, respectively, all of which describe a level of proficiency below the ICAO language proficiency requirement. Levels 5 and 6 describe Extended and Expert levels, at levels of proficiency more advanced than the minimum required Standard. As a whole, the scale will serve as benchmarks for training and testing, and in assisting candidates to attain the ICAO Operational Level (Level 4).

APPENDIX S Typical English Pronunciation Problems Faced by Students

The majority of students at the research site are Bahasa Malaysia speakers, although some are Mandarin and Tamil speakers. Bahasa Malaysia is the official language of Malaysia and is spoken by over eighty percent of the population, while Mandarin and Tamil are the main languages of the Chinese and Indian ethnic communities. However, it is reported that for many people, their first language is a local variety or dialect (Misbah et al., 2017). When Malaysian children start school, they will not necessarily be proficient in Bahasa Melayu, the official national language used in education (Misbah et al., 2017). Moreover, there are two distinct varieties of Bahasa Melayu which are referred to as the 'a' and 'schwa' varieties that differ in the pronunciation of certain word endings (Clynes & Deterding, 2011). In addition, it is said that "a wide range of regional varieties of Malay are spoken alongside the standard varieties, and they are often mutually unintelligible" (Deterding et al., 2022, p. 19). The Chinese ethnic communities also use different regional varieties of Chinese such as Hokkien or Cantonese before starting Mandarin Chinese at school. There are Hindi and Punjabi speakers within the Indian ethnic communities. These differences in linguistic background can all affect the way individuals pronounce individual vowels or consonants/ consonant clusters, which in turn can affect the length of syllables and the stress given to syllables. Rhythm and intonation are also influenced by the pronunciation of syllables and words in addition to prominence. Prominence is defined as "the expression of informational weight within utterances" which can be determined in various languages by word position, by loudness, duration or pitch, or by some combination of these features (Kember et al., 2021, p. 1). The rest of this appendix describes some of the typical pronunciation problems faced by students at the research site. Many of the problems can be attributed to a student's L1, or their L2 or a combination

of these. Differences between English vowels, consonants, consonant clusters and syllables affect whole-word pronunciation. Differences in rhythm and intonation are most likely to be related to differences at the whole-word, phoneme and syllable levels.

Whole word pronunciation

Differences in vowel systems leads to confusion such as those in the present study when 'skill' and 'scale' both being pronounced as /skɪl/ and one fluent speaker who pronounced 'fill' [the checklist] as [fiːl] ('feel'). This type of error is supported by Pillai and Ong (2018) who state that speakers of Bahasa Malaysia do not distinguish between vowel pairs when speaking English, although more fluent speakers "contrast for length" (p. 152). Bahasa Malaysia has only six vowel sounds (Clynes & Deterding, 2011; Deterding et al., 2022). Students also make errors when pronouncing English diphthongs. Examples from conversational English include mispronunciation of 'wait' [weɪt] as 'wet' [wɛt] or 'white' [wʌɪt], and 'take' [teɪk] as [tɛk]. In aviation vocabulary, examples include some students mispronouncing the diphthongs in 'aileron' and 'spoiler', both of which can be pronounced with [ʌɪ] instead of [eɪ] and [ɔɪ] respectively. A study by Kamarudin and Kamal (2021) also noted difficulty with /u:/ and diphthongs /əʊ/ among Bahasa Malaysia speakers. Such errors are reported by Pillai and Ong (2018) and explained by Deterding et al. (2022). Although some scholars have stated that three diphthongs exist in Bahasa Malaysia, others have explained that these are pronounced as two vowels, with the second vowel included in the end of the first, and thus differ from English diphthongs (Deterding et al., 2022, pp. 12-13).

Tamil has five pairs of short and long vowels, but none that correspond closely to the English vowels in cot, coat, caught or pat, pot, part (Narasimhan, 2001, pp. 244-245). Certainly the researcher has heard some students pronouncing the /ɔ/ sound in 'caught' as [kɒt] but has not specifically identified the student as a Tamil speaker. When speaking

English, Tamil speakers may lengthen or shorten vowels in accordance with the pattern found in Dravidian languages (Narasimhan, 2001, p. 245). Like Bahasa Malaysia speakers, they may have difficulties with English diphthongs because there are only two diphthongs in Tamil; English diphthongs may be pronounced as two short vowels with a glide in between (Narasimhan, 2001, p. 245).

Chinese phonology is very different from Bahasa Malaysia and Tamil as well as English phonology (Chang, 2001). Like English, Mandarin has a large vowel system which includes diphthongs, but tones are added to vowels (Huang, n.d.). Mandarin speakers have difficulty in distinguishing between some English vowels and tend to pronounce diphthongs as two very short vowels close together (Chang, 2001, p. 311).

Some students also have difficulties with consonants, consonant clusters and omitted vowel sounds, again due to features of their L1 or L2. Final consonants are sometimes omitted, as in GoogleForm (no 's') and Powerpoin (no 't'); these examples are common in Malaysia, where a word is borrowed from English but pronounced in accordance with the Bahasa Malaysia form of words. The researcher has noticed occasional mistakes with confusing [s] with [z] sounds and [ʃ] with [ʒ]. For example, in the present study there were times when some students pronounced the verb 'pronounce' as [prəʊnaʊnz] rather than [prəʊnaʊns]. Words with consonant clusters at the start or end of the word causes occasional difficulties for some students, who either omit a consonant or articulate all three consonants separately. Two examples are depth and length, both of which some students pronounce as ending in 't'. These examples also illustrate that some students find it difficult to pronounce 'th', such as saying 'trottle' instead of 'throttle'. According to Chang (2021), consonant clusters are a common problem for Chinese speakers. One study reported that seventh-grade Bahasa Malaysia speakers correctly pronounced consonant clusters in around 30% of 50 listed words, with more mistakes in word ending

clusters (Silitonga et al. 2021). Students sometimes insert an extra syllable where they find a consonant sequence difficult, such as *runaway* instead of *runway*.

Other mistakes include omitted vowel sounds, as in 'fuselage' which is often pronounced by Malaysians in the aviation engineering environment as [ˈfjuːzlɪdʒ] rather than [ˈfjuːzəlɑːʒ] (British English) or [ˈfjuːzəlɪdʒ] (American English). Thus the researcher's students hear this mispronounced in engineering classes even after correction by the teacher.

Further difficulties at the whole-word level occur as a result of L1 or L2 differences in syllable stress, meaning that students sometimes stress the wrong syllable. The example already given of 'pronounce' was also incorrectly stressed as well as mispronounced, [ˈprəʊnaʊnz], and sounded exactly the same as 'pronouns', [ˈprəʊnaʊnz]. The same student also stressed the first syllable in [ˈprəʊnaʊnzeɪʃən]. Whole-word mispronunciation sometimes leads to unintelligibility. Indeed, in the present study there were several examples of phrases and even whole sentences when neither the researcher nor the automated transcript of the interview could make sense of what was said, due to the absence of correct whole-word pronunciation.

In Tamil, syllables with long vowels and closed syllables ending in lengthened (geminated) consonants bear the stress (Narasimhan, 2001). This pattern may be carried over into English. In Chinese, there are distinct tones for each vowel sound in words that convey meaning, which may be associated with a tendency for speakers to stress too many syllables in English (Chang, 2021). In a study by Deterding (2011), Bahasa Malaysia speakers exhibited more regular timing of syllables than British speakers. Differences in the stress and length of syllables in the L1 and English are associated with problems with stress, rhythm and intonation faced by speakers of South Asian languages, including Tamil (Shackle, 2001).

Stress, rhythm and intonation

Speakers of South Asian languages are reported to have problems with stress, rhythm and intonation (Shackle, 2001). Shackle (2001) asserts that South Asian languages such as Tamil are syllable-timed, unlike English which is stress-timed (p. 231-232). However, Deterding et al. (2022) state that the division may not be so obvious, and that one language may be more or less syllable-timed than another (p. 17). Regarding Bahasa Malaysia, a study involving acoustic measurements of Malaysian and British speakers led Deterding to state cautiously that Bahasa Malaysia was more syllable-timed than British English (2011). Yong (2021) mentions that small differences in stress in Bahasa Malaysia can sound monotonous when the speaker applies the same variations in stress to English, and this is something the researcher has noticed with a good number of students. Pillai (2017) also links the intonation problems to the lack of vowel contrasts in Bahasa Malaysia. Moreover, the slight differences in stress may be accompanied by one or more breaks in a chunk of information; in the researcher's experience, this further disrupts the listener.

Tamil speakers are said to place stress incorrectly on the first syllable of a word (Shackle, 2001, p. 232). In the researcher's experience, Bahasa Malaysia speakers sometimes do the same, for example saying *ˈcomponent* rather than *comˈponent.* This can result in some cases in the omission of syllables, such as saying *ˈcasher* instead of *caˈshier.* As mentioned above, some Chinese speakers tend to stress too many syllables. This can affect the pronunciation of individual words, for example hangar. Some students stress both syllables in hangar and this leads them to lengthen the first syllable and say [ˈhɑːŋˈgɑː] instead of [ˈhaŋə]. If a student stresses too many syllables in a sentence, it can make it hard for the listener to pick out the information-carrying words. Moreover, Chinese speakers use pitch to distinguish between words, with relatively little variation across a sentence and thus may have problems with intonation in English (Chang, 2001).

The researcher has particularly noticed the difference when Hokkien or Mandarin speakers switch from English into their L1 and back again.

The examples provided here are taken from the classroom as well as the present study. They are in agreement with the literature. However, it is important to remember that the vast majority of the researcher's students only make occasional mistakes, although many would benefit from improving their rhythm and intonation to make it easier for them to convey information in a multilingual environment in the aviation industry.

APPENDIX T Visual Representations of Development

The process of educational design research has been visualized in a variety of ways by different scholars (Plomp, 2013, p. 17). Some emphasize the linear aspect and the timeline or flow, whilst others emphasize the cyclical nature of the process or, in some cases, use a spiral to indicate improvement. Examples of these diagrams, combinations of them, and additional more complex visualizations are given by Plomp (2013) and McKenney (2013). According to Plomp (2013), "all systematic educational and instructional design processes" involve iteration and are "cyclical in character" (p. 17).

In the present study, the process is visualized as mainly linear, moving through three phases of design and development, and indicating the overall flow. However, the process includes formative evaluation in each phase and revisions in the first and third stages. Moreover, in the second phase interviews, participants were asked about the time taken to complete the checklist and whether they considered it reasonable. Their views were confirmed by asking similar questions in the evaluation of usefulness questionnaires. Thus, there were elements of iteration in the essentially linear process.

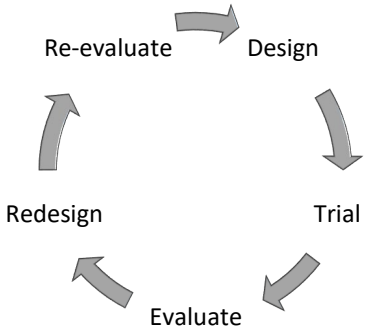
Future development of the study could include a cycle comparing the effectiveness of different types of feedback, as suggested in Section 5.12. This in turn could lead to repeated cycles assessing students' progress with their pronunciation, in which case a spiral could offer a better representation of the process.

The linear, cyclical and spiral versions are illustrated in Figure A-1. The stages in the linear diagram at A-1 (a) are named after the stages in the ADDIE model (Allen, 2006). Each stage could equally well be represented by a cyclical diagram such as the example in A-1 (b) if each stage is re-evaluated and confirmed before moving to the next. The names of cycle stages at A-1 (b) are representative of various development cycles, while

the cycles depicted in A-1 (c) could be development or learning cycles according to the type of improvement. Examples of improvements in the development of a checklist could include guidance for teachers and students on effective feedback or recording improvements in English pronunciation, among others.

(a)

(b)



(c)

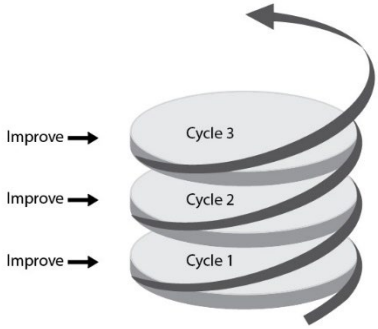


Figure A-1 Visualisations of checklist development

Source: researcher