

**Speaking to be understood: Indonesian
students' perceived and actual
understanding of Indonesian academics'
English speech**

by Douglas Evans

British Council's Master's Dissertation Awards 2020
Commendation

**SPEAKING TO BE UNDERSTOOD: INDONESIAN STUDENTS'
PERCEIVED AND ACTUAL UNDERSTANDING OF
INDONESIAN ACADEMICS' ENGLISH SPEECH**

MA in TESOL (by distance)

Dissertation

XXXXX XXXXX

Submitted: September 2019

Word count: 14,991 words

Abstract

The rapid growth of English-medium instruction (EMI) where English is spoken as a lingua franca (ELF) has resulted in numerous implementation issues, including lack of clarity regarding appropriate standards and training for instructor English language proficiency. This study aimed to provide insights on these issues for the Indonesian and wider ELF context by exploring factors influencing subjective and objective measures of student understanding of English academic speech. Its findings will help alleviate the paucity of ELF-domain research addressing understanding of monologic speech in the high stakes listening context of the EMI classroom.

Employing a quantitative design, this investigation elicited English speech samples from 20 native-speaking Indonesian academics that were used in assessments of perceived and actual understanding by 36 undergraduate students with the same language background. Three English native-speaker language teachers rated the samples for linguistic features. Students' perceptions of speakers as fairly comprehensible did not relate to their ability to objectively understand the academic speech. Higher proficiency speakers (as represented by their IELTS scores) were generally regarded as easier to understand, a finding that did not extend to listeners' actual understanding. Likewise, the better the English language ability of students, the more comprehensible they perceived EMI instructors to be. This difference between the two aspects of understanding was further confirmed by the finding that speakers' phonological errors influenced the perceived understanding, while their use of more diverse and technical vocabulary hindered objectively measured understanding.

One key conclusion drawn from these findings is the need for further development of relevant EMI instructor language proficiency standards in the face of standardized tests' unsuitability in predicting student comprehension. Furthermore, this research calls for the incorporation of pedagogical skill development in EMI educator training to assist in handling of new and unfamiliar lexical items that arise during the course of classroom instruction.

Acknowledgements

I would like to thank Professor XXXXX XXXXX for his supervision throughout this dissertation process. His well-timed encouragement kept me moving forward when I despaired of seeing the light at the end of the tunnel. I also appreciate his connecting me with key resources, including research articles and personnel, to assist me in this endeavour.

There are numerous friends and colleagues whose prayerful exhortations I would also like to acknowledge. Several were even willing to provide hours of their time in contributing to various aspects of this undertaking, including rating speech features, marking comprehension questions, and proofreading. Your input is much appreciated.

I wish to thank participating students at the university where I work for volunteering their time towards this research. I could not have arranged for their involvement without help from one of my colleagues, to whom I am indebted. I am also grateful to the lecturers who so willingly gave time from their busy schedules to contribute to this study. They are the hard-working heroes to whom I wish success as they pursue linguistic and pedagogical excellence in the midst of the maelstrom of life and work.

I could not have completed this two-year journey without my wife XXXXX. Her selfless support with our family, household, and a host of other matters made it possible to devote the required time and effort to see this dissertation through to the end. She also shares the credit for the finished product through her proofreading contributions. I am blessed to have her as my life partner.

Finally, I give thanks to God for providing health of mind and body to undertake this study process. His reminders that I never walk alone strengthened me at many points along this path.

Table of contents

Abstract	1
Acknowledgements	2
Table of contents	3
List of tables	7
List of figures	8
List of acronyms	9
1. Introduction	10
1.1 Background	10
1.2 Motivation for the study	11
1.3 Aims and rationale of the study	12
1.4 Study outline	13
2. Literature review	14
2.1 EMI challenges in Indonesia	14
2.2 Proposed ELF-domain solutions	15
2.3 Measures of L2 speech understanding	15
2.4 Comprehensibility and comprehension research findings	18
2.4.1 L2 speech assessed by NS listeners	18
2.4.1.1 Speaker factors	18
2.4.1.2 Stimulus factors	20
2.4.2 L2 speech assessed by NNS	22
2.4.2.1 Speaker factors	22
2.4.2.2 Stimulus properties	23
2.4.2.3 Listener factors	25
2.5 Research Questions	26
3. Methodology	28
3.1 Study design	28
3.2 Participants	28
3.2.1 Ethical considerations	28
3.2.2 Sampling	29

3.2.3	Speakers.....	29
3.2.3.1	Pilot	29
3.2.3.2	Main study	29
3.2.4	Listeners.....	30
3.2.4.1	Pilot	30
3.2.4.2	Main study	30
3.2.5	EFL teachers.....	31
3.3	Speech elicitation and recording.....	31
3.3.1	Speech elicitation task specifications	31
3.3.2	Interview process	32
3.3.3	Post-recording processing	32
3.4	Measuring comprehensibility and comprehension	33
3.4.1	Pilot.....	33
3.4.2	Main study	33
3.4.3	Comprehensibility ratings.....	34
3.4.4	Comprehension measurement.....	34
3.4.5	Post-tasks questionnaire	34
3.4.6	Post-survey processing	35
3.5	Measuring linguistic features	36
3.5.1	Audio-based measures	37
3.5.2	Transcript-based measures	37
3.5.3	Post-task questionnaire.....	37
4.	Results	39
4.1	Rater consistency.....	39
4.2	Descriptive statistics	40
4.2.1	Comprehensibility and comprehension: Descriptive statistics	40
4.2.2	Comprehensibility ratings per listener	41
4.2.3	Linguistic features: Descriptive statistics.....	41
4.3	Considerations for statistical analyses	43
4.4	RQ1: Do listeners' comprehensibility judgments relate to their assessed comprehension of L2 speech?	44

4.5	RQ2: Do speaker factors (i.e., L2 ability, EMI experience, L2 immersion experience, gender) relate to comprehensibility and comprehension?.....	44
4.5.1	L2 ability.....	45
4.5.2	EMI experience.....	46
4.5.3	L2 immersion experience.....	46
4.5.4	Gender.....	47
4.6	RQ3: Do stimulus properties (i.e., linguistic features, elicitation task, sample length) relate to comprehensibility or comprehension?	48
4.6.1	Linguistic features.....	48
4.6.2	Elicitation task.....	49
4.6.3	Sample length.....	49
4.7	RQ4: Does listener L2 ability relate to comprehensibility and comprehension?	50
4.8	Further analysis of specific cases of comprehensibility / comprehension	53
4.8.1	Most comprehensible.....	53
4.8.2	Least comprehensible.....	54
4.8.3	Best comprehended	55
4.8.4	Least comprehended.....	55
5.	Discussion	57
5.1	Linguistic ratings reliability	57
5.2	RQ1: Do listeners' comprehensibility judgments relate to their assessed comprehension of L2 speech?	57
5.3	RQ2: Do speaker factors (i.e., L2 ability, EMI experience, L2 immersion experience, gender) relate to comprehensibility and comprehension?.....	58
5.3.1	L2 ability.....	59
5.3.2	EMI experience.....	59
5.3.3	L2 immersion experience.....	60
5.3.4	Gender.....	61
5.4	RQ3: Do stimulus properties (i.e., linguistic features, elicitation task, sample length) relate to comprehensibility and comprehension?	61
5.4.1	Linguistic features.....	61
5.4.2	Elicitation task.....	62
5.4.3	Sample length.....	63
5.5	RQ4: Does listener L2 ability relate to comprehensibility and comprehension?	63

6. Conclusion.....	65
6.1 Summary of key findings	65
6.2 Implications for EMI program development.....	65
6.3 Implications for EMI lecturer L2 and pedagogical training.....	66
6.4 Study limitations.....	66
6.5 Areas for further study.....	67
References	68
Appendix A. Speech elicitation: Theoretical and methodological considerations .	79
A.1 Theoretical considerations	79
A.2 Target language usage domain speech task characteristics.....	80
A.3 Target L2 speech ability construct.....	81
Appendix B. Speaker information and consent form.....	82
Appendix C. Speaker interview: Background questions and elicitation prompts..	88
Appendix D. Speech task transcripts.....	90
Appendix E. Listener information and consent form.....	100
Appendix F. Listener survey: Comprehensibility and comprehension instruments and background information questions.....	102
Appendix G. Comprehension question answer key	115
Appendix H. EFL teacher information and consent form.....	118
Appendix I. EFL teacher survey: Audio- and transcript-based linguistic measures and background information questions.....	121
Appendix J. Additional analysis results	139

List of tables

Table 1 <i>Reliability Indices (Cronbach's Alpha) for Rated Linguistic Categories</i>	38
Table 2 <i>Average Comprehensibility Ratings and Comprehension Scores:</i>	
<i>Descriptive Statistics</i>	40
Table 3 <i>Linguistic Categories: Descriptive Statistics</i>	40
Table 4 <i>KSD Normality Tests for Linguistic Measures</i>	41
Table 5 <i>Effect of Speaker IELTS Score on Comprehensibility and</i>	
<i>Comprehension</i>	43
Table 6 <i>Effect of EMI Teaching Experience on Comprehensibility and</i>	
<i>Comprehension</i>	44
Table 7 <i>Effect of L2 Immersion Experience on Comprehensibility and</i>	
<i>Comprehension</i>	45
Table 8 <i>Effect of Speaker Gender on Comprehensibility and Comprehension</i>	45
Table 9 <i>Association of Linguistic Feature Ratings with Comprehensibility and</i>	
<i>Comprehension</i>	46
Table 10 <i>Effect of Sample Lengths on Comprehension</i>	48
Table 11 <i>Effect of Listener L2 Ability on Comprehensibility and Comprehension</i> ...	49
Table J1 <i>Comprehension Question Score Averages by Speaker</i>	137
Table J2 <i>Comprehensibility Ratings: Descriptive Statistics by Listener</i>	138

List of figures

Figure 1. Average comprehension versus comprehensibility per speaker50

List of acronyms

The following acronyms are used throughout this study.

ANOVA	Analysis of Variance	L1	First language
ASEAN	Association of South East Asian Nations	L2	Second or foreign language
CEFR	Common European Framework of Reference	M/C	Multiple Choice
EFL	English as a Foreign Language	MWU	Mann Whitney U
ELF	English as a Lingua Franca	NS	Native Speaker (of English)
ELP	English Language Proficiency	NNS	Non-Native Speaker (of English)
EMI	English as the Medium of Instruction	RQ	Research Question
ESL	English as a Second Language	T/F	True/False
HE	Higher Education	TLU	Target Language Usage
IELTS™	International English Language Test System	TOEFL (iBT)®	Test of English as a Foreign Language (internet-Based Test)
KS	Kolmogorov-Smirnoff	TOEIC®	Test of English for International Communication
KWH	Kruskal Wallis H	WM	Working Memory

1. Introduction

This chapter briefly explores the study's background, clarifies my motivation and rationale for conducting this research, and outlines the dissertation structure.

1.1 Background

The rise of English as a global lingua franca (ELF) has had profound impact on how the world conducts business, travels, and learns. Beyond the fact that English has come to be regarded as the language of academia (Doiz, Lasagabaster, & Sierra, 2012), its use as the medium of instruction in places where English is not the majority language (a practice hereafter referred to as English as the medium of instruction, or EMI) is being promoted as the “passport to the global world” (Deardan, 2014, p. 16).

Over the course of EMI's global expansion, several concerns have been raised about its implementation. EMI lecturers often lack the required English language (L2) competence and relevant pedagogical skills to teach in the L2, and trainings, when available, can be ineffective in addressing these deficiencies (Deardan, 2014). The issue of L2 competency is further compounded by the frequent absence of clearly defined standards of required L2 ability (Fenton-smith, Stillwell, & Dupuy, 2017). As Deardan laments in her global EMI survey, “in most countries there is currently no standardised English benchmark test for subject teachers teaching through EMI” (p. 27).

Conflicting perspectives on the desired end goal have exacerbated the confusion over EMI educators' target language competence. A prevalent view regarding native-speaker (NS) proficiency as the normative model for academic contexts tends to emphasize linguistics aspects such as pronunciation in developing EMI lecturer L2 oral proficiency (Fenton-smith et al., 2017). A contrasting perspective subscribing to the concept of localized English varieties, or ELF, argues for mutual understanding between interlocutors rather than achievement of NS ability as the target standard (Jenkins, 2006). Adoption of the latter position, however, only raises further questions regarding a suitable measure of English language proficiency (ELP) (given that available standardized

proficiency tests such as IELTS¹ and TOEFL² are based on the NS model) and appropriate linguistic features to target in prospective EMI educator L2 trainings. Additionally, the majority of ELF-domain comprehension studies to date have either focused on low stakes listening contexts (e.g., conversational interaction – Deterding & Kirkpatrick, 2006; Kaur, 2010; Smith & Nelson, 2006) and/or subjectively rated aspects of understanding such as perceived ease of understanding (e.g., Hellekjaer, 2010). Taking all of these issues into consideration, it is unsurprising that EMI program planners and policy makers struggle to address requirements and needs regarding lecturer ELP and training.

1.2 Motivation for the study

In Indonesia, the backdrop for my teaching context, increasing numbers of state and private universities are offering programs with English as the exclusive medium of instruction (known as ‘international programs’), and English/Indonesian mixed course offerings (referred to as ‘bilingual classes’) (Gill & Kirkpatrick, 2013; Simbolon, 2018). These are spurred on in part by national policy encouraging EMI programs at the tertiary level (Dewi, 2017).

Recently, several faculties at the state university where I work have sought to establish study tracks where EMI is used for teaching native speaking (L1) Indonesian students. My role has been to provide input on program specifications and to develop and undertake L2 and EMI pedagogy training for prospective and in-service non-native-speaking (NNS) EMI lecturers. Exploration of the aforementioned issues of sufficient instructor L2 ability for student understanding, appropriate ELP measure(s) for ELF contexts, and relevant foci for EMI lecturer L2 and pedagogical training would help to inform local EMI program planning and implementation.

¹ Internal English Language Testing System™ (Cambridge English Language Assessment, 1989)

² Test of English as a Foreign Language® (Educational Testing Service, 2005) – Note that the paper-based test (PBT) has been replaced by the internet-based test (iBT).

1.3 Aims and rationale of the study

This study aimed to:

- Explore factors influencing NNS's perceived and actual understanding of L1 Indonesian lecturers' L2 academic speech
- Assess a common standardized test's (i.e., IELTS) ability to predict association between speaker L2 ability and listener understanding
- Gain insights for EMI program policy and educator L2 training development

The investigation focused exclusively on L1 Indonesian lecturers as speakers given that most small- to mid-sized institutions in this context would rely on these rather than foreign instructors to implement EMI programs. Use of L1 Indonesian students as listeners is an accurate reflection of how 'bilingual courses', such as those offered by universities of this calibre, tend to primarily attract local rather than international students, a reality for many Asian HE institutions (Murata & Iino, 2018).

To date, relatively few studies have focused on understandability of NNS L2 academic speech, and those that have assume listener comprehension based on subjective measures. As one EMI researcher acknowledged, "a firmer conclusion [of listener understanding] would require a validation study where self-assessment item scores are correlated against a relevant listening test" (Hellekjaer, 2010, p. 238). This study's comparison of subjectively and objectively measured understanding could test the strength of earlier research's assumption of a correlative relationship.

Likewise, a dearth of research into whether standardized measures of speaker oral L2 proficiency predict NNS listener understanding, especially in high stakes listening contexts such as the EMI classroom, highlights how this investigation could contribute to the wider body of knowledge regarding EMI program policy formation. For this study, the IELTS test was selected as the speaker L2 ability measure due to its global and local prevalence.

A quantitative design was chosen to provide statistically significant rather than simply anecdotal evidence that could help inform EMI policy at both the local and international institutional level, specifically in the areas of EMI lecturers' ELP standards and training.

1.4 Study outline

This report begins with an introductory overview of the study's research topic and aims, as well as the motivation and justification for conducting this investigation. Chapter 2 examines the EMI- and ELF-related issues giving rise to this study, after which it defines constructs of perceived and actual understanding used in representing listener comprehension. It reviews research findings on factors that influence L2 speech understanding, paying particular attention to the methodology used, before concluding with a list of this study's research questions. Chapter 3 presents the study design and methodology, detailing employed samples and instruments. The results discussion of Chapter 4 describes collected data's characteristics before providing justification for statistical measures used in subsequent analyses. Analysis results are presented according to research question. Chapter 5 discusses results in the context of the wider body of research, comparing and contrasting with relevant literature findings. The paper concludes with Chapter 6, which summarizes this study's key findings, explores implications for development of EMI policy and educator L2 training, and notes limitations and potential areas for further investigation.

2. Literature review

This chapter details key issues contributing to the study's focus before reviewing relevant literature addressing speaker, stimulus, and listener factors influencing the understanding of L2 speech. It concludes with the study's research questions.

2.1 EMI challenges in Indonesia

Closer inspection of EMI's development within this study's focus context of Indonesia reveals a microcosm of some of the key issues that have been hampering its implementation worldwide. In 2013 the Indonesian government aborted a policy aiming to integrate EMI at the secondary educational level across the nation through the development of select institutions referred to as 'international standard' schools (Aritonang, 2013). The key reasons reported for this reversal included schools' lack of preparedness to implement EMI and teachers' lack of L2 ability (Dewi, 2017). At the tertiary level, educators' beliefs about EMI practice at one Indonesian university showed a similar lecturer ELP inadequacy as well as lack of understanding of relevant pedagogical practice (Simbolon, 2016). Additionally, local policy guiding EMI implementation failed to clarify minimum lecturer or student proficiency for program involvement. In the wake of this program's discontinuation after just a few years, Simbolon (2018, p. 123) noted, "the lack of clarity in the arrangement of the implementation of EMI seems to be common in several contexts, especially in [the] Asian region."

This lack of clarity is also reflected at the national level. Although the Indonesian Ministry of Education does not currently have an explicit written policy regarding inclusion of EMI programs, there is a stipulation encouraging higher education (HE) institutions to adopt at least one foreign language for content instruction, which tends to be English given its global ascendancy (Simbolon, 2018). However, in 2015 Indonesia's then-Minister of Education referred to EMI programs as a 'dual-language' vehicle for accommodating the Association of Southeast Asian Nation's (ASEAN) aim of skilled labour's free-flow amongst the member states (Nasir, 2015). It is unclear whether this key policy maker was

endorsing an ELF research-based proposal of combining the use of L1 along with L2 in HE EMI classrooms (e.g., Kirkpatrick, 2014; Murata & Iino, 2018) or suggesting that some program courses could be taught in L2 and others in the national L1. Although it is not this study's focus to explore a bilingual approach for EMI as espoused by ELF researchers, the aforementioned anecdotal evidence highlights the conflicting views regarding EMI implementation that confound the development of clear requirements for lecturers to participate in these programs.

2.2 Proposed ELF-domain solutions

Kirkpatrick's (2014) propounding of a multilingual versus L2-exclusive medium for instruction derives from the perspective that non-native 'Englishes' are "different rather than deficient" (Jenkins, Cogo, & Dewey, 2011, p. 284), or in other words, are valid in their own right without the need to conform to the norms of established English. Under the ELF construct, the NS is no longer the standard for successful communication, with the aim instead being mutual intelligibility between NNSs (Jenkins, 2006). Consequently, Jenkins along with other ELF researchers have proposed that a linguistic core comprised of phonological and lexicogrammatical forms intelligible across NNS English varieties be taught and practiced by interlocutors in contexts where such communication occurs. While proposing mutual understanding as an alternative standard to aim for in an EMI classroom context, ELF research has yet to offer concrete solutions to the issues of appropriate measures and standards of EMI lecturer ELP (cf., Harding & McNamara, 2018, for a recent survey of ELF-based proficiency test development).

2.3 Measures of L2 speech understanding

Measuring understanding of L2 speech first requires an awareness of how this concept is defined and operationalized by researchers. Smith and Nelson (1985) put forward a tripartite model of understanding, consisting of *intelligibility* (word/utterance recognition), *comprehensibility* (word/utterance meaning), and *interpretability* (word/utterance interpretation). Due to the inherent difficulty in measuring interpretability (Pickering, 2006) and a corresponding dearth of

literature addressing this aspect of understanding, discussion will be confined to the first two levels of understanding.

Armed with the presupposition that pronunciation errors are the highest frequency contributor to international communication breakdown (Jenkins, 2000, 2002, 2007), a majority of ELF studies have investigated understanding using intelligibility, often operationalizing the construct through transcription (Berns, 2008). The reasoning is that without a basic decoding of utterances as represented by the intelligibility construct, deeper levels of understanding are unattainable (Sewall, 2010). However, Smith and Nelson's (2006) subsequent research seeking to validate their earlier defined model revealed that although intelligibility was the most easily achieved aspect of understanding, the phonologically-oriented construct did not necessarily associate with deeper comprehension of L2 speech, the latter of which was operationalized through multiple choice (M/C) questions. These findings suggest that although utterance recognition may be a suitable comprehension measure for low stakes conversational interaction, other aspects of understanding are required for higher stakes listening contexts.

Operating in the English as a second language (ESL) research context, Derwing and Munro (1997; see also Munro & Derwing, 1999) proposed using intelligibility (extent of listener's understanding of an utterance, operationalized with transcription) and *perceived comprehensibility* (listener's perceived effort in understanding an utterance, operationalized with Likert scale ratings) as measures of listening comprehension. Their study amongst four L1 backgrounds of ESL students in Canada found intelligibility and comprehensibility to be correlated, but later empirical investigations suggested that the two constructs were measuring different things (e.g., Edwards, Zampini, & Cunningham, 2018; Kennedy & Trofimovich, 2008; Munro, Derwing, & Morton, 2006).

Given this study's focus on the high stakes comprehension context of the EMI classroom where students need to be able to grasp the meaning of lecturers' largely monologic speech, Smith and Nelson's (1985) comprehensibility

(henceforth referred to in this study as *comprehension*) and Derwing and Munro's (1997) perceived comprehensibility (henceforth referred to as *comprehensibility*, or perceived ease of understanding) were respectively selected as suitable measures of listeners' actual and perceived understanding of L2 academic speech.

Few investigations of listeners' understanding of L2 speech have examined the relationship between comprehensibility and comprehension as defined in this study, and those that have provide conflicting results. Van der Walt (2000) found that 140 NS/NNS students from Canadian, German, Dutch, and American universities were generally more cautious in rating their perceived understanding (using 4-point Likert scale) of various South African Englishes than reflected their actual comprehension (measured using an overarching short answer question). Smith and Nelson's (2006) study of various levels of understanding of NS/NNS mixed pairs' ($N=5$) conversational interaction by three groups of NNS ($n=10$, L1 Japanese), NS ($n=10$), and mixed NNS/NS ($n=9$, 9 L1 Asian, 1 NS) also showed that listeners' perceptions of ease of understanding (based on listeners' subjective judgment of having understood 60% of conversation) did not associate with their comprehension (operationalized as correctly answering two out of three M/C questions) of the 10 minute recordings. Contrastingly, a more recent investigation (Orikasa, 2016) of 37 L1 Japanese students' comprehensibility judgments (using 5-point Likert scale) and comprehension (labelled 'intelligibility', but operationalized with five short answer questions) of various L2 varieties ($N=8$, male/female pairs from NS, L1 Mandarin, Vietnamese) revealed varying-sized significant correlations ($.37 \leq r \leq .64$, $p < .05$) between perceived and actual understanding for five speakers. One explanation for the discrepancy in results could be the differences in implementation of comprehensibility/comprehension measures, while another is the varied targeted speech types, ranging from authentic radio and TV recordings (Van der Walt, 2000) to interview monologues (Orikasa, 2016) and conversational interactions (Smith & Nelson, 2006).

The latter issue reflects the need for elicited speech to reflect the target language usage (TLU) domain so as to allow for wider generalization of findings regarding listeners' perceived and actual understanding (cf., Appendix A for relevant theoretical and practical considerations of speech elicitation). The conflicting findings also point to the complexity of the relationship between comprehensibility and comprehension, justifying a more detailed literature review of factors influencing these two variables.

2.4 Comprehensibility and comprehension research findings

Following is a brief survey of research investigating speaker, stimulus, and listener factors' potential influence on perceived and actual understanding of L2 speech, beginning with studies using NS listeners to assess L2 speakers and followed by studies using NNS or mixed NS/NNS listeners.

2.4.1 L2 speech assessed by NS listeners

ESL domain research findings were explored to better understand speaker and stimulus factors influencing judgments of comprehensibility (typically measured using a 9-point Likert scale unless specified otherwise) and comprehension. Given this study's focus on NNS listeners, listener factors were not discussed in this section.

2.4.1.1 Speaker factors

A majority of studies addressing speaker ELP's association with comprehensibility measured its impact indirectly. Comparing results with a prior investigation (Munro & Derwing, 1995) of accentedness', comprehensibility's, and intelligibility's interrelatedness at a higher speaker oral proficiency (TOEFL score ≥ 550), Derwing and Munro (1997) concluded from their study of 26 NS undergraduates' ratings of 48 ESL learners' speech (L1 Cantonese, Japanese, Polish, and Spanish; mid- to high-intermediate ELP measured using university placement test results) that interlocutor ELP did not affect the semi-independent relationship between the three target measures, although speech phonological and grammatical error frequencies increased noticeably. In contrast, a study of speech linguistic feature correlates of comprehensibility ratings for 120 L1 Japanese speakers at three ELP levels

(grouped by average comprehensibility rating) concluded that higher proficiency interlocutors produce more comprehensible speech (Saito, Trofimovich, & Isaacs, 2016).

Likewise, an American study (Kraut & Wulff, 2013) of 6 NS and 18 L1 Asian, Hispanic, and Middle Eastern students' comprehensibility, accentedness and communicative ability found that the 78 NS listeners generally rated higher proficiency (ELP based on enrolment in beginner, intermediate, or advanced ESL course) speakers as more comprehensible [ANOVA: $F(2,41860)=226.936$; $p<.001$]. Taken together, these findings suggest that speakers with higher L2 ability are generally judged as more comprehensible, although other factors may also contribute to listeners' perceived understanding.

Although research amongst NSs has discovered differences between male and female phonological speech features (e.g., Byrd, 1994 – females' higher frequency of glottal stops; Whiteside, 1996 – males more likely to reduce vowels and consonant clusters), most ESL-domain studies have been concerned with gender effect on accentedness (e.g., Flege, Munro, & MacKay, 1995) rather than comprehensibility and/or comprehension, with the exception of Kraut and Wulff's (2013) aforementioned investigation. In their study, males ($n=9$) were rated as more comprehensible than females ($n=9$) at both low and intermediate speaker L2 abilities, regardless of their L1 background. This gender effect did not extend to high ELP speakers, suggesting a threshold proficiency where male/female pronunciation differences become less pronounced. However, considering the relatively small sample sizes and a lack of corroborative evidence from other investigations, Kraut and Wulff's findings should be treated with caution.

Regarding the factor of immersion experience in L2-speaking contexts, a Canadian longitudinal investigation of seven (L1 Mandarin, Nepali, Spanish) students' L2 speech development revealed that comprehensibility (rated by 19 NSs) increased for all but one speaker over the three-year time span (Kennedy, Foote, & Kurtz, 2012). Contrastingly, Derwing, Munro, and Morton's (2008)

longitudinal study of 32 (L1 Slavic, Mandarin) L2 learners' fluency and comprehensibility development over two years found that the Mandarin-background speakers' ratings (using 7-point Likert scale) remained relatively unchanged while the Slavic-background group experienced small but significant improvement. The authors attributed this difference to the varying quality of L2 exposure experienced.

2.4.1.2 Stimulus factors

Despite research demonstrating that task type and characteristics influence elicited L2 production features (e.g., Robinson, 2001 – task complexity related to lexicogrammatical accuracy and complexity; Krashen, 1977 – task type key factor influencing interlocutor lexicogrammatical usage), a significant number of ESL-domain comprehensibility studies have used elicitation tasks unreflective of wider L2 usage (e.g., picture description/narration task: Derwing & Munro, 1997; Munro & Derwing, 1999; Munro et al., 2006; Isaacs & Trofimovich, 2012; Saito, Trofimovich, & Isaacs, 2016, 2017). Crowther, Isaacs, Trofimovich, and Saito (2015) further highlighted this issue in their study of 10 NS applied linguists' comprehensibility and linguistic feature (using 1000-point slider scales) ratings of 60 L1 Chinese, Romance, Hindi, and Farsi university students' oral production from IELTS long turn (Part 2) and TOEFL iBT integrated speaking tasks. Results showed that IELTS- and TOEFL-prompted speech comprehensibility correlated with a combination of overlapping and distinct measures, thus underscoring the need to select task(s) eliciting speech reflective of the TLU domain so as to strengthen findings' generalizability.

Although many studies have explored associations between stimulus linguistic properties and listener understanding, only a handful will be reviewed here due to space limitations. The first stage of a dual-experiment study found that speech rate contributed to a small but statistically significant variance (second-order regression analysis: [$R^2=7\%$, $F(2,189)=8.46$, $p<.01$]) in 48 NS undergraduates' comprehensibility ratings for 48 (12 L1 backgrounds) intermediate ESL students (Munro & Derwing, 2001). Although the second stage, involving rate manipulation of 10 L1 Mandarin students' speech, confirmed speech rate's

moderate effect [$R^2=7\%$, $F(2,117)=7.92$, $p<.01$] on comprehensibility (assessed by 27 NS listeners), phonological errors' impact on perceived understanding was shown to be more significant [$R^2=41\%$, $F(3,116)=28.13$, $p<.01$]. Other suprasegmental features, such as stress and intonation, also impacted comprehensibility judgments in Kang, Rubin, and Pickering's (2010) investigation of 26 NNSs' TOEFL iBT speech performance. Using a multiple regression analysis, the researchers discovered that rate, stress and pitch characteristics accounted for half the variance [$R^2=.50$, $F(14,3609)=252.97$, $p<.001$] in 188 NSs' ratings.

Several studies discovered that varying linguistic features distinguished between L2 speakers at different comprehensibility levels. Isaacs and Trofimovich (2012) had 40 L1 French speakers' picture narrations rated by 60 NS non-linguistics students for comprehensibility, and then coded for measures of phonology, fluency, lexicogrammar, and discourse. Additionally, three ESL teachers provided retrospective reports on their comprehensibility ratings of the speech samples that were also coded for linguistic features. Limiting investigation to those linguistic measures that achieved a minimum correlation threshold ($r_p>.70$) and were reflected in teacher raters' coded feedback, the researchers ascertained that lexical richness and fluency distinguished between L2 speakers at lower comprehensibility rating levels, grammatical and discourse markers at higher levels, and word stress errors at all levels.

Building on these results and a precursor study narrowing down linguistic measures to eight categories (Saito et al., 2017), Saito et al.'s (2016) investigation of linguistic correlates of 120 L1 Japanese speakers' comprehensibility confirmed that features varied with rating levels. Five NS novice raters judged samples for comprehensibility, while five NS linguistics-background raters provided linguistic characteristics assessments. Interestingly, comprehensibility ratings associated with all eight linguistic measures: word stress and intonation factoring in all three rating levels; fluency and lexical appropriateness relevant at lower levels; and segmental and grammatical accuracy important at both beginner and advanced levels. However, in light of aforementioned findings

regarding task effect on comprehensibility ratings, results from the last two studies should be treated cautiously given their uncritical reliance on the sole task type of picture narration/description.

Although the above examples represent only a small sample of research addressing linguistic correlates of listener understanding, they clearly point to the potential for a wide variety of linguistic features to impact comprehensibility ratings, albeit mitigated by both speaker and listener factors.

2.4.2 L2 speech assessed by NNS

Research from the ELF domain was surveyed to balance and clarify ESL-domain findings on speaker, stimulus, and listener factors influencing perceived and actual understanding.

2.4.2.1 Speaker factors

Although studies have linked speaker ELP with NNS listeners' assessments of intelligibility (e.g., Bent & Bradlow, 2003; van Wuijngaarden, 2001; van Wuijngaarden, Steeneken, & Houtgast, 2002), none were found that examined speaker L2 ability in relation to NNS judgments of comprehensibility and/or comprehension of monologic L2 speech. Investigating achieved interactional understanding between NNS international students and their peers / instructors at a Malaysian university, Kaur (2010) concluded that an apparent lack of ELP did not appear to hinder participants' ability to make themselves understood. However, this claim was unsubstantiated, as only the minimum L2 ability level (IELTS band 5 / TOEFL score 550) of the 22 largely L1 Asian participants was provided. Despite Kaur's conflicting finding, which could be attributed to a differing construct of understanding and a lack of analytical proof, the majority of research reviewed earlier suggests it is intuitive to expect NNS listeners to perceive higher ELP speakers as more comprehensible.

Only one study was found that illuminated EMI experience's potential impact on listener understanding. It related lecturers' EMI experience with measures tangential to listener perceived understanding. Denver, Jensen, Mees, and Werther's (2013) exploration of business school students' (60% L1 Danish)

attitudes towards 31 NNS EMI lecturers' in-class L2 performance revealed that speakers' teaching experience did not associate with listener perceptions of either lecturing competence or lecturer L2 ability (both measured using 5-point Likert scales). However, neither perceived nor actual lecture understanding was measured directly.

None of the ELF-domain studies reviewed investigated the effects of speaker gender or L2 immersion experience.

2.4.2.2 Stimulus properties

Studies employing NNS listeners confirmed that speech task type could have significant impact on comprehensibility and comprehension. Jensen and Thøgersen's (2017) two-stage investigation of listener cognitive load (operationalized as response time to comprehension questions) as a function of task complexity discovered that L2 speech comprehensibility and comprehension decreased as elicitation task difficulty increased. In the first experiment, 20 L1 Danish raters demonstrated high comprehension ($M=95.2\%$) of True/False (T/F) statements recorded by 10 experienced EMI lecturers (80% NNS), with comprehensibility ratings correlating well with listener understanding ($r=.69, p<.05$). For the second experiment, two speakers (one NS, one L1 Japanese) were selected based on similar comprehension scores but varying listener response times, and recorded reading several academic texts aloud. Forty-two listeners then answered six text-based M/C questions and rated samples for comprehensibility. The NNS speaker scored much lower in comprehension and comprehensibility than the NS counterpart, suggesting that conclusions regarding listener understanding demonstrated in simple interactions do not necessarily apply to more complex interactions. Despite concerns of construct irrelevance (due to use of differing listener response tasks in experiments) and lack of generalizability (due to small sample sizes), this finding reinforces the need to select tasks appropriate to the TLU domain being investigated.

Although ELF research tends to focus more on pronunciation features' impact on listener understanding (Jenkins, 2000, 2002), a number of studies have confirmed that a variety of speech linguistic characteristics factor into perceived and actual comprehension. An investigation of Norwegian academics' comprehensibility, as rated (using a 4-point Likert scale) by the 364 mixed NNS/NS students who attended their EMI lectures, revealed that pronunciation errors and unfamiliar vocabulary were the top linguistic features influencing perceived understanding (Hellekjaer, 2010). However, the identification of these key factors was based on feedback from raters lacking linguistics training, thus raising doubt regarding the validity of these findings. A more recent study of Polish-accented L2 speech ($N=11$) as rated by 12 NNS (L1 Polish, Spanish) and 6 NS listeners revealed that segmental errors correlated moderately with NS ($r=.357, p<.050$) and Polish listeners' ($r=.350, p=.057$) comprehensibility judgments (Jułkowska & Cebrian, 2015). One might argue, though, that this latter finding lacks statistical power due to the small sample sizes.

Révész and Brunfaut (2013) investigated how increasing complexity of stimulus phonological, lexicogrammatical, and discourse properties affected 77 (77% L1 Mandarin) listeners' comprehension (operationalized with short answer response), finding that incidence of academic words (regression analysis: $R^2=.19, p=.04, f^2=.32$) and lexical diversity ($R^2=.17, p=.05, f^2=.28$) contributed the most to listening difficulty. The use of a NS to record speech samples and listeners' relatively high L2 ability (IELTS $M_{band}=6.5$) undoubtedly also influenced which stimulus properties featured more prominently in listener comprehension (cf., Sato et al., 2016). Taken together with the ESL-domain findings, it can be concluded that almost any linguistic feature can influence perceived understanding depending on other mitigating factors, while comprehension is more affected by lexical characteristics.

The impact of sample length on comprehension is related to a listener's working memory (WM), a key system that facilitates the storage and processing of information (de Jong, 1996). Prior research (e.g., Carpenter, Miyake, & Just, 1994; Just & Carpenter, 1992) has linked listening comprehension breakdowns

to the overtaxing of WM, suggesting that longer stimuli requiring listeners to store more aural information before responding could have an adverse affect on understanding. Thus, despite the majority of ELF/ESL studies on understanding of L2 speech failing to account for this factor, it is recommended to investigate sample length's potential affect on listener comprehension.

2.4.2.3 Listener factors

In her review of L2 listening comprehension research, Rubin (1994) posited that listener L2 proficiency is a major variable that should be accounted for in every comprehension study. Surprisingly though, findings on the interaction between listeners' ELP and their understanding are mixed, perhaps due to variation in how proficiency was measured. Smith and Nelson (2006) showed that although listener L2 ability affected each component of their tripartite model of understanding, it most impacted comprehension (labelled 'comprehensibility' in their study). Listener groups' (10 NNS; 10 NS; 9 NNS/NS) mean comprehension of five mixed NS/NNS pairs' conversational interactions (measured using M/C questions) increased with group's average ELP (NS group: $M=82$; NNS/NS group: $M=70$, TOEFL >600 ; NNS group: $M=62$, $300<TOEFL<600$), although no correlational analysis results were provided.

A Spanish-context investigation of factors affecting listening task processing times and comprehensibility judgments also points to listener ELP's impact on understanding (Ludwig & Mora, 2017). Six speakers (NS, L1 Catalan, German) recorded words and statements that were then used to conduct timed word association and T/F judgments by 50 listeners (20 L1 Catalan, 20 L1 German, 10 NS) with varying self-rated L2 abilities. The T/F samples were played again for listeners to judge comprehensibility using a 7-point Likert scale. Comparative analysis between L1 background, L2 proficiency, and comprehensibility ratings revealed that low-ELP listeners assessed shared-L1 speakers as easier to understand (t-test: $t(19)=-11.48$, $p<.001$, $r=.93$), while high-ELP listeners generally rated speakers as more comprehensible than their low-proficiency counterparts (ANOVA: $F=18.60$, $p<.001$, $\eta^2=.352$).

In contrast to the above findings, Orikasa' (2016) Japan-based study discovered inconsistent association between listeners' L2 ability and their comprehension. Thirty-seven students' TOEIC³ scores moderately correlated with comprehension results for half of the eight mixed NNS/NS interlocutors ($.37 < r < .47, p < .01-.05$), and not at all with comprehensibility ratings. Likewise, Jułkowska and Cebrian (2015) found no association between listeners' scores on an Oxford placement test and their ease of understanding of Polish-accented speech. However, the aforementioned ELP measure may be poorly suited for providing insight on L2 listening ability as it only addresses grammatical competence. In their exploration of NNS listener factors' impact on the comprehensibility ratings for 50 speech samples, Saito et al. (2019) also discovered a lack of association between the two variables. Similarly to the aforementioned Polish study, though, the 110 (8 L1 groups) listeners' ELP was measured by assessing productive (using a conflation of NS- and self-ratings of perceived comprehensibility) rather than receptive skills. These conflicting findings underscore the importance of using appropriate measures of L2 ability.

Other NNS listener factors, such as L1 familiarity, shared L1 background, age, and gender were not addressed in this review as they were deemed beyond the scope of this study.

2.5 Research Questions

This study aims to address some of the limitations in current research regarding understanding of L2 speech. First, findings demonstrating speech task effects on listener comprehension compellingly argue for the use of more real-world elicitation prompts that align with the TLU domain. Secondly, in higher stakes understanding contexts such as the EMI classroom, evidence of deeper listener comprehension of interlocutor communication is recommended to validate measures of perceived understanding. Thirdly, conflicting reports of the impact of speakers' and/or listeners' L2 proficiency on understanding arise in part due to the wide variety of L2 ability measures employed and occasional mismatches

³ Test of English for Internal Communication® (Educational Testing Service, 2019)

with the targeted language skills, underscoring the need for additional investigation into the relationship between commonly used proficiency test scores and levels of listener understanding of L2 speech. The current study thus addresses the following research questions (RQ) in the context of L1 Indonesian academic speakers and L1 Indonesian university student listeners:

- RQ1. Do listeners' comprehensibility judgments relate to their assessed comprehension of L2 speech?
- RQ2. Do speaker factors (i.e., L2 ability, EMI experience, L2 immersion experience, gender) relate to comprehensibility and comprehension?
- RQ3. Do stimulus properties (i.e., linguistic features, elicitation task, sample length) relate to comprehensibility and comprehension?
- RQ4. Does listener L2 ability relate to comprehensibility and comprehension?

3. Methodology

3.1 Study design

This study employed quantitative methods to examine factors influencing students' comprehensibility and comprehension of L2 academic speech from lecturers sharing the same L1. To accomplish this, appropriate elicitation methods were researched to obtain L2 speech samples representative of the TLU domain, the EMI classroom (see Appendix A for discussion of theoretical and practical considerations of L2 speech elicitation). The resulting speech tasks were used to elicit and record samples of academic L2 speech from L1 Indonesian lecturers. Online surveys facilitated collection of comprehensibility ratings and comprehension question responses from L1 Indonesian listeners, as well as ratings of the linguistic features of the L2 speech samples from English as a foreign language (EFL) teachers. Data was analysed using SPSS Version 25.

3.2 Participants

There are three samples from which data was gathered for this study: L1 Indonesian lecturers (*speakers*), L1 Indonesian student raters (*listeners*), and experienced language instructors with linguistics training (*EFL teachers*).

3.2.1 Ethical considerations

All participants were over 18 years of age, and confirmed that they had read and understood the information sheet. They were informed that:

- Data collected would be securely stored on an encrypted computer drive and/or password-protected university servers.
- No personal or identifying information would be included in the speech recordings or research report.
- Their data would be removed from the study should they withdraw their consent within two weeks of providing it.

All participants provided their consent for the storage and usage of their data, with speakers doing so in written form, while listeners and EFL teachers did so electronically. See Appendix B, Appendix E, and Appendix H for respective participant information and consent forms used.

3.2.2 Sampling

Given that the target context of the study is the EMI classroom, a purposive sampling strategy was used to select the speakers and listeners (Mackey & Gass, 2005). Selection criteria for speakers included a recent (within the last two years) IELTS score (≥ 5.5 to ensure speech tasks' interactivens; see Bachman & Palmer, 1996) and a mix of EMI teaching experience. For listeners, the requirement was mid- to upper-intermediate ELP so as to minimize the effects of L2 ability in listening for comprehension. Both could also be considered samples of convenience as they came from the local Indonesian university where I work. Likewise, EFL teachers selected represented a purposive sample of convenience based on requirements of EFL teaching experience and formal linguistics training.

3.2.3 Speakers

All speakers were L1 Indonesian lecturers from five different faculties (Engineering and IT, Education, Psychology, Law, and Economics) of a religious state university on the island of Sumatra, Indonesia.

3.2.3.1 Pilot

In order to test the speech elicitation tasks, two lecturers (one female, one male) were asked to participate in the recording process. Both were from the Engineering faculty, having a mean age of 38.5 years and the same IELTS score of 5.5. The female speaker was the only one who had spent time living abroad in an L2 speaking context, but neither had EMI classroom experience. Due to the challenge of finding sufficient participants with recent IELTS scores, the two pilot speakers also participated in the main study.

3.2.3.2 Main study

An initial 22 speakers participated in the speech capture process, but two speakers' data was excluded from subsequent rating procedures due to lack of recent IELTS scores, leaving a total of 20 speakers. Although three other participants' IELTS results fell below the required minimum (5.5), their samples were still included on account of having EMI experience. IELTS scores ranged from 4.0 – 7.5 ($M_{IELTS}=5.7$), with five obtained from predictive rather than official

tests. Speakers' ages ranged from 30-47 years, with an average of 37.2 years. There were almost twice as many females ($n=13$) as males ($n=7$) due to the sample being one of convenience.

Five speakers were active EMI lecturers, with 1 to 2.5 years of experience in the same EMI program. One other participant had two years of EMI experience as a teaching assistant while studying abroad. The remaining 14 speakers had no EMI experience.

Regarding L2 immersion experience, one speaker had done further studies in Australia for a number of years, while four others frequently used English in communicating with fellow international students while completing graduate degrees in NNS L2 speaking regions (i.e., Malaysia, Netherlands). Of those with significant time spent abroad, three mentioned that these experiences occurred 2 to 19 years ago. Several other lecturers also mentioned having spent time, ranging from several weeks to months, living abroad in ESL contexts, but 14 speakers had never travelled internationally.

3.2.4 Listeners

All members of this sample were L1 Indonesian students from the English department of the Education faculty at the same institution as the speakers.

3.2.4.1 Pilot

From the 10 students who were invited to participate in the pilot survey measuring comprehensibility and comprehension of speakers' L2 speech samples, only three completed the online form. All were male, in their sixth semester of English teaching studies, aged 20-21 years ($M=20.7$), and self-rated as having a high intermediate L2 ability.

3.2.4.2 Main study

For the main study, 36 students (32 females, 4 males) aged 20-22 years ($M=20.8$) and in their sixth semester of English teaching studies completed the online survey. Self-rated L2 ability ranged from elementary to advanced, but the mean rating ($M=3.39$; see listener survey in Appendix F for scale description) was in

the target range between intermediate (3) and upper intermediate (4). Although one third of respondents indicated having taken the TOEFL test within the last 12 months, this amount was insufficient for use in further analysis. Those completing the survey were issued a certificate of appreciation for their participation in the research.

3.2.5 EFL teachers

To carry out linguistic analyses of speech samples, three NSs ($M_{\text{age}}=49.3$ years; one female, two males) with 12 to 15 years ($M=14$) of EFL teaching experience were recruited. All EFL teachers had graduate-level linguistics training (ranging from post-graduate diploma to Masters) and two had high familiarity of Indonesian-accented L2 speech.

3.3 Speech elicitation and recording

This section describes the design and implementation of speech tasks.

3.3.1 Speech elicitation task specifications

Based on TLU domain tasks (see Appendix A for discussion of TLU domain tasks identification), four speech prompts (one for practice and three for speech data collection) were created for use in an interview setting. The tasks included (see Appendix C for speech prompt text):

- Practice Prompt – giving a brief promotion for a class subject
- Prompt 1 – providing reasons as to a subject’s importance
- Prompt 2 – listing key course learning objectives/outcomes
- Prompt 3 – giving details for a class assignment

Speakers were instructed to view these tasks as a type of role-play, where they were directly addressing a class of beginner students with little background knowledge in the focus subject. Based on insights gained from piloting elicitation prompts, each task was designed to get speakers to incorporate four discrete points in their response (e.g., “List four key course learning objectives”) so as to better facilitate subsequent listener comprehension measurement. Speakers were encouraged to take preparation time (two to three minutes) and

notes before responding to a prompt, and they were instructed to speak for approximately two minutes to generate sufficient speech data.

3.3.2 Interview process

Interviews for collecting speech and participant data were conducted individually in a relatively quiet environment (campus conference rooms). A digital Zoom H2 audio recorder (44.1-kHz sampling rate with 16-bit quantization) was used to capture session audio. Basic information questions were asked first in order to diminish speaker anxiety before moving on to a practice speech task to minimize the 'language mode' effect (Saito et al., 2016). Feedback on successful task completion and clarification of instructions were also provided.

3.3.3 Post-recording processing

Using Audacity Version 2.3.1 (2018), the WAV format recordings produced during the interviews were edited to remove 'dead air time' due to extended pauses, to filter out potentially distracting background noises, and to normalize clip volume so as not to distract from rating processes. Clips were bookended with one second of silence and then encoded to 80-120kbps mono MP3 format to minimize network load in the online surveys' audio playback.

To create comprehensibility samples ($N=20$), 25-39 seconds ($M=32s$; compare with $M=25s$ for Saito et al., 2016; $M=7s$ for Derwing & Munro, 1997) were extracted from the beginning of each response to Prompt 1. Samples ($N=40$, based on responses to Prompts 2 & 3) for the two comprehension questions (Q1 and Q2, respectively) were somewhat longer as each had to contain the four discrete information points for listeners to grasp: $M_{Q1}=113s$ ($Min=59s$; $Max=159s$); $M_{Q2}=111s$ ($Min=63s$; $Max=208s$). The speech samples used for rating linguistic categories were also taken from Prompt 1 recordings, but with a longer average length ($M=61s$; $Min=58s$; $Max=67s$) in order to provide EFL teachers with sufficient content to rate linguistic features (Saito et al., 2016).

Speakers' responses to all three speech prompts were transcribed using conventions from the Vienna Oxford International Corpus of English Version 2.1

(VOICE Project, 2007). See Appendix D for sample transcriptions and corresponding conventions used.

3.4 Measuring comprehensibility and comprehension

To facilitate collection of comprehensibility, comprehension, and relevant personal data from listeners, an online survey was created using Version May 2019 of Qualtrics (<https://lancasteruni.eu.qualtrics.com>). See Appendix F for a copy of the survey.

3.4.1 Pilot

The initial trial of the online survey for listeners revealed a need for several changes to be implemented. First, a low response rate (30%) necessitated organizing groups of students to take the survey during class time so as to ensure a larger number of completed responses. Second, comprehensibility rating scale values were reversed so as to align with instruction wording. Finally, the comprehension question answer key was refined based on the pilot participants' answers.

3.4.2 Main study

Data collection took place over several days during class time, although several listeners completed the survey outside of class. Most participants completed the tasks on their mobile phones rather than laptops as instructed due to the unavailability of WiFi connectivity for the latter. The increased difficulty of typing in answers using the smaller-screened devices may have negatively impacted comprehension task results. Although the environment where the majority did the survey was somewhat noisy, participants' use of earphones may have mitigated this factor's effect on listening tasks. Another problem noted was listeners' failure to take prescribed break(s) during the lengthier comprehension measurement portion [Time(40 samples)=121m38s] of the survey which may have introduced rater fatigue impact on scores. Finally, several listeners identified at least four instances of interrupted or distorted comprehension sample playback due to network connectivity issues, an issue that was accounted for in the marking process (cf., Section 3.4.6).

I will now examine each portion of the survey as it was implemented in the main study.

3.4.3 Comprehensibility ratings

Comprehensibility, defined as listeners' perceived ease of understanding of speakers' L2 speech, was operationalized using a 9-point Likert scale, with textual descriptions for only the lowest (1=very easy to understand) and highest (9=very difficulty to understand) rankings (Munro & Derwing, 2001; Saito et al., 2016). Before rating the 20 speech samples, participants were provided with a practice sample to rate so to familiarize them with the procedure. Task instructions clarified that ratings were to be based on degree of ease of understanding rather than an assessment of speaker's oral L2 proficiency. Sample playback was limited to a single listening only so as to remove the affect that repeat listening might have on ratings (Trofimovich & Isaacs, 2012). The samples were also presented in randomized order so as to minimize the potential impact of rater fatigue (Saito et al., 2017).

3.4.4 Comprehension measurement

Comprehension was defined as listeners' grasp of key points of L2 speech samples and operationalized through two short answer questions per speaker. Training consisted of several practice comprehension questions based on samples from two speakers whose lack of recent IELTS scores disqualified their involvement in the main study. Listeners were given opportunity to compare suggested answers with their own input. Similar to the comprehensibility rating process, speaker sample and question sets were presented in randomized order in order to minimize the impact of rater fatigue. Participants listened once to a sample, noting down the four key points mentioned.

3.4.5 Post-tasks questionnaire

The collection of listeners' personal data required less cognitive focus than evaluating comprehensibility and comprehension, and as such was placed at the end of the survey. Age, sex, study major, and current semester of study were gathered first, after which listeners were asked to provide a self-rating of their L2 ability and most recent scores from any standardized L2 proficiency test. The

former task provided six proficiency levels to choose from (Novice to Master), with detailed descriptions for each in L2 that were based on the Common European Framework of Reference's (CEFR) self-assessment grid for speaking and listening (Council of Europe, 2001).

3.4.6 Post-survey processing

The largest post-survey task was to mark the comprehension questions in an accurate and consistent manner. To accomplish this, I (having high familiarity with L1-accented L2) and a Hong Kong native with pedagogical training (having only several months' exposure to L1-accented L2) undertook a two-step process to arrive at a final comprehension score for each speaker. First, listeners' answers for each comprehension question were marked separately according to the answer key (cf., Appendix G), with one mark for each correctly identified key point (four marks for each question, for a total possible score of eight per speaker), 0.5 marks for points partially identified, and 0 for those points not identified. Neither spelling nor grammatical errors were considered in the scoring process as long as the words or phrases were deemed close enough to the answer. For example, the following listener responses were assessed as shown against an item answer key of "Know how to create the method for solving system problems".

1. How to create the method to solve the problem.	1 mark	Demonstrated understanding despite omission of "system"
2. Will know how to create the methode for solve the problem of the system.	1 mark	Spelling and grammatical errors ignored in scoring process.
3. Solve the problem.	0.5 marks	Correctly identified "solve" and "problem", but missed key words "create" and "method".

- | | | |
|------------------------|---------|---|
| 4. Create the problem. | 0 marks | Although correctly identified several key words from sample, response contradicts intended meaning. |
|------------------------|---------|---|

Due to some listeners' difficulty with sample playback, blank response, rather than assumed to represent a lack of understanding, were not scored or included in calculation of overall comprehension averages. The second step was to go through each question answer together, comparing the marks given. For each one that differed, a discussion ensued as to the reasons why before arriving at an agreed upon final mark. This process was helpful not only for catching marking errors but also in eliminating subjective marker bias.

3.5 Measuring linguistic features

Similar to data collection from listeners, measurements of stimuli linguistic properties and EFL teachers' background information were gathered through an online survey using Version May 2019 of Qualtrics (<https://lancasteruni.eu.qualtrics.com>). Appendix I contains a copy of the survey.

The instrument used for rating speech linguistic categories was developed and tested by Saito et al. (2017), and consequently no pilot was undertaken for the online survey facilitating this rating process. A key difference between the creators' and this study's implementation of the instrument is the latter's use of elicited speech targeting TLU domain language rather than pre-selected vocabulary to highlight phonological differences. Another contrast is this study's exclusive use of NNS (L1 Indonesian) speech samples (as opposed to also including a NS sample) to challenge the notion of the NS standard. Two practice samples and corresponding audio- and transcript-based measures were provided to familiarize teachers with the rating procedure.

EFL teachers were provided with a summary statement of what the speaker was talking about as context for speech excerpts. Similarly with other instruments,

the 20 samples were presented in random order so as to minimize the impact of rater fatigue. Recordings could be replayed to assist in the process of rating multiple categories at once (Saito et al., 2016). The mean time to complete the survey was three hours.

3.5.1 Audio-based measures

Four phonological categories were rated for each speech recording: segmental errors (missing, misplaced, or added vowels and consonants), word stress errors (misplaced or added pronunciation emphasis), intonation (appropriate variation of speech tone versus monotone delivery), and speech rate (listening comfortableness due to speed of delivery). For each category, a 1000-point slider scale (*Min=0; Max=1000*) was used to provide a numerical rating. Each endpoint had a brief textual description along with a frowning / smiling emoticon to denote the leftmost (negative) and rightmost (positive) ends of the continuum respectively. The slider's default position was in the middle of the scale, and EFL teachers were informed that even slight positional adjustments could represent significant changes in the rating value. The advantage of using such an instrument rather than a Likert scale is that it allows for greater precision (Saito et al., 2016).

3.5.2 Transcript-based measures

To remove influences of pronunciation and fluency on lexicogrammatical ratings, Saito et al. (2016) recommend using transcripts for evaluation of the four categories of lexical appropriateness (accurate use of vocabulary), lexical richness (appropriate variation in word usage), grammatical accuracy (errors in word order and morphological endings), and grammatical complexity (sophistication of grammatical structure). Transcripts were tidied up through removal of filler words such as "um" and "uh", pronunciation-related spelling errors, and obvious mispronunciations based on contextual information. The same rating procedure was used for audio-based measures (i.e., a 1000-point slider, textual description, and corresponding emoticon for each endpoint).

3.5.3 Post-task questionnaire

Beyond providing background information such as age, sex, and relevant linguistics training and experience, EFL teachers also used 9-point Likert scales to assess the extent to which they understood the linguistic categories being rated (0=don't understand at all; 9=understand very well) and how comfortable they were in using each of the corresponding measures (0=very difficult to use; 9=very comfortable to use). Overall understanding of the linguistic measures seemed fairly robust ($M=7.5$), ranging from a mean of 6.7 (speech rate) to a mean of 8.3 (grammatical accuracy). Similarly, confidence in using the scales for each measure was not far off EFL teachers' understanding ($M=7.3$), although this time lexical appropriateness ($M=6.7$) and richness ($M=6.7$) were deemed most difficult to use while grammatical accuracy was perceived as the easiest ($M=8.0$).

4. Results

This chapter details the results from analysis of listeners' assessments of comprehensibility and comprehension, EFL teachers' ratings of speech samples' linguistic features, and relevant speaker, stimulus, and listener factors.

4.1 Rater consistency

The 36 listeners were overall consistent in their ratings of the 20 speakers' comprehensibility, as demonstrated by the high Cronbach's alpha of .96. Likewise, their comprehension scores, based on correct identification of eight discrete informational points divided across two short answer questions, also showed a high level of agreement ($\alpha=.92$). Only 18 cases were considered in calculation of the latter's reliability index as 50% of students missed answering one or more comprehension questions which were left unmarked, thereby excluding their cases from the overall analysis. If these blank questions were scored as 0, thus allowing all cases to be considered, comprehension's reliability index would increase slightly ($\alpha=.93$). Regardless, both comprehensibility ratings and comprehension scores comfortably achieved the desired benchmark of 0.70-0.80 indicating high rater consistency (Pallant, 2016), and were averaged to achieve a single mean value per understanding measure for each speaker in keeping with standard L2 understanding research practice (e.g., Derwing & Munro, 1997; Isaacs & Trofimovich, 2012; Saito et al., 2016). Unanswered questions were excluded from calculations of comprehension averages, resulting in sample sizes ranging from 32 to 35 ($M=34.1$) for mean comprehension scores (see Table J1 in Appendix J for averaged comprehension scores and corresponding sample sizes).

As seen in Table 1, inter-rater reliability for linguistic feature ratings was generally lower than that for comprehensibility and comprehension, perhaps due to underlying linguistic categories' relative complexity compared with the latter measures (Saito et al., 2016). Nonetheless, six out of the eight linguistic dimensions assessed by EFL teachers for each speech sample ($N=20$) demonstrated sufficient consistency, and consequently the ratings for these

measures were able to be averaged across the three teachers to derive a single score per linguistic measure for each speaker.

Table 1

Reliability Indices for EFL Teacher-Rated Linguistic Categories

Linguistic measures	α (Cronbach's alpha)
Segmental errors	.85
Stress errors	.91
Intonation	-.40*
Speech rate	-.01*
Lexical appropriateness	.88
Lexical richness	.84
Grammatical accuracy	.93
Grammatical complexity	.86

Note: * $\alpha < .70$

Surprisingly, reliability indices were unacceptably low for ratings of intonation ($\alpha = -.40$) and speech rate ($\alpha = -.01$), suggesting variance in teachers' perspectives on what constituted natural-sounding pitch changes and speed of delivery (see Section 5.1 for further discussion of these inconsistencies). Due to ratings' unreliability, these two linguistic measures were not included in subsequent comparative analyses.

4.2 Descriptive statistics

This section explores the descriptive statistics for listeners' comprehensibility ratings and comprehension scores, followed by a similar detailing of EFL teachers' linguistic feature assessments.

4.2.1 Comprehensibility and comprehension: Descriptive statistics

The summary statistics in Table 2 show that speakers were largely perceived as easy to understand while overall comprehension of their speech was relatively robust, with a mean score of 75%.

The negative skewness statistics for comprehensibility and comprehension revealed data distributions to be mildly skewed towards higher values, while comprehension's negative kurtosis pointed to a flatter distribution than that of comprehensibility. A Lilliefors-corrected Kolmogorov-Smirnov (KSD) test of normality was conducted and the non-significant statistical results ($p>0.05$) suggested normal distributions for both comprehensibility [$D(20)=0.118$, $p=0.200$] and comprehension [$D(20)=0.163$, $p=0.170$] (Pallant, 2016).

4.2.2 Comprehensibility ratings per listener

Although inter-rater consistency was already established, further investigation of subjective judgments of perceived ease of understanding on a per listener basis (see Table J2 in Appendix J) revealed several noteworthy features.

Nine listeners showed little to no variation ($.00 \leq V \leq .10$) in their ratings, always judging speakers to be very comprehensible with few exceptions. On the other end of the spectrum, seven listeners used the full/almost full extent of the scale ($Range=7-8$). The contrast with those showing little to no variance suggests that listeners may have had varying levels of understanding and/or confidence in their usage of the comprehensibility instrument.

4.2.3 Linguistic features: Descriptive statistics

From the summary in Table 3, mean ratings for the six⁴ linguistic categories appear to cluster around the midpoint of the 1000-point scale used in the measurement procedure, with stress errors having the highest values and grammatical accuracy the lowest. Looking at the minimum and maximum ratings, it is clear that EFL teachers only used 25-45% of the 1000-point scale range despite being instructed to make use of the full scale. Several reasons for this could include the relative lack of variance amongst speakers' oral ELP (as indicated by IELTS scores) and EFL teachers' unconscious comparison of L2 speech performance with a NS standard.

⁴ Intonation and speech rate not included in analysis due to failing inter-rater reliability test (cf., Section 4.1).

Table 2

Average Comprehensibility Ratings and Comprehension Scores: Descriptive Statistics

Averaged values	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>Mdn</i>	<i>M</i>	<i>SD</i>	Skewness		Kurtosis	
							Statistic	Std. Error	Statistic	Std. Error
Comprehensibility ^a	20	1.00	3.00	2.33	2.28	.39	-.807	.512	.106	.992
Comprehension ^b	20	4.71	7.19	6.25	6.02	.69	-.401	.512	-.821	.992

Note: ^aComprehensibility scale: 1=very easy to understand, 9=very difficult to understand. ^bComprehension score is out of possible 8 points.

Table 3

Linguistic Categories⁴: Descriptive Statistics

Linguistic measures	<i>N</i>	<i>Min</i>	<i>Max</i>	<i>Mdn</i>	<i>M</i>	<i>SD</i>	Skewness		Kurtosis	
							Statistic	Std. Error	Statistic	Std. Error
Segmental errors	20	310.00	655.67	500.67	494.28	98.00	-.220	.512	-.804	.992
Stress errors	20	413.67	665.33	553.17	542.98	65.15	-.592	.512	.073	.992
Lexical appropriateness	20	356.00	719.67	500.33	518.92	104.13	.362	.512	-.747	.992
Lexical richness	20	278.33	722.67	475.67	494.50	123.23	.059	.512	-.839	.992
Grammatical accuracy	20	212.67	545.33	398.33	407.42	79.14	-.421	.512	.490	.992
Grammatical complexity	20	268.33	638.33	427.17	441.85	99.62	.471	.512	-.264	.992

Each linguistic feature's skewness and kurtosis statistics were relatively mild ($<\pm 2$), suggesting normal data distributions. Further investigation by means of KSD tests returned non-significant results ($p > .05$) for each category (see Table 4), which added to evidence of linguistic features' data normality.

Table 4

KSD Normality Tests for Linguistic Measures⁵

Linguistic measures	Statistic	df	<i>p</i>
Segmental errors	.116	20	.200
Stress errors	.176	20	.107
Lexical appropriateness	.157	20	.200
Lexical richness	.110	20	.200
Grammatical accuracy	.142	20	.200
Grammatical complexity	.105	20	.200

4.3 Considerations for statistical analyses

Selection of appropriate tests for data analysis depends on data characteristics such as type (i.e., continuous, ordinal, categorical), distribution normality, and sample size (Green, 2013; Pallant, 2016). The majority of studies investigating L2 speech comprehensibility use parametric tests for analysis of ordinal data produced from listeners' judgments of perceived understanding (e.g., Hellekjaer, 2010; Isaacs & Trofimovich, 2012; Munro & Derwing, 2001; Saito et al., 2016, 2017) despite parametric tests' assumption of continuous data. Some might argue that collapsing comprehensibility ratings into a single average per speaker satisfies parametric tests' requirement, but other researchers still regard this as invalid handling of ordinal data (cf., Kero & Lee, 2016, for further discussion on this issue). This potential validity concern, combined with the failure of this study's speaker sample size (N=20) to satisfy parametric tests' assumption of

⁵ Intonation and speech rate not included in analysis due to failing inter-rater reliability test (cf., Section 4.1).

sufficient data points (widely used rule of thumb is $N > 30$) (Green, 2013), necessitated the use of non-parametric analyses.

4.4 RQ1: Do listeners' comprehensibility judgments relate to their assessed comprehension of L2 speech?

Comprehensibility was selected as a common measure of listeners' perceived understanding, while comprehension was used to provide more solid evidence of this understanding. Given both variables' relatively few cases ($N=20$), the non-parametric Spearman's rho test was employed to investigate any association between the two. However, no significant correlation was discovered between averaged comprehensibility and comprehension values ($r_s = -.12, p = .626$), showing that listeners' perceptions of how easy a speaker was to understand did not reflect their actual understanding of corresponding L2 speech.

Applying the same analysis to each listener's comprehensibility ratings and comprehension scores provided a similar result, revealing only two instances of significant correlation being discovered [$r_s(\text{SR25}) = -.60, p = .005$; $r_s(\text{SR33}) = .49, p = .43$]. The negative rho value for the first case indicates that listener SR25's perceptions of how easy a speaker was to understand reflected her/his actual understanding of their speech with moderate accuracy. The opposite occurred with listener SR33, who achieved higher comprehension scores for those speakers perceived as being more difficult to understand. Despite these exceptions, the general lack of association between comprehensibility and comprehension at both the overall and individual levels suggests that factors influencing listeners' perceived ease of understanding differ from those influencing their actual comprehension.

4.5 RQ2: Do speaker factors (i.e., L2 ability, EMI experience, L2 immersion experience, gender) relate to comprehensibility and comprehension?

In addressing the second research question, the impact of speakers' L2 proficiency, prior EMI teaching experience, time spent in L2-speaking contexts, and gender on listeners' perceived and actual understanding were considered.

4.5.1 L2 ability

One of the study's aims was to determine if standardized L2 proficiency scores were a reliable predictor of how well an EMI lecturer's L2 academic speech could be understood, or be perceived to be understood, by listeners. Speakers were divided into two groups based on their IELTS scores (Upper: 6.0-7.0; $n=10$; Lower: 4.0-5.5; $n=10$), after which a Mann Whitney U (MWU) test was conducted with the IELTS score as the grouping variable. The effect sizes of the comparisons were calculated using the following recommended formula: $r=Z/\sqrt{N}$ (Pallant, 2016). Reported statistics reflect that the MWU test compares group medians rather than distributions, and as such mean and standard deviation values were not included (see Table 5 for statistics for each group).

Table 5

Effect of Speaker IELTS Score on Comprehensibility and Comprehension

		Comprehensibility				
IELTS scores	<i>n</i>	<i>Mdn</i>	<i>U</i>	<i>Z</i>	<i>p</i>	<i>r</i>
Upper	10	2.15	24.00	-1.968	.049*	-.440
Lower	10	2.50				
		Comprehension				
IELTS scores	<i>n</i>	<i>Mdn</i>	<i>U</i>	<i>Z</i>	<i>p</i>	<i>r</i>
Upper	10	6.26	49.00	-.076	.940	-.017
Lower	10	6.15				

Note: * $p<.050$

There was no significant difference in comprehension marks between upper and lower groups of speakers' IELTS scores. Comprehensibility ratings, on the other hand, were slightly better for speakers with higher versus lower IELTS scores, with the MWU test reporting a moderate-sized (defined as $.30 \leq r < .50$ by Cohen, 1988) statistical difference. Thus, although speakers with higher IELTS scores were generally perceived to be slightly easier to understand, their L2 speech was not necessarily comprehended better than those speakers with lower IELTS scores.

4.5.2 EMI experience

Only one quarter of participating lecturers had any significant experience teaching in an EMI classroom, and as such the division of speakers based on a minimum of six months of L2 instruction experience resulted in disparate-sized groups as seen in Table 6. Surprisingly, MWU test results revealed that EMI teaching experience had no significant impact on listeners' perceived ease of understanding or actual comprehension. It should be noted, however, that of those who reported having L2 instruction experience, the average time teaching in an EMI classroom was only 1.8 years, suggesting that the discovered lack of association should be treated cautiously.

Table 6

Effect of EMI Teaching Experience on Comprehensibility and Comprehension

Comprehensibility						
EMI experience	<i>n</i>	<i>Mdn</i>	<i>U</i>	<i>Z</i>	<i>p</i>	<i>r</i>
Yes	5	2.44	30.00	-.655	.512	-.146
No	15	2.33				
Comprehension						
EMI experience	<i>n</i>	<i>Mdn</i>	<i>U</i>	<i>Z</i>	<i>p</i>	<i>r</i>
Yes	5	5.58	32.00	-.480	.631	-.107
No	15	6.32				

4.5.3 L2 immersion experience

Another speaker attribute to explore with regards to listeners' perceived and actual understanding was time spent abroad in L2-speaking regions. Speakers were divided into groups based on whether or not they had spent two or more years living in an area where L2 was widely used. Those with the required amount of experience ($n=5$) had spent an average of 3.6 years in L2-speaking contexts. However, MWU tests showed that this factor did not significantly affect comprehensibility ratings or comprehension scores (see Table 7).

Table 7

Effect of L2 Immersion Experience on Comprehensibility and Comprehension

L2 immersion experience	Comprehensibility					
	<i>n</i>	<i>Mdn</i>	<i>U</i>	<i>Z</i>	<i>p</i>	<i>r</i>
Yes	5	2.61	29.50	-.699	.484	-.156
No	15	2.25				

L2 immersion experience	Comprehension					
	<i>n</i>	<i>Mdn</i>	<i>U</i>	<i>Z</i>	<i>p</i>	<i>r</i>
Yes	5	6.32	34.00	-.306	.760	-.068
No	15	5.96				

4.5.4 Gender

Female speakers outnumbered male speakers almost two to one. MWU tests comparing their respective averaged comprehensibility ratings and comprehension scores showed that gender did not significantly impact listeners' subjective or objective measurements of understanding (as summarized in Table 8); in fact, ratings for female and male speakers' comprehensibility were almost identical.

Table 8

Effect of Speaker Gender on Comprehensibility and Comprehension

Speaker gender	Comprehensibility					
	<i>n</i>	<i>Mdn</i>	<i>U</i>	<i>Z</i>	<i>p</i>	<i>r</i>
Female	13	2.33	40.50	-.397	.692	-.089
Male	7	2.33				

Speaker gender	Comprehension					
	<i>n</i>	<i>Mdn</i>	<i>U</i>	<i>Z</i>	<i>p</i>	<i>r</i>
Female	13	5.98	40.00	-.436	.663	-.097
Male	7	6.32				

4.6 RQ3: Do stimulus properties (i.e., linguistic features, elicitation task, sample length) relate to comprehensibility or comprehension?

To further investigate potential hindrances to NNS listeners' understanding of L2 academic speech, the impact of speech linguistic properties, elicitation task and sample length on comprehensibility and comprehension were examined.

4.6.1 Linguistic features

EFL teachers' assessments of L2 speech samples in six⁴ linguistic categories and listeners' judgments of comprehensibility and comprehension were submitted to a Spearman's rho test. This analysis, the results of which are summarized in Table 9, revealed an inverse moderately-sized correlation between comprehensibility values and segmental errors, and a similar mid-sized negative correlation between comprehension scores and lexical richness. In other words, listeners' perceived understanding of speakers' L2 speech improved with the fewer vowel and consonant pronunciation errors committed, while their actual understanding appeared to benefit from speakers' usage of simpler vocabulary.

Table 9

Association of Linguistic Feature Ratings with Comprehensibility and Comprehension

Linguistic features	N	Comprehensibility		Comprehension	
		<i>r_s</i>	<i>p</i>	<i>r_s</i>	<i>p</i>
Segmental errors	20	-.449	.047*	-.093	.696
Stress errors	20	-.229	.331	-.271	.248
Lexical appropriateness	20	-.264	.260	-.316	.175
Lexical richness	20	-.054	.820	-.468	.038*
Grammatical accuracy	20	-.221	.348	-.395	.084
Grammatical complexity	20	-.137	.563	-.247	.294

Note: * $p < .050$

4.6.2 Elicitation task

Comprehensibility judgments were based on a single speech sample per speaker, and as such no measurement of task effect on perceived understanding was required. Listener comprehension, on the other hand, was assessed using two short answer questions (Q1 & Q2) based respectively on speech tasks Prompts 2 and 3, thus facilitating comparisons between task impact on elicited speech and/or listener performance.

Mean comprehension question scores revealed only slight difference between listeners' assessed understanding of Q1 ($M=2.85$, $SD=.54$) and Q2 ($M=3.18$, $SD=.40$). Following Crowther et al.'s (2015) approach in exploring task effect through comparison of listener assessments' linguistic correlates, a Spearman's correlation test was conducted on Q1 and Q2 scores along with rated speech linguistic features. It showed that listeners' understanding of Prompt 3-elicited speech (as measured by Q2) had the same moderate negative association with lexical richness ($r_s=-.46$, $p<.050$) as overall comprehension scores (discussed in Section 4.6.1), while Q1 scores did not correlate with any linguistic features. Taken together, these findings demonstrate that although Prompt 3 may have pushed speakers to produce more lexically diverse L2 speech than Prompt 2, the overall effect on listener comprehension was negligible.

4.6.3 Sample length

Although the stimuli factor of recording length is less relevant to the EMI classroom setting where monologic speech segments tend to be quite lengthy, its consideration in this investigation serves to account for potential unintended effects resulting from methodological design. Due to the small variance in speech sample lengths used for the comprehensibility rating process ($M=32s$, $SD=3.52$), a Spearman's correlation test was conducted but no significant association was found between the two variables [$r_s(20)=.12$, $p=.608$]. Comprehension question speech sample lengths had greater variance ($M_{Q1}=124s$, $SD_{Q1}=25.75$; $M_{Q2}=133s$, $SD_{Q2}=27.22$), and as such were divided into three groups for each question set (see Table 10). Although the first question's median comprehension scores appeared to decrease with increasing recording

length, a Kruskal Wallis H (KWH) test did not find any significant effect, a finding similarly reflected with recording lengths' impact on second comprehension question's scores.

Table 10

Effect of Sample Lengths on Comprehension

Sample lengths	Comprehension Question 1					
	<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>X</i> ²	<i>p</i>
59-99s	6	3.11	3.21	.49		
100-118s	7	2.71	2.97	.63	2.87	.238
119-159s	7	2.76	2.74	.46		
Sample lengths	Comprehension Question 2					
	<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>X</i> ²	<i>p</i>
63-96s	7	3.15	3.21	.28		
97-119s	6	3.40	3.52	.26	2.90	.234
120-208s	7	3.02	3.10	.55		

4.7RQ4: Does listener L2 ability relate to comprehensibility and comprehension?

Based on students' self-ratings of their L2 ability, one identified as CEFR level A2, 24 as B1, seven as B2, and four as C1. No one selected either extreme of A1 or C2. A KWH test showed that listener ELP had a small but significant effect on comprehensibility ratings [$N=720$, $X^2(4)=59.32$, $p<.001$, $\eta^2=.083$] but not on comprehension scores [$N=720$, $X^2(4)=4.28$, $p<.233$, $\eta^2=.006$]. Post hoc MWU tests were conducted for all possible L2 ability level pairings for comprehensibility and comprehension, a total of six comparisons between four groups (A2-C1). To control for Type 1 errors, a Bonferroni adjustment was applied (Pallant, 2016), decreasing the acceptable significance threshold from .05 to $.05/6=.008$. Table 11 summarizes the results of the post hoc analysis.

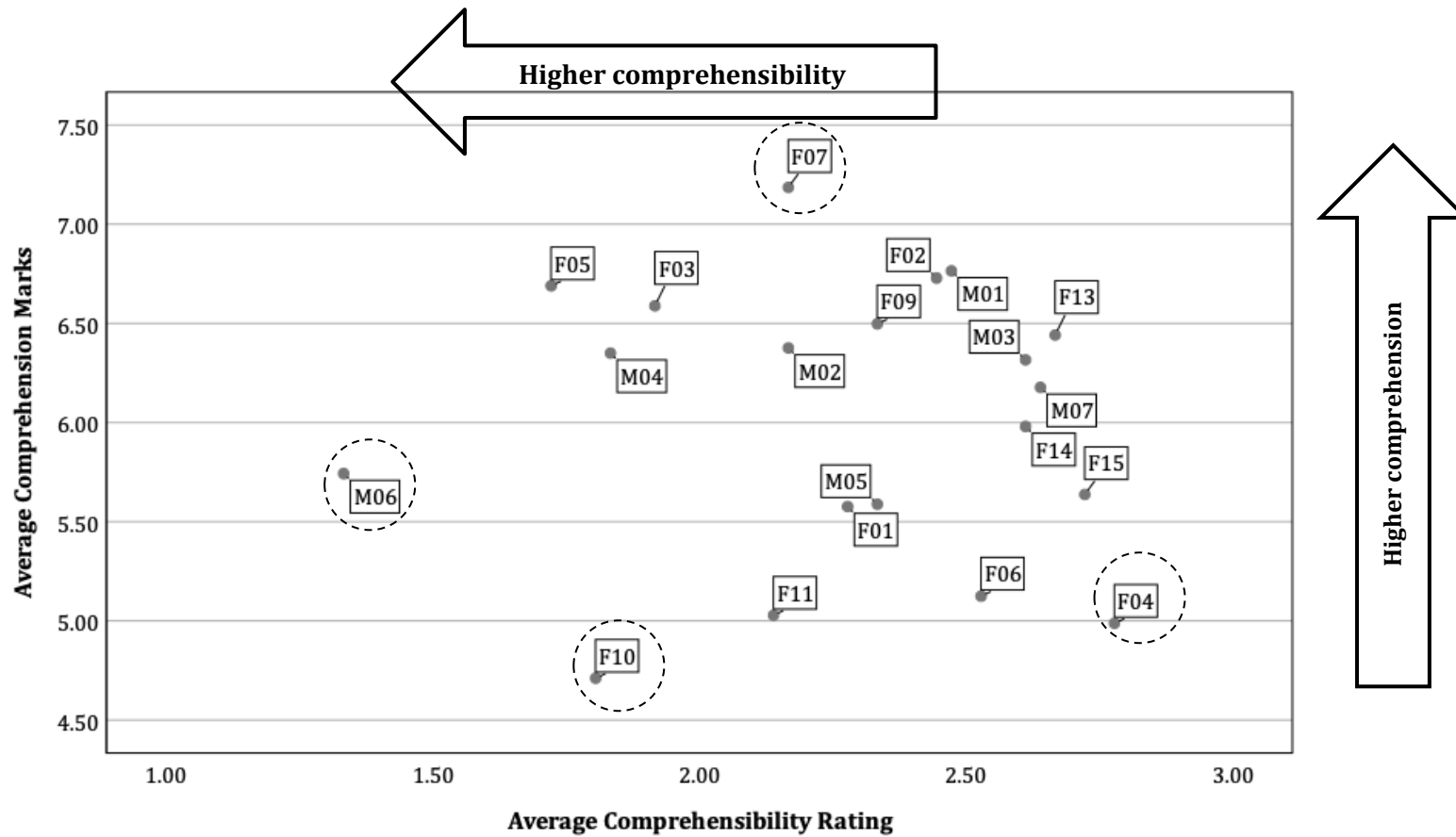
Table 11

Effect of Listener L2 Ability on Comprehensibility and Comprehension

		A2	B1	B2	C1	Comprehensibility
A2	<i>U</i>		1239.50	610.50	88.50	
	<i>p</i>		.000	.000	.000	
	<i>r</i>		-.268	-.329	-.655	
B1	<i>U</i>	4495.50		26621.00	16389.50	
	<i>p</i>	.629		.000	.024	
	<i>r</i>	-.022		-.159	-.096	
B2	<i>U</i>	1168.50	29969.50		3516.00	
	<i>p</i>	.230	.050		.000	
	<i>r</i>	-.095	-.079		-.325	
C1	<i>U</i>	734.50	18874.00	5078.50		
	<i>p</i>	.570	.807	.249		
	<i>r</i>	-.057	-.010	-.078		
Comprehension						

Note: Highlighted cells indicate $p < .008$ (Bonferroni corrected)

With the exception of the C1-B1 pairing, comprehensibility ratings differed significantly between CEFR levels with effect sizes ranging from small to large. Comprehension scores, on the other hand, did not vary noticeably with L2 ability. Generally, the higher the listeners' self-rated ELP, the easier they perceived understanding speakers' L2 speech. However, this perception may have been deceiving, as all listeners appeared more evenly matched in their actual understanding of the L2 speech, casting doubt on the veracity of their L2 ability self-assessments.



Note: Comprehensibility scale: 1=very easy to understand, 9=very difficult to understand. Comprehension marks: 0-8 points.

Figure 1. Average comprehension versus comprehensibility per speaker

4.8 Further analysis of specific cases of comprehensibility / comprehension

To better understand how the explanatory power of the aforementioned correlates with perceived and actual understanding of speakers' L2 oral production, an in-depth examination of several cases was undertaken. A scatterplot graph (see Figure 1) was used to identify cases (labelled by speaker ID) for further investigation, with the specific ones of interest circled.

4.8.1 Most comprehensible

Speaker M06 was perceived as the easiest to understand ($M_{\text{comprehensibility}}=1.33$) and yet exemplifying comprehensibility's lack of association with comprehension, listeners struggled more in grasping his main points ($M_{\text{comprehension}}=5.74$ or 71.8%) than with most other speakers. M06's high comprehensibility can be explained in part by a higher ELP (IELTS score=6.0), which was weakly associated with better comprehensibility ratings. His L2 speech was also deemed to have fewer segmental pronunciation errors than 80% of other speakers' L2 oral production, a linguistic feature factoring significantly in participants' ratings of perceived understanding. These clues provide a solid explanation as to the speaker's exemplary comprehensibility, but understanding the lacklustre comprehension result will require item analysis of the testing questions.

Respondents were required to identify four key items in their answers to each comprehension question based on a corresponding L2 speech sample. In M06's case, the items with the lowest facility values were from Comprehension Question 1 (Q1): Item 3 ($P=.27$) and Item 2 ($P=.53$). For Item 3, the required answer incorporated the technical phrase "factor analysis" (see Appendix G. Answer Key: M06.Q1.Item3), which caused many students to either not respond or to supply a phonetically similar but incorrect word (e.g., "vector"). The speaker listed "matrix calculations" (see Appendix G. Answer Key: M06.Q1.Item2) for Item 2's answer, which half of respondents struggled to correctly hear and transcribe. Although M06 only achieved a mid-level rating for lexical richness, his use of discipline-specific academic vocabulary causing

listener confusion appears to confirm the earlier-discovered negative association between speakers' breadth of vocabulary usage and listeners' actual understanding.

4.8.2 Least comprehensible

In comparison with other speakers, listeners subjectively rated F04's L2 speech as the most difficult to understand, although in the context of the 9-point measurement scale she was still regarded as fairly comprehensible ($M_{\text{comprehensibility}}=2.78$). In contrast, listeners appeared to comprehend less than two thirds of what F04 said ($M_{\text{comprehension}}=4.99$ or 62.4%). Despite F04 being one of the few participants with prior EMI teaching experience, this factor that did not appear to benefit listeners' perceived or actual understanding of her instruction. The speaker's relatively poor comprehensibility rating is somewhat surprising in light of her L2 ability (IELTS=6.0) ranking in the upper 50% of speakers' test scores, but it suggests that the association discovered between IELTS scores and perceived understanding ratings should be treated with caution.

EFL teachers assessed F04's L2 speech as having the highest frequency of segmental errors, which appears to provide explanation as to the low comprehensibility given its negative association with the linguistic feature. However, a detailed examination of the speech sample used for comprehensibility testing (see Appendix D for transcript) reveals a paucity of vowel and/or consonant errors, with only two that could be regarded as noteworthy in influencing subjective ease of understanding ratings (/u:/ in "duty" pronounced as /æ/; /l/ in "firstly" replaced with /r/). Thus, comprehensibility's negative association with mispronunciation frequency should be handled with caution as it appears that other factors may have also contributed to listener judgments of perceived understanding.

F04's L2 speech ranked fourth for breadth of vocabulary usage, which provides some explanation of listeners' failure to grasp one third of what she said. Item analysis supported this claim, revealing that the lowest-facility item [P(Q1.

Item3)=.42] incorporated rich lexical items such as “psychopathological rule” in its answer (see Appendix G. Answer Key: F04.Q1.Item3).

4.8.3 Best comprehended

While listeners regarded F07’s L2 speech as requiring less effort to understand than a slight majority of other speakers ($M_{\text{comprehensibility}}=2.17$), their actual understanding of the key information she communicated was better than for any other speaker ($M_{\text{comprehension}}=7.19$ or 89.9%). The speaker’s ‘upper tier’ IELTS band of 6.0 aligns with the above average perceived understanding rating, but conflicting evidence, in the form of F07’s relatively frequent mispronunciation of vowels and consonants (as denoted by fifth highest rating for segmental errors), suggests that a mix of factors influenced listeners’ judgments of comprehensibility.

Listeners’ comprehension of F07’s L2 speech appeared to have benefited from her usage of basic vocabulary, as evidenced by her receiving the second lowest rating for the linguistic feature of lexical richness. Interestingly, F07 also rated poorly in appropriate lexical usage (third lowest), grammatical complexity (third lowest), and grammatical accuracy (fourth lowest) and yet was the best understood. These findings suggest that complex lexical usage does more to hinder comprehension than a lack of lexicogrammatical accuracy.

4.8.4 Least comprehended

For the final case examined, the speaker’s L2 speech was subjectively rated as relatively easy to understand ($M_{\text{comprehensibility}}=1.81$) but objectively assessed as the most difficult, with a mean comprehension score of 4.39. F10’s IELTS band of 5.5 places her in the lower L2 ability group, providing yet another exception to the association between listener comprehensibility and standardized measure of speaker ELP. Likewise, her speech’s mid-level rating for segmental errors suggests that factors other than the ones identified earlier can explain the relatively high comprehensibility.

At first glance, the poor listening comprehension cannot appear to be attributed to the speaker’s breadth of vocabulary usage given that F10’s speech was rated

the fifth lowest for lexical richness. However, analysis of several of the lowest-scored comprehension items [$P(Q1.Item2)=.087$; $P(Q2.Item1)=.098$] reveals that the use of an unfamiliar technical term “forecast” largely explains listener confusion, thus confirming the inverse relationship between understanding and lexical richness.

5. Discussion

This chapter addresses inter-rater reliability issues encountered with measurement of linguistic features. It then discusses the results in light of the study's research questions investigating perceived and actual comprehension of L2 academic speech within an Indonesian EMI context, and speaker, stimulus and listener factors influencing those variables.

5.1 Linguistic ratings reliability

Although EFL teachers' assessments of L2 speech linguistic features were reliably consistent in six categories ($.85 \leq \alpha \leq .93$), they failed to align sufficiently for measurement of intonation ($\alpha = -.40$) and speech rate ($\alpha = -.01$) despite raters' feedback (based on 9-point Likert scales) that they both understood ($M_{\text{intonation}} = 7.33$; $M_{\text{speechrate}} = 6.67$) and were comfortable ($M_{\text{intonation}} = 7.33$; $M_{\text{speechrate}} = 7.00$) using these measures. A possible explanation for the inconsistencies could be that vague scale descriptions led to subjective interpretations based on each EFL teacher's perceived standard for intonation and speech rate. For example, the instrument's explanation for intonation defines it as "the natural pitch changes that occur when we speak" (Saito et al., 2016, p. 235), but this speech prosody can vary significantly even for NSs based on factors such as English variety and gender (cf., Clopper & Smiljanic, 2011). Likewise, the target standard for speech rate is defined within the instrument as "sounding natural" or "comfortable to listen to" (Saito et al., 2016, p. 235), both of which could elicit widely different ratings depending on factors such as EFL teachers' familiarity with the English variety being spoken and personal preference.

5.2 RQ1: Do listeners' comprehensibility judgments relate to their assessed comprehension of L2 speech?

The lack of association between listeners' perceived ease of understanding and their actual comprehension of L2 speech [$r_s(20) = -.12, p = .626$], a result largely reflected at the individual level, contrasted with Orikasa's (2016) observation of a positive trend between L1 Japanese students' subjective (5-point Likert scale) and objective (five short answer questions) comprehension judgments of NNS/NS speech, with effect size ranging from moderate to large ($.37 \leq r \leq .64$,

$p < .05$). However, her finding extended to only five out of the eight speakers, two of whom were NS, suggesting that her evidence is somewhat inconclusive. Conversely, this study's finding confirmed Van der Walt's (2000) conclusion, derived from her comparison of 140 mixed NS/NNS listeners' perceived comprehensibility (4-point Likert scale) and comprehension (short answer question) of South African accented L2 speech, that, "a subjective rating...of whether the respondent thinks the speaker is easy to understand is not always indicative of real understanding" (p. 147). Likewise, Smith and Nelson (2006) showed that for more than 60% of mixed NS-NNS pairs' ($N=5$) conversations, NNS and NS listeners' ($N=29$) perceptions of understanding failed to match their performance in answering three M/C questions based on the 10-minute samples.

Although methodological issues with data collection (i.e., rater fatigue, playback problems) may have accounted for some variance in comprehension scores, they lack sufficient power to explain the extent of the two measures' divergence. A more likely reason for the result is that given the listeners' shared L1 background, their familiarity with speakers' L2 variety caused them to "believe that they [understood] most of what they [heard]" (Smith & Nelson, 2006), while other more powerful factors such as average L2 listening ability (self-rated as mid- to high-intermediate, but see Section 5.5 for discussion on accuracy of these self-assessments) and L2 speech characteristics influenced their actual comprehension. Regardless, this result, combined with wider research's conflicting findings call into question an underlying assumption made in many studies that subjective judgments of comprehensibility are representative of deeper understanding (e.g., Hellekjaer, 2010). Levis (2006, p. 254) concurs with this conclusion, stating that "comprehensibility...is not a measure of what is actually understood. It is instead a measure of how comfortable a speaker is to listen to."

5.3 RQ2: Do speaker factors (i.e., L2 ability, EMI experience, L2 immersion experience, gender) relate to comprehensibility and comprehension?

The second research question was primarily concerned with investigating whether speakers' L2 proficiency (as denoted by IELTS scores) was a predictor

of how well their speech was understood by listeners, although other independent factors' potential impact was also examined for thoroughness' sake.

5.3.1 L2 ability

Speakers with higher IELTS scores ($N=10$; 6.0-7.0) were perceived to be marginally easier to understand than those from the lower-scoring group ($N=10$; 4.0-5.5) (but see Sections 4.8.2 and 4.8.4 for exceptions to this finding), while no significant association was found between listeners' comprehension and speakers' IELTS scores. Given that the scalar judgments used in IELTS's L2 oral proficiency testing have been likened to the comprehensibility scale (Crowther et al., 2015; Isaacs & Trofimovich, 2012), it follows that in their comprehensibility assessments this study's listeners were drawn to notice similar linguistic features as their professional IELTS rater counterparts, albeit on a more basic level. Research identifies these linguistic correlates of IELTS oral proficiency measurements as pronunciation and fluency (Crowther et al., 2015), the first of which also associated with comprehensibility in this study (see Section 4.6.1 for analysis results). Although no empirical research corroborated this IELTS-comprehensibility relationship, Kraut and Wulff (2013) confirmed that speaker ELP (as denoted by 24 NS/NNS L1 Hispanic, Asian, Middle-Eastern speakers' placement in either beginner, intermediate, or advanced ESL classes), in interaction with factors such as speaker gender and L1 family, associated positively [ANOVA: $F(2,41860)=226.936$; $p<.001$] with comprehensibility ratings (7-point Likert scale).

The fact that listener comprehension did not noticeably improve for higher IELTS-band speakers suggests that other factors, including non-phonological linguistic features, had greater influence on scores. Similar to comprehensibility, no studies were discovered that explored the IELTS-comprehension dynamic.

5.3.2 EMI experience

Surprisingly, the benefit of prior EMI experience did not result in improved student understanding of speakers' L2 speech. In fact, some speakers who had spent several years teaching in EMI classrooms received the lowest comprehension scores and comprehensibility ratings (see Section 4.8 for

examples). One explanation is that increased confidence gained from using L2 to instruct on their academic discipline may have caused some in-service EMI lecturers to introduce more technical terms and details, thereby lowering overall listener comprehension. Another reason may be that despite participation in EMI teaching initiatives, lecturers have not been provided with adequate training and support to improve their L2 competence, resulting in further fossilization of their phonological and lexicogrammatical errors. No research was found that either confirmed or challenged the lack of association between EMI experience and listener perceived/actual understanding, although in a similar vein a Danish study (Jensen et al., 2016) also found a lack of association between L2-mediated teaching experience and students' perception of EMI instructors' lecturing competence.

5.3.3 L2 immersion experience

Speakers with two or more years' experience living in an L2-speaking context ($n=5$) did not produce noticeably more comprehensible or comprehensible speech than peers lacking this background characteristic. This finding could be partially explained by the fact that 60% of those with immersion experience spent their time in Malaysia where the local language is quite similar to speakers' L1, and consequently may have not have experienced as much quality ELP development. This reason aligns with Derwing et al.'s (2008) hypothesis that quality rather than quantity of L2 exposure determined gains in comprehensibility based on their longitudinal study finding of L1 Mandarin and L1 Slavic learners' differing progress in producing comprehensible speech.

Another factor that may have influenced this outcome is the time that elapsed between speakers' immersion in L2-speaking contexts and this study's investigation of comprehensibility and comprehension of their L2 speech. Three speakers reported that their time abroad took place 3 to 19 years ago, which could explain the contrast with a three-year study finding that 85% of ESL students' comprehensibility improved over the course of their time in Canada (Kennedy et al., 2012).

5.3.4 Gender

No statistical difference was found between listeners' perceived and actual understanding of male ($n=7$) and female ($n=13$) speakers' speech. This result appears to disagree with Kraut and Wulff's (2013) finding that speaker gender had a statistically significant effect on comprehensibility judgments [ANOVA: $F(1,41860)=93.754$; $p<.001$]. However, the American-context study also found this factor's impact to be mitigated by speaker ELP such that at higher L2 ability, males and females were rated similarly. Given that many of this study's speakers were upper intermediate L2 ability ($M_{IELTS}=5.7$), it is plausible that any gender-related pronunciation variability was muted due to higher ELP. Another explanation could be that the NNS listeners' familiarity with speakers' L2 variant and their mid-level self-rated ELP led them to notice other more salient features in their evaluations rather than gender-based pronunciation differences.

5.4RQ3: Do stimulus properties (i.e., linguistic features, elicitation task, sample length) relate to comprehensibility and comprehension?

According to the relevant literature, speech characteristics factor prominently in judgments of listener understanding and thus were deemed important to address in this study.

5.4.1 Linguistic features

In rating speakers' comprehensibility, listeners appear to have been more sensitive to incorrectly omitted, added, or changed vowels and/or consonants. This moderate association ($r_s=-.45$, $p=.047$) generally reflects broader research's discoveries in this area, including the impact of segmental errors on both NSs' (Munro & Derwing, 2001) and NNSs' (Jułkowska & Cebrian, 2015) judgments of perceived understanding, the association of comprehensibility with pronunciation features for IELTS's monologic speech task (Crowther et al., 2015), and the impact of segmental accuracy on comprehensibility ratings at all levels of speaker ELP (Saito et al., 2016). However, these and other studies have also linked comprehensibility to almost all of the other linguistic factors investigated, including word stress (Isaacs & Trofimovich, 2012; Kang et al., 2010; Saito et al., 2016), intonation (Kang et al., 2010; Saito et al., 2016), speech rate (Munro & Derwing, 2001), lexical richness (Isaacs & Trofimovich, 2012), and

lexicogrammatical accuracy (Munro & Derwing, 1999; Trofimovich & Isaacs, 2012). There are several reasons that could explain this study's finding that comprehensibility ratings associated solely with segmental errors. First, listeners' mid-level L2 ability (self-rated average of intermediate) may have drawn their attention more towards conspicuous pronunciation features rather than noticing lexicogrammatical accuracy, a supposition that requires further investigation. Second, the shared L1 background between speakers and listeners may have had a similar affect to listeners' ELP in that they unconsciously missed or ignored non-phonological linguistic features in assessing speaker comprehensibility. A final reason is that the use of only three EFL teachers in rating linguistic features may have resulted in individual biases being more pronounced in linguistic categories, thereby distorting potential associations.

This study also found that speakers' breadth of vocabulary usage (denoted by linguistic category lexical richness) inversely associated with listeners' comprehension ($r_s = -.47, p = .038$), which seems intuitive given that listeners' processing load increases when encountering complex or unfamiliar lexical items, thereby reducing overall understanding (Révész & Brunfaut, 2013). Surprisingly, grammatical complexity did not relate to comprehension, perhaps because speakers opted for simpler grammatical constructions in their communication. Also, other pronunciation, fluency, and lexicogrammatical accuracy speech characteristics may not have had significant impact on listener comprehension due to the shared L1 with speakers.

5.4.2 Elicitation task

Listeners' actual understanding of Prompt 3-elicited L2 speech ($M_{Q2} = 3.18, SD_{Q2} = .40$) was marginally higher than that of Prompt 2-elicited speech ($M_{Q1} = 2.85, SD_{Q1} = .54$) despite the former's negative correlation with speakers' breadth of vocabulary usage ($r_s = -.46, p < .050$). Although the differing linguistic correlates suggest varying speech task complexities, as posited by Crowther et al. (2015) in their finding of differing associating linguistic features between rater perceived understanding judgments of IELTS long turn and TOEFL speaking tasks, the impact on listener comprehension, as measured by respective question scores,

was negligible. This result does not seem to reflect Jensen and Thøgersen's (2017) finding that task complexity significantly impacted listening comprehension.

One reason for these seemingly divergent findings is that speech tasks used in this study did not significantly differ from each other in complexity, thus minimizing task effect on comprehension scores (see Appendix C for complete task descriptions). The comprehension questions' differing linguistic correlates despite similar task complexity could be explained by speakers' increased use of discipline-specific jargon for Prompt 3 (Question 2) speech.

5.4.3 Sample length

Statistical analyses revealed that recording length had no significant effect on judgments of comprehensibility or comprehension. Although this finding appears to conflict with research suggesting that listening comprehension benefits from shorter L2 stimuli due to WM constraints (cf., Carpenter, Miyake, & Just, 1994; Just & Carpenter, 1992), one explanation could be that variation in this study's sample lengths was too small to noticeably impact comprehensibility or comprehension assessments.

5.5 RQ4: Does listener L2 ability relate to comprehensibility and comprehension?

Listeners with higher ELP tended to gauge perceived understanding more leniently than those with lower ability, while no such distinction was discovered for comprehension scores. In their study of the processing time mixed NNS/NS raters took in making comprehensibility judgments of speech samples from six (NS, L1 Catalan, German) speakers, Ludwig and Mora (2017) also found that higher ELP listeners assessed speakers as easier to understand than did lower ELP peers (ANOVA: $F=18.60$, $p<.001$, $\eta^2=.352$). However, this finding of listener ELP's relationship with comprehensibility ratings was not reflected in other studies' results (e.g., Jułkowska & Cebrian, 2015; Saito et al., 2019); this difference could be explained by these investigations' operationalization of L2 ability based on raters' productive rather than receptive skills.

The lack of correlation between listener ELP and comprehension appears counter-intuitive, as one would expect higher L2 proficiency listeners to comprehend more than their lower proficiency counterparts, as other studies have found (e.g., Smith & Nelson, 2006). One explanation could be that L2 ability self-ratings were not entirely accurate as some participants may have either been over- or under-confident in their reporting, leading to discrepancies with their actual comprehension task performance.

6. Conclusion

This chapter presents a summary of the key findings, discusses implications for developing EMI programs and training EMI instructors, notes the research limitations, and recommends areas for further investigation.

6.1 Summary of key findings

As expected, L1 Indonesian lecturers' English academic speech was fairly comprehensible to shared L1 background undergraduate students. Listeners' actual grasp of what was communicated was lower, however, and did not associate with subjectively-rated ease of understanding. Speaker ELP, as measured by IELTS scores, generally predicted perceptions of their speech comprehensibility, but did not relate to listener comprehension. Likewise, listeners' self-estimates of L2 ability showed fairly strong association with perceived understanding but not at all with deeper understanding. Surprisingly, neither EMI teaching nor L2 immersion experience factored into either perceived or actual understanding of speakers' L2 oral production. The two measured aspects of understanding related to different linguistic features, with comprehensibility moderately associating with segmental accuracy, and comprehension inversely associating with breadth of lexical usage.

6.2 Implications for EMI program development

NS model-based standardized ELP tests appear to be inappropriate for judging whether NNS students will understand EMI lecturers' classroom speech. Rather than using IELTS or TOEFL to measure pre- or in-service lecturer L2 competence, serious EMI program implementations should undertake an appropriate assessment of instructors' linguistic and pedagogical abilities (Huang & Singh, 2014). One such measure from the Danish EMI context is Kling and Staehr's (2012) Test of Oral English Proficiency for Academic Staff (TOEPAS). Consisting of three parts (warm-up, mini-lecture, question and answer), the testing procedure involves a combination of peer lecturers and trained L2 examiners to assess test-takers' fluency, pronunciation, lexicogrammar, and interaction skills. In their validation study of TOEPAS, Dimova and Kling (2018) list other examples of European universities initiating in-house test procedures for EMI teachers.

Given that lecturer EMI experience did not impact student comprehension, program planners should avoid assuming that prior teaching experience signifies higher instructor ELP (e.g., Vinke, Snippe, & Jochems, 1998 – recommended using veteran versus novice EMI educators based only on in-class experience). Instead, regulation of lecturers' ELP should be undertaken using testing procedures such as TOEPAS. Additionally, both experienced instructors and L2 teachers could mentor newer EMI lecturers (Fenton-smith et al., 2017).

6.3 Implications for EMI lecturer L2 and pedagogical training

Comprehensibility's correlation with a pronunciation-related speech characteristic and lack of association with comprehension suggest that addressing lecturers' L2 pronunciation may not benefit students' actual understanding of instructional content. Although this may appear to contradict ELF research identifying phonology as core to NNSs' mutual intelligibility, it should be noted that in this study, speakers' and listeners' shared L1 negated the impact of pronunciation variation arising from diverse L1 backgrounds.

In contrast, comprehension's negative correlation with perceived speech lexical richness suggests that instructors would benefit from EMI pedagogical training focusing on skills such as explaining new and unfamiliar vocabulary, and paraphrasing complex statements to encourage listener understanding. This recommendation concurs with other findings that greater methodological awareness, rather than increased L2 pronunciation competence, is the more acute need for EMI lecturers' professional development (e.g., Ball & Lindsey, 2013).

6.4 Study limitations

Perhaps the most significant limitation of this study is the uniformity of speaker and listener L1 backgrounds. Although this may prevent generalization to more linguistically diverse EMI classrooms, uniform L1 background was selected as representative of regional HE institutions' EMI classroom realities (Murata & Iino, 2018). Furthermore, listeners' and speakers' academic discipline diversity necessitated elicitation of more generalized EMI classroom speech, which may

have benefited resulting understanding. Finally, methodological limitations of listener gender bias (88.9% female), data collection issues (e.g., noisy environment, network-induced sample playback errors), and EFL teachers' less rigorous training in instrument usage (as compared with prescribed procedure in Saito et al., 2016) may have impacted findings and their generalizability.

6.5 Areas for further study

Although this study contributed to awareness of factors influencing EMI lecturers' understandability in a shared L1 EMI context, a repetition of the investigation with listeners from varied L1 backgrounds could help to strengthen conclusions' generalizability within the EMI domain. In light of EMI research highlighting the dual importance of lecturer L2 and pedagogical competences, it could be beneficial to explore actual in-class comprehension breakdowns compared with linguistic and instructional methodology factors. Finally, this study's finding of comprehensibility's lack of association with a measure of deeper understanding calls for more research clarifying the relationship between subjective and objective measures of understanding in high stakes contexts.

References

- Airey, J. (2011). Talking about teaching in English: Swedish university lecturers' experiences of changing teaching language. *Iberica*, 22, 35–54. Retrieved from http://www.aelfe.org/documents/02_22_Airey.pdf
- Aritonang, M. S. (2013, January 14). Int'l-standard schools must disband by April. *The Jakarta Post*. Retrieved from <http://www.thejakartapost.com/>
- Audacity [Software application]. (2018). Retrieved from <https://audacityteam.org/>
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Ball, P., & Lindsay, D. (2013). Language demands and support for English-medium instruction in tertiary education: Learning from a specific context. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-medium instruction at universities: Global challenges* (pp. 44–61). Retrieved from <https://ebookcentral.proquest.com>
- Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, 114(3), 1600–1610. doi:10.1121/1.1603234
- Berns, M. (2008). World Englishes, English as a lingua franca, and intelligibility. *World Englishes*, 27(3), 327–334. doi:10.1111/j.1467-971X.2008.00571.x
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, 15, 39–54. doi:10.1016/0167-6393(94)90039-6
- Cambridge English Language Assessment. (1989). IELTS. Retrieved from <https://www.ielts.org/>

- Carpenter, P. A., Miyake, A., & Just, M. A. (1994). Working memory constraints in comprehension: Evidence from individual differences, aphasia, and aging. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.
- Chaudron, C. (2003). Data collection in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 762–828). doi:10.1002/9780470756492
- Clopper, C. G., & Smiljanic, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics*, 39(2), 237–245. doi:10.1016/j.wocn.2011.02.006
- Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Council of Europe. (2001). Self-assessment grid (CEFR). Strasbourg: Council of Europe. Retrieved from <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168045bb52>
- Crookes, G. (1991). Second language speech production research: A methodologically oriented review. *Studies in Second Language Acquisition*, 13, 113–132. doi:10.1017/S0272263100009918
- Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2015). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly*, 49(4), 814–837. doi:10.1002/tesq.203
- Dearden, J. (2014). *English as a medium of instruction – a growing global phenomenon*. London: British Council. Retrieved from <https://www.britishcouncil.org/>

- Denver, L., Jensen, C., Inger, M. M., & Werther, C. (2016). Good enough to teach? A study of EMI lecturers' language skills and metadiscourse. *Moderna Språk, 2*, 46–72. Retrieved from <http://ojs.uib.no/ojs/index.php/modernasprak/>
- Deroey, K. L. B., & Taverniers, M. (2011). A corpus-based study of lecture functions. *Moderna Språk, 105*(2), 1–22. Retrieved from <http://ojs.uib.no/ojs/index.php/modernasprak/>
- Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*(1), 1–16. doi:10.1017/S0272263197001010
- Derwing, T. M., Munro, M. J., & Ron, T. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics, 29*(3), 359–380. doi:10.1093/applin/amm041
- Deterding, D., & Kirkpatrick, A. (2006). Emerging South-East Asian Englishes and intelligibility. *World Englishes, 25*(3), 391–409. doi:10.1111/j.1467-971X.2006.00478.x
- Dewi, A. (2017). English as a medium of instruction in Indonesian higher education: A study of lecturers' perceptions. In B. Fenton-Smith, P. Humphreys, & I. Walkinshaw (Eds.), *English medium instruction in higher education in Asia-Pacific: From policy to pedagogy* (pp. 241–258). doi:10.1007/978-3-319-51976-0
- Doiz, A., Lasagabaster, D., & Sierra, J. M. (2012). Future challenges for English-medium instruction at the tertiary level. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-medium instruction at universities: Global challenges* (pp. 213–221). Retrieved from <https://ebookcentral.proquest.com>
- Educational Testing Service. (2005). TOEFL iBT: About the test. Retrieved from <https://www.ets.org/toefl/ibt/about>

- Educational Testing Service. (2019). The TOEIC program. Retrieved from <https://www.ets.org/toEIC>
- Edwards, J. G. H., Zampini, M. L., & Cunningham, C. (2018). The accentedness, comprehensibility, and intelligibility of Asian Englishes. *World Englishes*, *37*, 538–557. doi:10.1111/weng.12344
- Eisenbeiss, S. (2010). Production methods in language acquisition research. In E. Blom, & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 11–34). Retrieved from <https://ebookcentral.proquest.com>
- Fenton-smith, B., Stillwell, C., & Dupuy, R. (2017). Professional development for EMI: Exploring Taiwanese lecturers' needs. In B. Fenton-Smith, P. Humphreys, & I. Walkinshaw (Eds.), *English medium instruction in higher education in Asia-Pacific: From policy to pedagogy* (pp. 195–217). doi:10.1007/978-3-319-51976-0_11
- Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*(5), 3125–3134. doi:10.1121/1.413041
- Gass, S. M., & Mackey, A. (2011). *Data elicitation for second and foreign language research*. New York: Routledge.
- Gill, S. K., & Kirkpatrick, A. (2013). English in Asian and European higher education. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1999–2002). doi:10.1002/9781405198431.wbeal0383
- Green, R. (2013). *Statistical analyses for language testers*. New York: Palgrave Macmillan.

- Harding, L., & McNamara, T. (2017). Language assessment: The challenge of ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca* (pp. 570–582). Retrieved from <https://ebookcentral.proquest.com>
- Hellekjaer, G. O. (2010). Assessing lecturer comprehension in Norwegian English-medium higher education. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 233-258). Philadelphia, PA: John Benjamins Publishing Company.
- Huang, D., & Singh, M. (2014). Critical perspectives on testing teaching: Reframing teacher education for English medium instruction. *Asia-Pacific Journal of Teacher Education*, 42(4), 363–378. doi:10.1080/1359866X.2014.956046
- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic Influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34, 475–505. doi:10.1017/S0272263112000150
- Jenkins, J. (2000). *The phonology of English as an international language: New models, new norms, new goals*. Oxford: Oxford University Press.
- Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23(1), 83–103. doi:10.1093/applin/23.1.83
- Jenkins, J. (2006). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly*, 40(1), 157–181. doi:10.2307/40264515
- Jenkins, J. (2007). *English as a lingua franca: Attitude and identity*. Oxford: Oxford University Press.

- Jenkins, J., Cogo, A., & Dewey, M. (2011). Review of developments in research into English as a lingua franca. *Language Teaching*, 44(3), 281–315.
doi:10.1017/S0261444811000115
- Jensen, C., & Thøgersen, J. (2017). Foreign accent, cognitive load and intelligibility of EMI lectures. *Nordic Journal of English Studies*, 16(3), 107–137. Retrieved from <http://ojs.ub.gu.se/ojs/index.php/njes/>
- Jonge, P. De, & Jong, P. F. De. (1996). Working memory, intelligence and reading ability in children. *Personality and Individual Differences*, 21(6), 1007–1020.
doi:10.1016/S0191-8869(96)00161-4
- Jułkowska, I. A., & Cebrian, J. (2015). Effects of listener factors and stimulus properties on the intelligibility, comprehensibility and accentedness of L2 speech. *Journal of Second Language Pronunciation*, 2(2015), 211–237.
doi:10.1075/jslp.1.2.04jul
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149. doi:10.1037/0033-295X.99.1.122
- Kang, O., Rubin, D. O. N., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94(4), 554–566. doi:10.1111/j.1540-4781.2010.01091.x
- Kaur, J. (2010). Achieving mutual understanding in World Englishes. *World Englishes*, 29(2), 192–208. doi:10.1111/j.1467-971X.2010.01638.x
- Kennedy, S., Foote, J. A., & Kurtz, L. (2012). Second language speakers at university: Longitudinal development and rater behaviour. *TESOL Quarterly*, 49(1), 199–209. doi:10.1002/tesq.212

- Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 Speech: The role of listener experience and semantic context. *The Canadian Modern Language Review*, 64(3), 459–489.
doi:10.3138/cmlr.64.3.459
- Kero, P., & Lee, D. (2016). Likert is pronounced "LICK-urt" not "LIE-kurt" and the data are ordinal not interval. *Journal of Applied Measurement*, 17(4), 502–509.
Retrieved from <http://jampress.org/>
- Kirkpatrick, A. (2014). English as a medium of instruction in East and Southeast Asian universities. *The Asian Journal of Applied Linguistics*, 1(1), 4–15.
doi:10.1007/978-94-007-7972-3_2
- Kling, J., & Staehr, L. S. (2012). *The development of the Test of Oral English Proficiency for Academic Staff (TOEPAS)*. Copenhagen, Denmark: University of Copenhagen. Retrieved from https://cip.ku.dk/forskning/cip_publicationer/CIP_TOEPAS_Technical_Report.pdf
- Krashen, S. (1977). The Monitor Model for adult second language performance. In M. Burt, H. Dulay, & M. Finocchiaro (Eds.), *Viewpoints on English as a second language* (pp. 152-161). New York: Regents.
- Kraut, R., & Wulff, S. (2013). Foreign-accented speech perception ratings: A multifactorial case study. *Journal of Multilingual and Multicultural Development*, 34(3), 249–263. doi:10.1080/01434632.2013.767340
- Levis, J. M. (2006). Pronunciation and the assessment of spoken language. In R. Hughes (Ed.), *Spoken English, TESOL and applied linguistics: Challenges for theory and practice* (pp. 245-270). New York: Palgrave Macmillan.

- Ludwig, A., & Mora, J. C. (2017). Processing time and comprehensibility judgments in non-native listeners' perception of L2 speech. *Journal of Second Language Pronunciation*, 3(2), 167–198. doi:10.1075/jslp.3.2.01lud
- Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- 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. doi:10.1111/j.1467-1770.1995.tb00963.x
- Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49, 285–310. doi:10.1111/0023-8333.49.s1.8
- Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 23, 451–468. Retrieved from <https://www.cambridge.org/core>
- Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28, 111–131. doi:10.1017/S0272263106060049
- Murata, K., & Iino, M. (2017). EMI in higher education: An ELF perspective. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca* (pp. 400–412). Retrieved from <https://ebookcentral.proquest.com>
- Nasir, M. (2015, November 30). Bilingual curriculum to be compulsory in universities starting from 2016. *The Jakarta Post*. Retrieved from <https://www.thejakartapost.com/>

- Orikasa, M. (2016). The intelligibility of varieties of English in Japan. *World Englishes*, 35(3), 355–371. doi:10.1111/weng.12209
- Pallant, J. (2016). *SPSS survival manual* (6th ed.). Maidenhead, England: Open University Press.
- Pickering, L. (2006). Current research on intelligibility in English as a lingua franca. *Annual Review of Applied Linguistics*, 26, 219–233. doi:10.1017/S0267190506000110
- Révész, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, 35, 31–65. doi:10.1017/S0272263112000678
- Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57. doi:10.1093/applin/22.1.27
- Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, 78(2), 199–221. doi:10.2307/329010
- Saito, K., Tran, M., Suzukida, Y., Sun, H., Magne, V., & Ilkan, M. (2019). How do second language listeners perceive the comprehensibility of foreign-accented speech?: Roles of first language profiles, second language proficiency, age, experience, familiarity, and metacognition. *Studies in Second Language Acquisition*, 1–17. doi:10.1017/S0272263119000226
- Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37(2), 217–240. doi:10.1017/S0142716414000502

- Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, 38(4), 439–462. doi:10.1093/applin/amv047
- Sewell, A. (2010). Research methods and intelligibility studies. *World Englishes*, 29(2), 257–269. doi:10.1111/j.1467-971X.2010.01641.x
- Simbolon, N. E. (2016). *Lecturers' perspectives on English medium instruction (EMI) practice in Indonesian higher education* (Doctoral thesis, Curtin University). Retrieved from <https://pdfs.semanticscholar.org/f5ff/d40da46bca87298e380b59bc99b110f60b78.pdf>
- Simbolon, N. E. (2018). EMI in Indonesian higher education: Stakeholders' perspectives. *TEFLIN Journal*, 29, 108–128. doi:10.15639/teflinjournal.v29i1/108-128
- Smith, L. E., & Nelson, C. L. (1985). International intelligibility of English: Directions and resources. *World Englishes*, 3, 333–42. doi:10.1111/j.1467-971X.1985.tb00423.x
- Smith, L. E., & Nelson, C. L. (2006). World Englishes and issues of intelligibility. In B. B. Kachru, Y. Kachru, & C. L. Nelson (Eds.), *The handbook of World Englishes* (pp. 428-445). Malden, MA: Blackwell Publishing Ltd.
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, 15(4), 905–916. doi:10.1017/S1366728912000168
- van der Walt, C. (2000). The international comprehensibility of varieties of South African English. *World Englishes*, 19(2). doi:10.1111/1467-971X.00165

- Van Wijngaarden, S. J. (2001). Intelligibility of native and non-native Dutch speech. *Speech Communication*, 35(1–2), 103–113. doi:10.1016/S0167-6393(00)00098-4
- Van Wijngaarden, S. J., Steeneken, H. J., & Houtgast, T. (2002). Quantifying the intelligibility of speech in noise for non-native listeners. *The Journal of the Acoustical Society of America*, 111(4), 1906-1916. doi: 10.1121/1.1456928
- Vinke, A. A. (1995). *English as the medium of instruction in Dutch engineering education* (Doctoral thesis, Delft University of Technology). Retrieved from <http://resolver.tudelft.nl/uuid:491b55f9-fbf9-4650-a44d-acb9af8412a8>
- Vinke, A. A., Snippe, J., & Jochems, W. (1998). English-medium content courses in non-English higher education: A study of lecturer experiences and teaching behaviours. *Teaching in Higher Education*, 3(3), 383–394. doi:10.1080/1356215980030307
- VOICE Project. (2007). Mark-up conventions. *VOICE Transcription Conventions [Version 2.1]*, Retrieved from http://www.univie.ac.at/voice/documents/VOICE_mark-up_conventions_v2-1.pdf
- Whiteside, S. P. (1996). Temporal-based acoustic-phonetic patterns in read speech: some evidence for speaker sex differences. *Journal of the International Phonetic Association*, 26(1), 23–40. doi:10.1017/S0025100300005302

Appendix A. Speech elicitation: Theoretical and methodological considerations

This appendix overviews relevant theory considered in the creation of speech elicitation tasks for this research, details the target speech tasks from the TLU domain of an EMI classroom, and concludes with a brief definition of the target L2 speech construct for this study.

A.1 Theoretical considerations

Munro, Derwing, and Morton (2006, p. 113) observe that, “speech elicitation techniques and the nature of the listening task have important effects on the kinds of conclusions that can be drawn about [listeners’ understanding].” As such, it is necessary to explore the theoretical considerations of eliciting the kind of monologic speech representative of an EMI classroom before examination of any other factors impacting listening comprehension. Given this study’s targeting of context-specific language, the speech elicitation method should account for a variety of sociolinguistic-based factors potentially influencing language data, typically categorized into the following three types: setting, task and participant (Gass & Mackey, 2011; Mackey & Gass, 2005). The selected setting should allow for production of L2 speech with high *authenticity* (degree to which reflects language actually used in target context) while also allowing for control of topic and discourse complexity to minimize the effect of background knowledge on measurements of listeners’ understanding. For this scenario, the *semi-spontaneous* data collection setting of an interview provides the right balance between controlled and natural speech (Chaudron, 2003; Eisenbeiss, 2010).

In addressing task-related factors, it is helpful to view speech elicitation as a type of language testing, as in the context of this research both share the aim of aligning elicited L2 performance with target language usage (TLU) domain tasks. Measuring the extent to which the test or elicitation tasks accomplish said purpose, also termed *usefulness*, requires evidence of the following relevant task attributes: *reliability* (task variation accurately reflects TLU task differences), *construct validity* (interpretations based on task performance reflect actual language ability in TLU domain), *authenticity* (extent to which tasks reflect TLU

tasks), and *interactiveness* (extent to which tasks tap into participants' linguistic and topical knowledge) (Bachman & Palmer, 1996). Reliability and authenticity are demonstrated through specification and comparison of elicitation tasks with corresponding TLU domain tasks, while construct validity is strengthened through detailing the target language ability construct. Interactiveness, on the other hand, can be proven through examination of each task's characteristics in light of the aforementioned construct.

One other noteworthy task-related factor is the methodological choice in elicitation task design of allowing opportunity for preparation and/or note taking. Although research has shown the inclusion of either to have ramifications for validity, authenticity, and interactiveness (Chaudron, 2003 – planning influences fluency and complexity), EMI domain studies have found these practices to be common practice for lecturers (cf., Airey, 2011; Vinke, Snippe & Jochems, 1998) and as such could contribute to overall task usefulness.

Potential participant-related factors influencing elicited speech include status inequality between researcher and participant due to gender difference, role disparity (i.e., interviewer versus interviewee) (Mackey & Gass, 2005) and speaker L2 ability (Crookes, 1991). Suggested respective solutions include ensuring a balanced gender ratio amongst participants, using role play in speech tasks to reduce the tension of perceived status inequality in an interview setting (Kormos, 1999) (as cited in Chaudron, 2003), and specifying a minimum L2 proficiency test score as selection criteria to ensure adequate performance (Crookes, 1991).

A.2 Target language usage domain speech task characteristics

Although it was beyond the scope of this study to undertake extensive analysis of all possible speech tasks performed by L1 Indonesian EMI lecturers in their classrooms, several relevant ones have been identified from a brief perusal of pertinent literature, which was largely from the European context given the paucity of research detailing Asian EMI classroom practices. Vinke (1995) discovered that 'presenting new information' was the most common speech task performed by EMI lecturers in a Dutch university context, while Deroey and

Taverniers' (2011) corpus investigation of British lecturers' oral skills produced similar findings, with key production skill of 'informing' and related subskills of describing, interpreting, and demonstrating. A survey of TLU tasks for the development of a test of oral proficiency for EMI lecturers in Copenhagen University also identified the tasks of presenting material to students, and describing details of an assignment (Kling & Staehr, 2012).

A.3 Target L2 speech ability construct

This elicitation process aims to capture speech samples that are representative of L1 Indonesian lecturers' monologic L2 speech used in undergraduate EMI classrooms. Given that listeners may lack in-depth background knowledge of the speakers' various academic disciplines, tasks will target general rather than discipline-specific academic speech and focus on specific communicative functions common to EMI classrooms:

- provide reasons for class subject's importance and relevance to students' learning
- list key objectives of a class
- give instructions about an assignment

Speakers' L2 subskills in the domains of fluency, pronunciation, lexis and grammar will be engaged through the speech tasks, and also used as criteria in subsequent sample analysis.

Appendix B. Speaker information and consent form

Participant information sheet (*Lembar informasi peserta*)

Project Title (*Judul Proyek*): “Speaking to be understood: Indonesian academics’ English oral comprehensibility to Indonesian listeners”

Researcher (*Peneliti*): XXXX XXXX

I would like to invite you to take part in a research study examining how well Indonesian lecturers’ English speech can be understood by students from Indonesian language backgrounds.

Please take time to read the following information carefully before you decide whether or not you wish to take part, and if anything is unclear please ask me.

Saya ingin mengundang Anda untuk mengambil bagian dalam studi penelitian yang memeriksa seberapa baik pembicaraan bahasa Inggris dosen Indonesia dapat dipahami oleh mahasiswa dari latar belakang bahasa Indonesia.

Luangkan waktu untuk membaca informasi berikut dengan seksama sebelum Anda memutuskan apakah Anda ingin mengambil bagian atau tidak.

What is the study about? (*Tentang apa penelitian ini?*)

The aim of this study is to discern whether Indonesian lecturers’ standard proficiency test results relate to Indonesian listeners’ understanding of their English speech, and how other factors such as grammatical accuracy and listeners’ language background affect comprehension.

Tujuan dari penelitian ini adalah untuk mengetahui apakah hasil tes kecakapan standar dosen Indonesia terkait dengan pemahaman pendengar orang Indonesia tentang pidato bahasa Inggris mereka, dan bagaimana faktor-faktor lain seperti akurasi tata bahasa dan latar belakang bahasa pendengar memengaruhi pemahaman.

Why have I been invited? (*Mengapa saya diundang?*)

I have approached you as a lecturer who either has had experience in teaching your subject matter in English or may have opportunity to do so in the next several years. You also have taken either the official IELTS or TOEFL exam in the last 2 years.

Saya telah mendekati Anda sebagai dosen yang memiliki pengalaman dalam mengajar mata pelajaran Anda dalam bahasa Inggris atau mungkin memiliki kesempatan untuk melakukannya dalam beberapa tahun ke depan. Anda juga telah mengikuti ujian IELTS atau TOEFL resmi dalam 2 tahun terakhir.

What will I be asked to do if I take part? (Apa yang akan saya lakukan jika saya ambil bagian?)

If you decide to take part in the study, you will participate in a 20-30 minute recorded interview conducted in English. You will be asked to answer a few questions regarding background information and then provided with several speaking tasks related to teaching. The interview would be held in the university's Faculty of Science & Technology, on the 2nd floor of the administrative / classroom building, at a date and time that we agree on. Several samples of your recorded speech will then be extracted and played for groups of Indonesian university students who will answer comprehension questions based on the content and rate for ease of understanding. Your personal details (ie. age, name, role) will not be included in the recording.

Jika Anda memutuskan untuk mengambil bagian dalam studi ini, Anda akan berpartisipasi dalam wawancara direkam selama 20-30 menit yang dilakukan dalam bahasa Inggris. Anda akan diminta menjawab beberapa pertanyaan mengenai informasi latar belakang, lalu diberikan beberapa tugas berbicara berkaitan dengan mengajar. Wawancara akan diadakan di Fakultas Sains & Teknologi universitas, di lantai 2 gedung administrasi / ruang kelas, pada tanggal dan waktu yang kami sepakati. Beberapa sampel pidato Anda yang direkam akan diekstraksi dan dimainkan untuk kelompok mahasiswa Indonesia yang akan menjawab pertanyaan pemahaman berdasarkan konten dan tingkat kemudahan pemahaman. Detail pribadi Anda (mis. usia, nama, peran) tidak akan dimasukkan dalam rekaman.

What are the possible benefits from taking part? (Apa manfaat yang mungkin didapat dari ikut serta?)

If you take part in this study, insights gained from analysis of your speech and listeners' comprehension of it will contribute to our understanding of how best to equip lecturers to deliver their subjects in English to students from a variety of language backgrounds. We will also be able to better discern a suitable oral proficiency standard for those preparing to teach their subjects in English. You will also be provided access to the research report with opportunity to give feedback.

Jika Anda mengambil bagian dalam penelitian ini, wawasan yang diperoleh dari analisis pidato Anda dan pemahaman pendengarnya akan berkontribusi pada pemahaman kami tentang cara terbaik untuk membekali dosen untuk menyampaikan mata pelajaran mereka dalam bahasa Inggris kepada siswa dari berbagai latar belakang bahasa. Kami juga akan dapat membedakan dengan lebih baik standar kecakapan lisan yang cocok bagi mereka yang bersiap mengajar mata pelajaran mereka dalam bahasa Inggris. Anda juga akan diberikan akses ke laporan penelitian dengan kesempatan untuk memberikan umpan balik.

Do I have to take part? (Apakah saya harus ambil bagian?)

No. It's completely up to you to decide whether or not you take part. Your participation is voluntary and you are free to withdraw within the first two weeks after giving your consent, without giving any reason.

Tidak. Terserah Anda untuk memutuskan apakah Anda akan ambil bagian atau tidak. Partisipasi Anda bersifat sukarela dan Anda bebas untuk menarik dalam dua minggu pertama setelah memberikan persetujuan Anda, tanpa memberikan alasan apa pun.

What if I change my mind? (Bagaimana jika saya berubah pikiran?)

As explained above, you are free to withdraw within the first two weeks after giving your permission, and should you choose to do so, I will extract any data you contributed to the study and destroy it. Data means the information, views, ideas, etc. that you and other participants will have shared with me. Following the initial two week period, however, any data you have contributed to this study will not be able to be extracted.

Seperti dijelaskan di atas, Anda bebas untuk menarik dalam dua minggu pertama setelah memberikan izin Anda, dan jika Anda memilih untuk melakukannya, saya akan mengekstrak setiap data yang Anda berkontribusi untuk penelitian dan menghancurkannya. Data berarti informasi, pandangan, gagasan, dll. Yang akan Anda dan peserta lainnya bagikan dengan saya. Namun, setelah periode dua minggu awal, data apa pun yang telah Anda sumbangkan untuk studi ini tidak akan dapat diekstraksi.

What are the possible disadvantages and risks of taking part? (Apa kerugian dan risiko yang mungkin terjadi jika mengambil bagian?)

One disadvantage of participating in this study is that you will need to invest 20-30 minutes for an interview. Aside from this time cost, there are no other anticipated risks or disadvantages of taking part in this research.

Salah satu kelemahan berpartisipasi dalam penelitian ini adalah Anda perlu berinvestasi 20-30 menit untuk wawancara. Selain biaya waktu ini, tidak ada risiko atau kerugian yang diantisipasi untuk ikut serta dalam penelitian ini.

Will my data be identifiable? (Apakah data saya dapat diidentifikasi?)

After the data is collected, only I, the researcher conducting this study, will have access to it.

I will keep all personal information about you (e.g. your name and other information about you that can identify you) confidential, that is I will not share it with others. I will anonymise any audio recordings and hard copies of any data. This means that I will remove any personal information.

Setelah data dikumpulkan, hanya saya, peneliti yang melakukan penelitian ini, yang akan memiliki akses ke sana.

Saya akan merahasiakan semua informasi pribadi tentang Anda (mis. Nama Anda dan informasi lain tentang Anda yang dapat mengidentifikasi Anda), yaitu saya tidak akan membagikannya kepada orang lain. Saya akan menganonimkan rekaman

audio dan salinan data apa pun. Ini berarti bahwa saya akan menghapus informasi pribadi apa pun.

How will my data be stored? (*Bagaimana data saya disimpan?*)

Your data will be stored in encrypted files (that is no-one other than me, the researcher, will be able to access them) and on a password-protected computer.

I will keep data that can identify you separately from non-personal information (e.g. your views on a specific topic).

In accordance with University guidelines, I will keep the data securely for five years, after which it will be destroyed.

Data Anda akan disimpan dalam file terenkripsi (yang tidak lain adalah saya, peneliti, akan dapat mengaksesnya) dan di komputer yang dilindungi kata sandi.

Saya akan menyimpan data yang dapat mengidentifikasi Anda secara terpisah dari informasi non-pribadi (mis. Pandangan Anda tentang topik tertentu).

Sesuai dengan pedoman Universitas, saya akan menyimpan data dengan aman selama lima tahun, setelah itu akan dihancurkan.

How will we use the information you have shared with us and what will happen to the results of the research study? (*Bagaimana kami akan menggunakan informasi yang telah Anda bagikan dengan kami dan apa yang akan terjadi pada hasil studi penelitian?*)

I will use the data you have shared with me in the following ways:

- I will use it for academic purposes only, in order to create a dissertation paper as part of my MA program.
- Several samples of your recorded speech will be played for groups of Indonesian university students who will answer comprehension questions based on the content and rate for ease of understanding. Your identity as the speaker in the recordings will be kept anonymous, but there may be a small chance that some student listeners might recognize your identity based on your voice.

Saya akan menggunakan data yang telah Anda bagikan dengan saya dengan cara berikut:

- *Saya akan menggunakannya untuk tujuan akademik saja, untuk membuat makalah disertasi sebagai bagian dari program MA saya.*
- *Beberapa sampel pidato Anda yang direkam akan dimainkan untuk kelompok mahasiswa Indonesia yang akan menjawab pertanyaan pemahaman berdasarkan konten dan menilai untuk memudahkan pemahaman. Identitas Anda sebagai pembicara dalam rekaman akan tetap anonim, tetapi mungkin ada kemungkinan kecil bahwa beberapa pendengar siswa dapat mengenali identitas Anda berdasarkan suara Anda.*

Who has reviewed the project? (*Siapa yang telah meninjau proyek?*)

This study has been reviewed and approved by my supervisor in line with the Faculty of Arts and Social Sciences and Lancaster Management School's Research Ethics regulations.

Studi ini telah ditinjau dan disetujui oleh penyelia saya sesuai dengan Fakultas Seni dan Ilmu Sosial dan peraturan Etika Penelitian Sekolah Manajemen Lancaster.

What if I have a question or concern? (Bagaimana jika saya memiliki pertanyaan atau masalah?)

If you have any queries or if you are unhappy with anything that happens concerning your participation in the study, please contact myself using the following information:

Jika Anda memiliki pertanyaan atau jika Anda tidak puas dengan apa pun yang terjadi sehubungan dengan partisipasi Anda dalam penelitian ini, silakan hubungi diri saya menggunakan informasi berikut:

Douglas David Evans
Cell/Mobile: ///////////////
Email: //////////////@lancaster.ac.uk

If you have any concerns or complaints that you wish to discuss with a person who is not directly involved in the research, you can also contact:

Jika Anda memiliki masalah atau keluhan yang ingin Anda diskusikan dengan orang yang tidak terlibat langsung dalam penelitian, Anda juga dapat menghubungi:

Dr. Luke Harding
Phone: ///////////////
Email: //////////////@lancaster.ac.uk

**Thank you for considering your participation in this project.
(Terima kasih telah mempertimbangkan partisipasi Anda dalam proyek ini.)**

CONSENT FORM
(Formulir Persetujuan)



Project Title (Judul Proyek): "Speaking to be understood: Indonesian academics' English oral comprehensibility to Indonesian listeners"

Name of Researcher(s) (Nama Peneliti): Douglas David Evans

Email: ■■■■■@lancaster.ac.uk

Please tick each box (Silahkan centang setiap kotak)

1. I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.
(Saya mengkonfirmasi bahwa saya telah membaca dan memahami lembar informasi untuk studi di atas. Saya memiliki kesempatan untuk mempertimbangkan informasi, mengajukan pertanyaan yang telah dijawab dengan memuaskan.)

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. If I withdraw within 2 weeks of commencement of the study my data will be removed. If I withdraw after 2 weeks my data may remain part of the study.
(Saya memahami bahwa partisipasi saya bersifat sukarela dan bahwa saya bebas untuk mundurkan diri kapan saja, tanpa memberikan alasan apa pun. Jika saya mundurkan diri dalam waktu 2 minggu sejak studi dimulai data saya akan dihapus. Jika saya mundurkan diri setelah 2 minggu, data saya dapat tetap menjadi bagian dari penelitian.)

3. I understand that any information given by me may be used in an MA program research report by the researcher, but my personal information will not be included and I will not be identifiable.
(Saya mengerti bahwa informasi apa pun yang diberikan oleh saya dapat digunakan dalam laporan penelitian program MA oleh peneliti, tetapi informasi pribadi saya tidak akan dimasukkan dan saya tidak akan dapat diidentifikasi.)

4. I understand that my name/my organisation's name will not appear in this MA research report without my consent.
(Saya mengerti bahwa nama saya / nama organisasi saya tidak akan muncul dalam laporan penelitian MA ini tanpa persetujuan saya.)

5. I consent to be audio-recorded during the interview.
(Saya setuju untuk direkam audio selama wawancara.)

6. I understand that any interviews will be audio-recorded and transcribed and that data will be protected on an encrypted device and kept secure.
(Saya mengerti bahwa setiap wawancara akan direkam audio dan ditranskrip, dan bahwa data akan dilindungi pada perangkat terenkripsi dan tetap aman.)

7. I understand that data will be kept according to University guidelines for a minimum of five years after the end of the study.
(Saya mengerti bahwa data akan disimpan sesuai dengan pedoman Universitas untuk minimum lima tahun setelah akhir penelitian.)

8. I agree to take part in the above study.
(Saya setuju berpartisipasi dalam studi tersebut.)

Name of Participant
(Nama Peserta)

Date
(Tanggal)

Signature
(Tanda Tangan)

I confirm that the participant was given an opportunity to ask questions about the study, and all the questions asked by the participant have been answered correctly and to the best of my ability. I confirm that the individual has not been coerced into giving consent, and the consent has been given freely and voluntarily.

Signature of Researcher /person taking the consent _____ **Date** _____ **Day/month/year**

One copy of this form will be given to the participant and the original kept in the files of the researcher at Lancaster University

☒

☒

Appendix C. Speaker interview: Background questions and elicitation prompts

A. Background information

The following questions focus on your personal information, and are not a test of your language ability.

1. What year were you born?
2. What is your ethnicity?
3. Where did you grow up in Indonesia?
4. What is your last completed level of education?
5. What faculty do you work at?
6. What subjects do you normally teach?
7. Do you have any experience in teaching your subject matter using English?
8. What is your most recent IELTS/TOEFL score? When did you achieve it?
9. Have you ever spent time living abroad in an English-speaking region? If yes, where, when, and for how long?

B. Semi-spontaneous speaking prompts

You will be given several speaking tasks related to your role as a lecturer. For each, you will have a few minutes of preparation to plan what you are going to say before you begin.

Practice prompt: Think about a course you taught that you feel was very successful. What were the reasons for its success? Take 1-2 minutes to plan your talk, including 4 reasons why the class was successful, and then take 2 minutes to talk about the reasons for your course going so well.

Prompt 1: Imagine you are giving a brief promotion for your favourite subject / class to new students who have no background knowledge in the subject. Take 1-2 minutes to plan your talk, which should include 4 key features to promote this subject, and then talk for 2 minutes about your class.

Prompt 2: Imagine you are listing 4 key course learning objectives or outcomes for a class of new students. Take 1-2 minutes to plan your talk, and then talk for 2 minutes about the course objectives. Remember to explain any words that might be unfamiliar to your listeners.

Prompt 3: Imagine you are giving instructions on a class project / final task to a group of students. Take 1-2 minutes to plan what you want to say, including:

- a. What the task is
- b. Any resources the students will need to complete the task
- c. Whether students must work on it individually or in groups
- d. When it is due

Take 2 minutes to give these instructions.

Appendix D. Speech task transcripts

The speech task response transcripts for speakers F04, F07, F10, and M06 were included for reference and illustrative purposes. See Appendix C for speech task/prompt descriptions.

Transcripts were simplified through removal of hesitation markers (e.g., um, uh) and most phonological feature indicators (e.g., pauses, lengthened sounds), although basic punctuation was retained for ease of reading.

Legend⁶

()	Words, fragments, or phrases that cannot be reliably identified
-	Repetition (e.g., t- t- today) or word fragment (e.g., partic-)
[]	Added comments or clarifications by researcher
<L#xx> </L#xx>	Non-English utterances, either in speaker's native tongue (i.e., L1) or in other language. The specific language two letter abbreviation is included (i.e., id = Indonesian; ar = Arabic)
{ }	Translated word or phrases

Speaker F04

Prompt 1: 3 minutes, 12 seconds

So, in our clinical psychology we want to aim- we want to learn together between you and me in our contrary time now. Why we s- why we should learn about clinical psychology as the (duty) in our faculty because of from the (firstly) from the main reason is there many disorder growth, not only

⁶ Based on Vienna Oxford International Corpus of English Version 2.1 transcription conventions (VOICE project, 2007)

that- that state in DSM 5 or 4 – you know about DSM 5, DSM 5 or 4 is diagnostic system- diagnostic system- of diagnostic system manual that created by APA. APA is American Psychology Association. Now we still have to DSM 4, because the DSM 5 is not declared by all psychologists in the world.

The second one is because psychologists is become rare in our society, because of it's not a habit or because the psychology is close with the disorder or with the mental sickness people, so the therapies that given by psychologists is very expensive, because it's take too long time, and it's only give one by one. But in our globalizations demand, the psychology- the clinical psychology is not only give one by one but it must change to macro clinical psychology. So we must have a contribution to our society, to our nation. For the example, when we have elections now, clinical psychology can give some treatment to the legislative candidate who failed. So the- the (value) to go to mental illness is minimized.

The third one is because of mental illness is culturally bound, so is- so ex- so it's so interesting because of our Indonesian city- our Indonesian (nations) is consist of many various ethnicity. For the example, for the Javanese people, mental illness in (chance) is pro- prohibited but in some regents is prohibited to give chance for mental illness people.

And the last important- the last but not least, because of Indonesia is consists of many islands, so there are many remote area that didn't psychologists, so it's very challenges when the psy- the clinical psychologists can give therapy from the remote area, like- like give the therapies by intelligence method by website.

Prompt 2: 2 minutes, 9 seconds

This is four key course learning objectives or outcomes for our abnormal psychology class. The first one is describe the pathological symptoms, because the pathological symptoms is a raw material to- to master, sorry, a

raw material to mastery in our class. For the example, is the students who- who come to our class must mastery about the (differences) between distress or (au-stress). Is it ok to categorize in stress or not?

The second one is analyze the psychopathological group of disorders because there are individual differences – it means that every person have a different background of life, and have different backgrounds of the disorder learning cycle. So, in the third of our outcomes class, you can diagnose the certain disorder based on the analysis of the psychopathological (role). For the example is some study case can determine the symptoms but you can confuse between anorexia, nervosa or bulimia, or (binge) eating. So, before we have or (gain) to the third objective you must mastery first and the second one.

And the finally, you must arrange the predictions of appropriate psychological therapy based on the disorders diagnosing, because not all persons- not every persons can get every- not can get general therapy. So, it's like a tailor-made that therapy just appropriate with a certain person based on the life background and life growth of pathological symptoms.

Prompt 3: 3 minutes, 27 seconds

The final task or our main class project is due to one month again, on- I mean that one month we have midterm, because if we will finish the theoretical framework in- until midterm, and in the nin- ninth meeting we must arrange the final task that is psychoeducations. The task is- the title of task is health promotion about mental health awareness. So, it will submit due to September, because our mental health awareness month is on October. So, from the first- at the first week we must go to the society and collect the data assessment from society, and you must choose one setting. Because of clinical psychology not in micro perspective, but it's must combine in macro perspective, it means that you can choose for the example in clinical psychology which use educational background as your setting. So, it's about

mental health awareness in school wellbeing or others as your little (title) or your themes.

So you must interview the teenager or students, interview the teachers and interview the parents. And you must find the topics that appropriate with that educational setting, for the example, in the school. And after collecting the data assessment, you must go to the certain problem statement. And the second one is you can combine the second steps is combine that problem statement with your handbook or your meta-analysis of journals, what kind of- what kind of perspective that will be used in given- in giving the psychoeducations. You can use the handbook of- is- must use handbook, or it's the- handbook <L1id> yang wajib dibaca {that must be read} </L1id> that must be read is from professor [person's name]. The title is Macro Clinical Psychology, and the second handbook is applied psychology too.

And the second steps is y- the second steps is you make the certain group that consists of three person who find the same data assessment. So when collect the data assessment by interview uh individually, read and giv- read and do the meta-analysis of journal is individually and but in the second step you must join to the other three person who have- who find the similar data from the society, and make a group discussions, and in your group you must arrange the model of psychoeducation.

Speaker F07

Prompt 1: 1 minute, 58 seconds

Hi students. I think in choosings your subjects material or subjects lessons you want to think about four reasons. I want to say that's maybe you must to choose my subjects which is name is Database. Because the first is in our life, many kinds of information systems or applications to have, and all those

information systems or applications use database to store their informations. So the informations can manage by database. That's the first reasons.

The second reasons, when you has able in database so you can also learn about other subjects that the foundations is database. The other subjects such as data warehouse, data mining, business intelligence, and etcetera.

And the sec- , the third, my subjects lessons is not only theory but also practice. So I guess that you will be more funs when your subjects can you see directly and can you practice it.

And the last maybe if you interest to learn more about the database you also can make some course and get certificates from the institutions of database. Ok, I think there's four reasons.

Prompt 2: 1 minute, 58 seconds

Hi students. There are four objective or outcomes what you have if you join my class. There is subject is a database. The first, the outcome is you will understands why the database is importants and why the database is used in every is- information system or applications. That's the theory. But you must understand why you use it.

The second is you can modeling your data. So our data has related each other. And the rela- the relations is a model. The relations is representings by a model so you must to modeling your data. That's the second objective.

The third is you can create your data modelings physically to the database. So your theory or your model that you designs you will create in database.

And the last is you can operate the (datarbase)- the database. Not only create, but you can also insert the data. Insert is activity you will store data to database. And then you can update, it means you can edit the data. The

third you can delete or you must delete your data if something don't work. And the second is you can select the data. Select, what will you show in your report and what do not will you show in your report. So there are four or five operations in database you wa- must capable to do it, it for the objective of my lessons.

Prompt 3: 1 minute, 38 seconds

Ok students, we have final test in our subjects. The final test is you must design a database in an institutions or organizations. For this task, you can get some of resource of data, such as interview or from the explanations from their websites.

The final test must done by three groups of persons, and I will wait it in the end of our class, especially in June, 15 of Junes maybe after the <L2ar> Eid-ul-Fitri [religious holiday] </L2ar>, and I will explain you somes about this final test, some explanations.

The first is the institutions you choose, you will ask if they has store their data manually or digitially. If they have digitially, so you must choose other institu- institutions. And seconds, you must model the data, and design the database. And the last is you must create the report from the institutions. Example is the report of transactions in one month, the- the last one month, or the last one years. So that's about the final test. I hope you understand and do it with your best.

Speaker F10

Prompt 1: 1 minute, 58 seconds

Today we will ta- we will learn about the production system. This course really important to our department and to our course because this is the main course of industrial engineering. You know that production system

teach us about how to see the production as a system. So you can see the relate- the relation between all element in production. So that's why the production system is really important.

And then you know that nowadays business very- develop very- very fast. So we- we need to- we need to create or develop the production system that can, you know, win the- the situation.

And then by learning this course, you will face different problem in production system with different approaches, so you will learn about how to make the best decision between the alternative, very- a lot of alternative. And then if you finish this course very well, it will make you easy to, to find job, because your idea, your point of view of system, especially in production system, will be different. That will- that will support you in make decision in the real industrial world.

Prompt 2: 1 minute, 56 seconds

Ok, I would like to explain about what is the objective or outcome for our subject. So the first thing that you will, you'll get from this study, you will be very, you will- you will have ability to define what is the most popular product now in the market.

And then after you know what is the- the product that become the most favourite, and then you can make the forecast about the number of the product, how many product that we will be produce by the factory. And then based on the forecast result, you will make a plan for the production about how many- how many products that will be produced, and when the process exactly executed to the shop floor.

And then the- the last thing that you- you will get from this sh- course, you will be- you will have ability to make sure the capacity or the resources that you have in your factory. And then you will make the decision about what-

what- what you will do next if the capacity overload or below the capacity. So that's the four objective outcomes, or outcomes from this <L1id> (apa) {what} </L1id> subject.

Prompt 3: 1 minute, 26 seconds

Ok class. I will give you s- a task to do in your group. So the task is you need to make a forecast for the product. I will give you what is the product, the product. So you need to make the forecast with- and then you have to choose what is the- the- the method that can be used to- to make the forecast related to the- the- the- the- the data. So the- you- you need to complete this task. You- you need to prepare, of course, the data itself. And then you need the- to- you need to use the software KM, and then I think because this is a really big problem to do, so you have to finish it in your groups and I give you a week to finish this product. And the deadline, of course, the next Tuesday. So that's the task for this course of forecasting.

Speaker M06

Prompt 1: 2 minutes, 15 seconds

Good morning students, ok. Today I will introduce you about electromagnetic engineering. This is a special subject for electrical engineering where you interested in telecommunication. The- the reason why you need to learn about this subject is, first this is the basic skill- the basic skill of telecommunication, special for radio- radio engineering. If you want to do anything in radio, you must know how the electromagnetic should be- can be- can be describe. And the sec- the second reason is, most of radio models presented this skill. I mean need this skill. So, without this subject you cannot understand what they talking or what the model can be explain and how to analyze the model for the next- for the next steps.

And the third reason is, with this subject you can understand the mathematic skill of the radio- of the radio models. So many research or engineer talk about radio but they don't talk about the radio model in mathematics. So, this is a problem. They cannot understand what's going on in the- the model. S- and the last thing why you should learn this subject is there are so many research available according to this subject. Specially in <L1id> medis- {medicine} </L1id> (.) medis-electromagnetic, like biolo- biomedic and maybe for, k maybe for social activity, another social activity and many things. Just like my next research should be in biomedic I think, and this need electromagnetic engineering skill of course.

Prompt 2: 2 minutes, 30 seconds

Ok good day students. For the discourse we- we should be able to learn, I mean the outcomes- the outcomes from this course is should be, first you can understand any questions and can solve any question according to coordinate system, just like a (Cartesian) coordinate and then ball coordinate and (cylindris) coordinate.

And three things is the basic coordinate system you sh- you must know. So, by this course you can master this, all of the question how can I make any (trap) on the questions you- you should be know after that.

And the second ones, after the course- after the course- k, after the course you can able to understand how the metric [matrix] calculation can be done. So, this is a mathematics skill you need to expands and expands more for telecommunication engineering.

And the third things is should be, factor analysis. After you know about the metric [matrix], you know about the coordinate system, then you must know also about (fracture), how the (fracture) can men- can- can be mentioned in coordinate system how they can make range between two points and how to calculate the absolute and other things.

And the last things by the course, by this- this course, you should be able to understand how the Ohm law can be applied. You know the Ohm law. Ohm law should be the basic- the basic- the basic skill- you- the basic knowledge you must know in electrical engineering. You can explain to all the people how they, how can- how they can calculate the intensity electromagnetic fields and other things. So I hope by follow this course you can master four key course learning objective. Thank you.

Prompt 3: 2 minutes, 6 seconds

Ok, good morning class. Today we- I want to- I want you to make a some homework, but this is should be like final task for our course. This homeworks- the title this homework is should be literature review. So you need to- to have a rev- to have a literature, it means a journal, technical journal, by (term) radio telecommunications. You can find this resource from anything maybe by using Google, by using (Saihab) or another special web engine you can do. Just find an English journal related to radio telecommunication. And then, you should be analyze review what's they do- what's they have complete for they task. And then you should say what they weaks also, their weak also. Sorry.

This task should be done in five groups of student. I mean the member of groups should be five students. So we have forty in this class, so we have eight groups should be done in May.

This report should be available on my desk next weeks, next weeks. Not more than this date. So, I believe you can work together and can find the research by using internet, you have smart phone, you have computer, you have any things. You can use any things to complete this tasks by five weeks, five of your friends. Thank you that's all I think. If there is any question, please ask me later.

Appendix E. Listener information and consent form

Participant Info

Welcome to the research study!

Project Title: “Speaking to be understood: Indonesian academics’ English oral comprehensibility to Indonesian listeners”

Researcher: Douglas David Evans

You are invited to participate in a web-based questionnaire on understanding of Indonesian academics’ English speech by non-native English speakers. This is a research project being conducted by Douglas David Evans, a student at Lancaster University, UK. It should take approximately two 45-60 minute sessions (ie. 1.5-2 hours total) to complete.

PARTICIPATION

Your participation in this survey is voluntary. You may refuse to take part in the research or exit the survey at any time without penalty. You are free to decline to answer any particular question you do not wish to answer for any reason.

BENEFITS

You will receive a certificate of appreciation for participating in this research study. In addition, your responses may help us learn more about factors that influence how well academic non-native English speech is understood.

RISKS

There are no foreseeable risks involved in participating in this study.

CONFIDENTIALITY

Your survey answers will be sent to a link at Lancaster university where data will be stored in a password protected electronic format. This survey does not collect identifying information such as your name, email address, or IP address. Therefore, your responses will remain anonymous. No one will be able to identify you or your answers, and no one will know whether or not you participated in the study.

CONTACT

If you have questions at any time about the study or the procedures, you may contact myself using the following information:

Douglas David Evans

Cell/Mobile ■■■■■■■■■■

Email: ■■■■■■■■■■

If you have any concerns or complaints that you wish to discuss with my research supervisor, you may contact him using the following information:

Dr. Luke Harding

Phone: ■■■■■■■■■■

Email: ■■■■■■■■■■

Informed Consent

ELECTRONIC CONSENT: Please select your choice below. Clicking the “I consent” button indicates that:

- You have read the provided information about the research study
- You voluntarily agree to participate
- You are 18 years of age or older

I consent, begin the study

I do not consent, I do not wish to participate

Appendix F. Listener survey: Comprehensibility and comprehension instruments and background information questions

Note that only several instrument samples are presented here due to their uniformity.

There are three sections to this questionnaire. In Section 1, you will listen to brief recordings and rate for how easily you could understand the speech. In Section 2, you will listen to more recorded speech samples, and answer questions regarding the content. This is a longer section, and you may need to take a break so as to be able to give your full attention to these tasks. Finally in Section 3, you will answer questions about your background information. Let us begin!

S1 - Speech Comprehensibility - Instructions & Example

Section 1 - Comprehensibility of speech samples

Instructions:

1. Please find a quiet place as you will need to listen carefully to some speech recordings.
2. If you have earphones / headphones available, it would be good to use them for this section.
3. You will listen to a total of 20 recorded speech samples, each between 25-39 seconds long.
4. After each recording, rate how easy or difficult it was to understand by selecting an appropriate value on the provided scale.
5. You will first listen to an example recording to practice rating your ease of understanding.

Example:

You will hear a lecturer explaining why his/her subject is important for students to learn. Click on the button to listen to the recording, and then answer the question below.

(Note: Please wait, as the recording may take a few moments to load. It will only play once.)

Play Example Recording

How easy or difficulty was it to understand the speaker's English?

(Note: This question is for practice only.)

Very easy to understand 1 2 3 4 5 6 7 8 Very difficult to understand 9

Transition

Now that you have completed the example, we will move on to rating difficulty / ease in understanding for each of the 20 speakers, whom are presented in a randomized order.

Comprehensibility - F1

Speaker F1:

You will hear a lecturer explaining why his/her subject is important for students to learn. Click on the button to listen to the recording, and then answer the question below.

(Note: Please wait, as the recording may take a few moments to load. It will only play once.)

Play Recording

How easy or difficulty was it to understand the speaker's English?

Very easy to understand 2 3 4 5 6 7 8 Very difficult to understand

Comprehensibility - M5

Speaker M5:

You will hear a lecturer explaining why his/her subject is important for students to learn. Click on the button to listen to the recording, and then answer the question below.

(Note: Please wait, as the recording may take a few moments to load. It will only play once.)

Play Recording

How easy or difficulty was it to understand the speaker's English?

Very easy to understand 1	2	3	4	5	6	7	8	Very difficult to understand 9
------------------------------------	---	---	---	---	---	---	---	---

Comprehensibility - M6

Speaker M6:

You will hear a lecturer explaining why his/her subject is important for students to learn. Click on the button to listen to the recording, and then answer the question below.

(Note: Please wait, as the recording may take a few moments to load. It will only play once.)

Play Recording

How easy or difficulty was it to understand the speaker's English?

Very easy to understand 1 2 3 4 5 6 7 8 Very difficult to understand 9

Comprehensibility - M7

Speaker M7:

You will hear a lecturer explaining why his/her subject is important for students to learn. Click on the button to listen to the recording, and then answer the question below.

(Note: Please wait, as the recording may take a few moments to load. It will only play once.)

Play Recording

How easy or difficulty was it to understand the speaker's English?

Very easy to understand 1 2 3 4 5 6 7 8 Very difficult to understand 9

S2 - Speech Comprehension - Instructions & Examples

Section 2 - Comprehension of speech samples

Instructions:

1. You will listen to 20 lecturers talking about information and tasks related to their courses that they teach. There are two samples for each speaker, for a

total of 40 recordings, each between 59-208 seconds long.

2. While listening to each recording, type in English the main points that you hear in the speech.
3. Because there are a lot of recordings to listen to, you are encouraged to take at least one break during this section. Your survey progress is saved online for up to one week, and you can access it again by clicking on the link provided.
4. Several examples are given below for you to practice on.

Example 1:

In this recording, you will hear the speaker giving four reasons why the course they taught was successful. After you listen, write the four reasons for the course's success in the text box below.

Click on the button to listen to the recording.

(Note: Please wait, as the recording may take a few moments to load. It will only play once.)

Play Example Recording

In this example, the four reasons for the course's success have been completed for you in the text box below.

Please write the four reasons for the course's success in the text box below.

(Note: You DO NOT have to write in full sentences.)

1. can deliver content so that students understand well
2. students can implement content in real applications
3. changed students' attitudes towards discipline, working hard, presentation skills, etc.
4. students' marks were higher

Example 2:

In this recording, you will hear the speaker giving four reasons as to why the course they taught was successful. After you listen, write the four reasons for the course's success in the text box below.

Click on the button to listen to the recording.

(Note: Please wait, as the recording may take a few moments to load. It will only play once.)

Play Example Recording

Please write the four reasons for the course's success in the text box below.

(Note: You DO NOT have to write in full sentences.)

Click "Next" to see some possible answers for this question.

Possible answers for Example 2:

1. Good results (70-90; average 80%); due to good student motivation
2. Interactive class; students allowed to ask questions anytime; students pushed to ask questions
3. Gave videos about physics
4. Gave students experiments about theory taught

Transition to Comprehension

Now that you have had some practice, we will begin listening to samples of lecturers' classroom speech and answering related comprehension questions. Don't forget to take a break if you are feeling tired with these tasks.

Comprehension - F1

Speaker F1:

Speech Sample 1

In the following recording, you will hear a speaker giving four learning objectives / outcomes for a course they teach. After you listen, write the four learning objectives / outcomes mentioned in the text box below.

(Note: Please wait, as the recording may take a few moments to load. It will only play once.)

Play Recording

Please write the four learning objectives / outcomes mentioned in the text box below.

(Note: You DO NOT have to write in full sentences.)

Speaker F1:

Speech Sample 2

In the following recording, you will hear a speaker giving instructions for a course project or final task. After you listen, write the following key details for the course project instructions in the text box below:

1. What the task is
2. What resources are needed to do the task
3. Whether to work individually or in groups
4. When the task is due (eg. task deadline)

(Note: Please wait, as the recording may take a few moments to load. It will only play once.)

Play Recording

Please write the following key details for the course project instructions in the text box below:

1. What the task is
2. What resources are needed to do the task
3. Whether to work individually or in groups
4. When the task is due (eg. task deadline)

(Note: You DO NOT have to write in full sentences.)

Speaker M7:

Speech Sample 1

In the following recording, you will hear a speaker giving four learning objectives / outcomes for a course they teach. After you listen, write the four learning objectives / outcomes mentioned in the text box below.

(Note: Please wait, as the recording may take a few moments to load. It will only play once.)

Play Recording

Please write the four learning objectives / outcomes mentioned in the text box below.

(Note: You DO NOT have to write in full sentences.)

Speaker M7:

Speech Sample 2

In the following recording, you will hear a speaker giving instructions for a course project or final task. After you listen, write the following key details for the course project instructions in the text box below:

1. What the task is
2. What resources are needed to do the task
3. Whether to work individually or in groups

4. When the task is due (eg. task deadline)

(Note: Please wait, as the recording may take a few moments to load. It will only play once.)

Play Recording

Please write the following key details for the course project instructions in the text box below:

1. What the task is
2. What resources are needed to do the task
3. Whether to work individually or in groups
4. When the task is due (eg. task deadline)

(Note: You DO NOT have to write in full sentences.)

S3 - Participant Background Information

Section 3 - Background information

Please answer the following questions.

What is your age?

What is your sex?

Male

Female

What is your nationality?

Indonesian

Thai

Other (Please specify)

What is the highest level of education you have completed?

High school or earlier

Some college but no degree

Associate degree in college (2-year)

Bachelor's degree in college (4-year)

Master's degree

Doctoral degree

What is your current occupational status?

(Select all that apply)

Student

Employed

Retired

None of the above

What is your current semester of study?

What is your study program major or discipline?

What is your first language?

Indonesian

Thai

Other (Please Specify)

How would you rate your English language ability?

Beginner

- I can understand simple common phrases.
- I can talk a little about myself.
- I can ask and answer simple personal questions.
- I can communicate a little with help.

Elementary

- I can understand sentences.
- I can exchange some familiar information.
- I can describe my life a little.
- I can describe what I can see near me.

Intermediate

- I can understand the important points when I listen to something about everyday life.
- I am usually OK if I travel somewhere people speak English.
- I can talk about my interests.
- I can describe what has happened.
- I can give my basic opinions.

Upper intermediate

- I can understand the important points when I read something complex.
- I can talk with first language English speakers with flow and without discomfort.
- I can explain a clear, detailed argument.

Advanced

- I can understand long sentences and implicit meaning.
- I can explain my ideas with flow and without much hesitation while I think.
- I can adapt my English for effective social or professional interaction.
- I can explain complex topics with well-structured, connected language.

Master

- I can understand almost everything easily.
- I can reconstruct information from different sources coherently.
- I can find slight differences in meaning in complex language.
- I can use English precisely, fluently and spontaneously.

If you have taken one of the following English proficiency exams, please note the year when you last took it and the score that you got.

	Year	Score
IELTS	<input type="text"/>	<input type="text"/>
TOEFL	<input type="text"/>	<input type="text"/>
Other (Please specify) <input type="text"/>	<input type="text"/>	<input type="text"/>

Survey Feedback:

I would appreciate any feedback you have regarding this survey (eg. clarity of instructions, any issues with listening to audio recordings, difficulty of comprehension questions, etc). Please write your feedback in the textbox below .

Powered by Qualtrics



Appendix G. Comprehension question answer key

Comprehension question answer keys for speakers F04, F07, F10, and M06 have been included here for reference and illustrative purposes.

Comprehension Question Answer Key

- Comp1 – Comprehension question 1, based on Prompt 1
- Comp2 – Comprehension question 2, based on Prompt 2
- _X## - speaker code (e.g., _F01 = Female Speaker 1)
- **Bold Text** – word(s) (or equivalent meaning) required in answer
- [] – acceptable alternative/additional words not mentioned in recording

Marking Procedure:

- 1 mark for each key point adequately identified
- Half marks (0.5) awarded where some but not all key words present, indicating that answer is on the right track
- Judgment can be used to award half or full marks if synonyms are used in place of key words identified from recording
- Spelling or grammar errors are not penalized

Question	Answer Key
Comp1_F04	<ol style="list-style-type: none">1. describe the pathological symptoms2. analyze the psychopathological group of disorders3. can diagnose the certain disorder based on analysis of psychopathological [rule]4. arrange predictions of appropriate psychological therapy based on diagnosis<ul style="list-style-type: none">- tailor-made therapy- based on person's background and pathological symptoms

Question	Answer Key
Comp2_F04	<ol style="list-style-type: none"> 1. Health promotion about mental health awareness <ul style="list-style-type: none"> - choose social setting for research and educational setting to deliver health promotion - collect data through interviews (of teenagers, teachers, parents) - create problem statements 2. Resources: handbook “Macro Clinical Psychology”; second handbook in applied psychology; meta-analysis of journals 3. Group of 3 people for presentation <ul style="list-style-type: none"> - collect data, read books and do journal meta-analysis individually 4. Due one month after midterm / theory <ul style="list-style-type: none"> - Submit in September, because October is mental health month
Comp1_F07	<ol style="list-style-type: none"> 1. understand theory/importance of databases <ul style="list-style-type: none"> - why databases are important, why used in every information system / application 2. can model data 3. can create physical database from data models 4. can operate database <ul style="list-style-type: none"> - insert/store, update, edit, select data
Comp2_F07	<ol style="list-style-type: none"> 1. design a database for an institution/organization <ul style="list-style-type: none"> - select institution without digital database - model data and design database - create report from institution’s transactions 2. resources: interviews, explanations from websites 3. done by groups [of 3 persons] 4. 15 June; after <i>Idul Fitri</i>; end of class

Question	Answer Key
Comp1_F10	<ol style="list-style-type: none"> 1. able to define most popular product in the market 2. able to make forecast of production <ul style="list-style-type: none"> - how many product that will be produce[d] [in] the factory 3. able to make production plan 4. able to ensure [sufficient] resources in production process; make decision about what to do if capacity overload or low capacity
Comp2_F10	<ol style="list-style-type: none"> 1. Task: make forecast for product <ul style="list-style-type: none"> - Teacher specifies product - choose method to make forecast 2. Resources: data, KM software 3. Group [task] 4. 1 week; due next Tuesday
Comp1_M06	<ol style="list-style-type: none"> 1. understand any questions according/[related] to coordinate systems 2. understand how matrix calculations can be done 3. how to do factor analysis 4. able to understand how Ohm's law can be applied
Comp2_M06	<ol style="list-style-type: none"> 1. Task: literature review <ul style="list-style-type: none"> - analyze review - describe articles' weaknesses 2. Resources: [online] journals 3. Groups of 5 students 4. Due: next week

Appendix H. EFL teacher information and consent form

Participant Info

Welcome to the research study!

Project Title: “Speaking to be understood: Indonesian academics’ English oral comprehensibility to Indonesian listeners”

Researcher: Douglas David Evans

You are invited to participate in a web-based analysis of Indonesian academics’ English speech. This is a research project being conducted by Douglas David Evans, a student at Lancaster University, UK. It should take approximately 80-90 minutes to complete.

PARTICIPATION

Your participation in this analysis is voluntary. You may refuse to take part in the research or exit the online questionnaire at any time without penalty.

BENEFITS

You will receive no direct benefits from participating in this research study. However, your responses may help us learn more about factors that influence how well non-native academic English speech is understood.

RISKS

There are no foreseeable risks involved in participating in this study.

CONFIDENTIALITY

Your analysis ratings will be sent to a link at the University of Lancaster-hosted Qualtrics website where data will be stored in a password protected electronic

CONTACT

If you have questions at any time about the study or the procedures, you may contact myself using the following information:

Douglas David Evans

Cell/Mobile: ■■■■■■■■■■

Email: ■■■■■■■■■■

If you have any concerns or complaints that you wish to discuss with my research supervisor, you may contact him using the following information:

Dr. Luke Harding

Phone: ■■■■■■■■■■

Email: ■■■■■■■■■■

→

Powered by Qualtrics

ELECTRONIC CONSENT: Please select your choice below. Clicking the “I consent” button indicates that:

- You have read the provided information about the research study
- You voluntarily agree to participate

I consent, begin the study

I do not consent, I do not wish to participate

→

Powered by Qualtrics

Appendix I. EFL teacher survey: Audio- and transcript-based linguistic measures and background information questions

Note that only several samples of audio- and transcript-based measures are presented here due to their uniformity.

Section 1 - Rating academics' English speech features

You will be analyzing 20 Indonesian academics' speech according to eight categories addressing pronunciation, fluency, and lexicogrammatical linguistic features, and using rating scales developed by Saito, Trofimovich and Isaacs (2014). There is one speech sample per speaker, each between 58-67 seconds in length, for an approximate total of 23 minutes of recording.

Instructions:

1. Please find a quiet place to listen carefully to some speech recordings. If you have earphones / headphones available, it would be good to use for this task.
2. One speech sample will be provided for each speaker. After listening carefully to the recording, rate speech for segmental errors, word stress, intonation and speech rate.
3. A transcript of the speech sample will also be provided. After analyzing it carefully, rate speech for lexical appropriateness, lexical richness, grammatical accuracy and grammatical complexity.
4. Several example tasks are provided in order to familiarize yourself with the rating procedure.

A. Pronunciation and fluency categories	
Segmental errors	This refers to errors in individual sounds. For example, perhaps somebody says "road" "rain" but you hear an "l" sound instead of an "r" sound. This would be a consonant error. If you hear someone say "fan" "boat" but you hear "fun" "bought," that is a vowel error. You may also hear sounds

	missing from words, or extra sounds added to words. These are also consonant and vowel errors.
Word stress	When an English word has more than one syllable, one of the syllables will be a little bit louder and longer than the others. For example, if you say the word “computer”, you may notice that the second syllable has more stress (comPUter). If you hear stress being placed on the wrong syllable, or you hear equal stress on all of the syllables in a word, then there are word stress errors.
Intonation	Intonation can be thought of as the melody of English. It is the natural pitch changes that occur when we speak. For example, you may notice that when you ask a question with a yes/no answer, your pitch goes up at the end of the question. If someone sounds “flat” when they speak, it is likely because their intonation is not following English intonation patterns.
Speech rate	Speech rate is simply how quickly or slowly someone speaks. Speaking very quickly can make speech harder to follow, but speaking too slowly can as well. A good speech rate should sound natural and be comfortable to listen to.

Source: Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37(2), 217–240.

(Note: The definition for a given measure can be reviewed by resting the mouse cursor over the related term in rating instrument)

Rating Procedure:

1. Listen to the recording as many times as you need to in order to be able to rate the speech.
2. Adjust the position of the slider to the location that reflects your rating of the speech feature. Feel free to use the entire range of the scale. The scale has 1000 discrete points, so even the slightest shift of the slider could represent a

significant change in rating.

Example 1:

Listen to the following speech sample of a lecturer giving reasons as to why her subject is important for students to study, and then rate according to pronunciation and fluency measures defined above.

-1:00

1. Vowel and/or consonant errors



Frequent



Infrequent or absent

2. Word stress errors affecting stressed and unstressed syllables



Frequent



Infrequent or absent

3. Intonation (ie. pitch variation)



Too varied or not varied enough



Appropriate across stretches of speech

4. Speech rate



Too slow or too fast



Optimal

Following are the lexicogrammatical measures and related definitions. After reading the transcript of the earlier speech sample, rate according to these measures.

B. Lexicogrammar categories	
Lexical appropriateness	This dimension refers to the appropriateness of the vocabulary words used by the speaker. If the speaker uses incorrect or inappropriate words, including words from the speaker's native language, lexical accuracy is low. On the other hand, lexical accuracy is high if the speaker has all the lexical items required to accomplish the speaking task and does so using frequently-used and/or precise lexical expressions.
Lexical richness	This dimension also refers to the vocabulary used by the speaker. What is important here, however, is how sophisticated this vocabulary is, taking into account the demands of the speaking task. If the speaker uses a few simple, unnuanced words, the speech lacks lexical richness. However, if the speaker's language is characterized by varied and sophisticated uses of English vocabulary, the speech is lexically rich.
Grammatical accuracy	This refers to the number of grammar errors that the speaker makes, including errors in word order and morphological ending.
Grammatical	This dimension is about the complexity and sophistication of the speaker's grammar. If the speaker uses basic, simple or fragmented structures or

complexity

sentences, grammatical complexity is low. Grammatical complexity is high if the speaker uses elaborate and sophisticated grammar structures.

Source: Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37(2), 217–240.

(Note: The definition for a given measure can be reviewed by resting the mouse cursor over the related term in rating instrument)

Rating Procedure:

1. Use the simplified transcript of the speech sample to rate the lexical and grammatical features.
2. Adjust the position of the slider to the location that reflects your rating of the speech feature. Feel free to use the entire range of the scale. The scale has 1000 discrete points, so even the slightest shift of the slider could represent a significant change in rating.

Example 1 - Simplified Transcript:

My favourite subject is math, because you know that math is very powerful...because some reason. The first one, with math you can predict something. For example, when we want to know about prediction about rainfall or about the earthquake, we can use math, because, as we know that math is a mother of language. The second one, with math we can- we can also predict, but in the different way. I mean, when we want to- for example, when we want to produce- to know how many advantage that we have when we will build a houses, if you have a marketing about house...

1. Lexical appropriateness



Many inappropriate words used



Consistently appropriate vocabulary

2. Lexical richness



A few simple words used

Varied vocabulary

3. Grammatical accuracy



Poor grammar accuracy

Excellent grammar accuracy

4. Grammatical complexity



Simple & fragmental grammar

Elaborate grammar



Powered by Qualtrics

A. Pronunciation and fluency categories	
Segmental errors	This refers to errors in individual sounds. For example, perhaps somebody says “road” “rain” but you hear an “l” sound instead of an “r” sound. This would be a consonant error. If you hear someone say “fan” “boat” but you hear “fun” “bought,” that is a vowel error. You may also hear sounds missing from words, or extra sounds added to words. These are also consonant and vowel errors.
Word stress	When an English word has more than one syllable, one of the syllables will be a little bit louder and longer than the others. For example, if you say the word “computer”, you may notice that the second syllable has more stress (comPUter). If you hear stress being placed on the wrong syllable, or you hear equal stress on all of the syllables in a word, then there are word stress errors.
Intonation	Intonation can be thought of as the melody of English. It is the natural pitch changes that occur when we speak. For example, you may notice that when you ask a question with a yes/no answer, your pitch goes up at the end of the question. If someone sounds “flat” when they speak, it is likely because their intonation is not following English intonation patterns.
Speech rate	Speech rate is simply how quickly or slowly someone speaks. Speaking very quickly can make speech harder to follow, but speaking too slowly can as well. A good speech rate should sound natural and be comfortable to listen to.

(Note: The definition for a given measure can be reviewed by resting the mouse cursor over the related term in rating instrument)

Rating Procedure:

1. Listen to the recording as many times as you need to in order to be able to accurately rate the speech.
2. Adjust the position of the slider to the location that reflects your rating of the speech feature. Feel free to use the entire range of the scale. The scale has 1000 discrete points, so even the slightest shift of the slider could represent a significant change in rating.

Example 2:

Listen to the following speech sample of a lecturer giving reasons as to why her subject is important for students to study, and then rate according to pronunciation and fluency measures defined above.

-0:59

1. Vowel and/or consonant errors



Frequent

Infrequent or absent

2. Word stress errors affecting stressed and unstressed syllables



Frequent

Infrequent or absent

3. Intonation (ie. pitch variation)



Too varied or not varied enough

Appropriate across stretches of speech

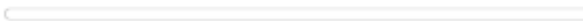


4. Speech rate



Too slow or too fast

Optimal



Following are the lexicogrammatical measures and related definitions. After reading the transcript of the earlier speech sample, rate according to these measures.

B. Lexicogrammar categories	
Lexical appropriateness	This dimension refers to the appropriateness of the vocabulary words used by the speaker. If the speaker uses incorrect or inappropriate words, including words from the speaker's native language, lexical accuracy is low. On the other hand, lexical accuracy is high if the speaker has all the lexical items required to accomplish the speaking task and does so using frequently-used and/or precise lexical expressions.
Lexical richness	This dimension also refers to the vocabulary used by the speaker. What is important here, however, is how sophisticated this vocabulary is, taking into account the demands of the speaking task. If the speaker uses a few simple, unnuanced words, the speech lacks lexical

	richness. However, if the speaker's language is characterized by varied and sophisticated uses of English vocabulary, the speech is lexically rich.
Grammatical accuracy	This refers to the number of grammar errors that the speaker makes, including errors in word order and morphological ending.
Grammatical complexity	This dimension is about the complexity and sophistication of the speaker's grammar. If the speaker uses basic, simple or fragmented structures or sentences, grammatical complexity is low. Grammatical complexity is high if the speaker uses elaborate and sophisticated grammar structures.

(Note: The definition for a given measure can be reviewed by resting the mouse cursor over the related term in rating instrument)

Rating Procedure:

1. Use the simplified transcript of the speech sample to accurately rate the lexical and grammatical features.
2. Adjust the position of the slider to the location that reflects your rating of the speech feature. Feel free to use the entire range of the scale. The scale has 1000 discrete points, so even the slightest shift of the slider could represent a significant change in rating.

Example 2 - Simplified Transcript:

Our subject today is about the history of Islam in Southeast Asia. I want- this- during this semester we would like to study about history of Islam in this area- in this region. And this subject is very important. And I- I want to tell you that- the first one- that the- Islam is the main religion in this country, so you should know the history of Islam, and so why many of people were Muslim in this region. So

the- the reason why many of people were converted to be Muslim.

1. Lexical appropriateness



Many inappropriate words used

Consistently appropriate vocabulary

2. Lexical richness



A few simple words used

Varied vocabulary

3. Grammatical accuracy



Poor grammar accuracy

Excellent grammar accuracy

4. Grammatical complexity



Simple & fragmental grammar

Elaborate grammar



Speaker F3:

Listen to the following speech sample of a lecturer giving reasons as to why her subject is important for students to study, and then rate according to pronunciation and fluency measures below.

-1:04

1. Vowel and/or consonant errors



Frequent

Infrequent or absent

2. Word stress errors affecting stressed and unstressed syllables



Frequent

Infrequent or absent

3. Intonation (ie. pitch variation)



Too varied or not varied enough

Appropriate across stretches of speech

4. Speech rate



Too slow or too fast

Optimal



Using the transcript, rate the sample according to the lexical and grammatical measures below.

Simplified Transcript:

There are some of the key features that maybe I will introduce to my students. The first, I will explain to- to my students that how is the significant subjects that she will be taken for this new era. For the example is the DSS subject course, and DSS, or Decision Support System, is a very usable subject in this era revolution industry, where all the industry become to be digitalized and then they have to make some applications that have managements in making the digitali- the digitalizations everythings – something like that. And then, the second one, I will show to the my students about how is the subject can be applied to the real world. And how is the subject- and how is the subject can solving the problems.

1. Lexical appropriateness



Many inappropriate words used

Consistently appropriate vocabulary



2. Lexical richness



A few simple words used

Varied vocabulary

3. Grammatical accuracy



Poor grammar accuracy



Excellent grammar accuracy

4. Grammatical complexity



Simple & fragmental grammar



Elaborate grammar



Powered by Qualtrics

Section 2 - Understanding and Confidence in Rating Categories

Please select the extent to which you understood the categories used in rating the speech samples.

	I did not understand at all								I understand this concept well
	1	2	3	4	5	6	7	8	9
1. Segmental errors	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Word stress	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Intonation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Speech rate	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Lexical appropriateness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Lexical richness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Grammatical accuracy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Grammatical complexity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please select the extent to which you were able to comfortably use the categories in rating the speech samples.

	very difficult								very comfortable
	1	2	3	4	5	6	7	8	9
1. Segmental errors	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Word stress	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Intonation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

4. Speech rate	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Lexical appropriateness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Lexical richness	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Grammatical accuracy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Grammatical complexity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

→

Powered by Qualtrics

Section 3 - Background information

Please answer the following questions.

What is your age?

What is your sex?

Male

Female

What is the highest level of education you have completed?

High school or earlier

Some college but no degree

Associate degree in college (2-year)

Bachelor's degree in college (4-year)

Master's degree

Doctoral degree

Please specify the discipline of your highest level of completed education.

How many years of experience do you have in language teaching (if none, leave blank)?

What is your current occupational status?

(Select all that apply)

Student

Employed

Retired

None of the above

→

Appendix J. Additional analysis results

The following tables of analysis results are referenced within the main report but included here due to space constraints.

Table J1

Comprehension Question Score Averages by Speaker

Speakers	Comprehension question 1			Comprehension question 2		
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
F01	33	2.56	1.07	32	3.02	.76
F02	35	3.51	.85	35	3.21	.72
F03	34	3.63	.69	34	2.96	.78
F04	33	2.29	.97	35	2.70	.91
F05	35	3.37	.97	33	3.32	.72
F06	34	1.87	.61	35	3.26	.67
F07	35	3.60	.78	35	3.59	.55
F09	35	3.09	.90	34	3.41	.58
F10	33	2.17	.95	33	2.54	.86
F11	35	3.06	.87	35	1.97	.62
F13	35	2.97	.80	33	3.47	.56
F14	34	2.81	.97	35	3.17	.65
F15	34	2.31	.87	35	3.33	.56
M01	34	3.16	.97	34	3.60	.65
M02	34	3.16	1.13	35	3.21	.76
M03	35	2.74	.73	34	3.57	.60
M04	32	3.25	.70	35	3.10	1.03
M05	34	2.21	1.63	34	3.38	.55
M06	32	2.13	1.02	34	3.62	.60
M07	35	3.07	.94	33	3.11	.92

Table J2

Comprehensibility Ratings: Descriptive Statistics by Listener

Listener ID	<i>Min</i>	<i>Max</i>	<i>Med</i>	<i>Range</i>	<i>V</i>
SR01	1	1	1	0	.00
SR02	1	8	4.5	7	4.05
SR03	2	4	3	2	.62
SR04	1	3	2	2	.56
SR05	1	3	1	2	.67
SR06	1	2	1	1	.10
SR07	1	8	1	7	4.31
SR08	1	5	2	4	1.38
SR09	1	2	1	1	.25
SR10	2	3	2	1	.05
SR11	1	2	1	1	.05
SR12	1	9	5	8	5.63
SR13	2	5	3	3	.75
SR14	1	2	1	1	.05
SR15	1	3	2	2	.48
SR16	1	3	1	2	.46
SR17	1	8	1	7	3.29
SR18	1	5	2	4	.80
SR19	1	2	1	1	.10
SR20	1	8	4	7	4.41
SR21	1	9	2	8	3.52
SR22	1	1	1	0	.00
SR23	1	1	1	0	.00
SR24	1	9	8	8	9.74
SR25	1	3	2	2	.43
SR26	1	4	2	3	1.04
SR27	1	5	2.5	4	2.04
SR28	1	7	4	6	2.17

Listener ID	<i>Min</i>	<i>Max</i>	<i>Med</i>	<i>Range</i>	<i>V</i>
SR29	1	3	2	2	.36
SR30	1	2	1	1	.05
SR31	1	7	3	6	1.63
SR32	1	3	1.5	2	.72
SR33	1	2	1	1	.25
SR34	1	4	2	3	.62
SR35	2	7	4	5	2.37
SR36	1	3	1	2	.35
