



L2 Listener Attitudes Towards Speaker Intelligibility, Comprehensibility, and Teaching Quality

Jennifer Payne-Wheeler

Exeter College

Candidate Number

1038600

Submitted in partial fulfilment of the requirements for the degree of
MSc Applied Linguistics for Language Teaching

August 2021

Abstract

Intelligibility and the associated concepts of comprehensibility and accentedness are not objective features of speech, but are at least partially constructed by listener factors. Studies using a matched-guise technique, where listeners are led to believe that they are listening to two different speakers of different ethnicities, while they actually listen to the same native speaker, have found that listeners who believe a speaker to be non-Caucasian may judge them to sound less like a native speaker, to be a poorer quality teacher, and to be less intelligible. While the effects of native listener biases against non-native speakers and marginalised ethnicities have been well explored, fewer studies have focused on non-native listeners, especially in English as a foreign language (EFL) contexts. This study therefore aimed to determine whether the speech and teaching evaluations of English learners in Korea were affected by biases regarding the speaker's ethnicity. 74 Korean L1 university students listened to a speech sample by the same native English speaker. Photographs displayed while listening implied to the listeners that the speaker was either Asian or Caucasian. Listeners rated the speaker on various measures, including personal attributes, comprehensibility, accentedness, and teaching quality, and completed tasks to measure comprehension and intelligibility. It was found that listeners who saw an Asian face were more likely to rate the speaker as having a less native-like accent, and found him less intelligible than listeners who saw a Caucasian face. However, no differences were found in listeners' attitudes towards the speaker's personal qualities or teaching ability. Listeners' knowledge of L2s other than English and their exposure to English speakers were not found to mediate the effects of speaker ethnicity on ratings or listening performance.

Acknowledgements

First and foremost, I must begin by thanking our course director and my supervisor, Dr Heath Rose. His academic guidance and expertise has been invaluable, as has his calming influence and advice throughout the research process. I am also eternally grateful for the support he provided me when I, somewhat naively, decided not to suspend my studies after my baby was born in the middle of my second year. As well as helping me to navigate the complexities of applying for extensions on assignments, allowing me flexibility during the term, and being willing to include a newborn in our supervisions, Heath's confidence that I could handle this challenge gave me the confidence I needed in myself.

I would also like to thank the other people whose academic advice has helped to shape this dissertation, including Naosuke Amano for his input on scoring methods and Sihan Zhou for her guidance on statistics, both of whom helped me to cut through the knot of confusion in which I had found myself. Thanks also to Okim Kang, who kindly sent me some of the instruments she and her colleagues used in the studies which inspired this project.

Other people upon whom this research depended are my fellow teachers who enabled me to expand my pool of participants by recruiting their own students. Thank you to all who helped me in this area, especially to those who agreed to send out my questionnaire at their own universities further afield.

Thanks are also due to my ALLT cohort. Despite only meeting in person for one week and the subsequent distance between all of us, our group chat has been a great source of academic discussion, as well as moral support, humour and camaraderie.

I could not have written this dissertation without the unwavering support and patience of my husband, Mike. Thank you for holding down the fort at home during my many hours beavering away at my desk. Your belief in me and my abilities has kept me going in the most difficult of times.

Last but not least, thank you to my son, Grady, who came into the world in the middle of all this. For the first several months of his life, he attended most of my supervisions, slept on me as I wrote and studied, and was read research articles instead of bedtime stories. Thank you, Grady, for sharing me with this dissertation. It's a shame that you will never remember the free Oxford education you received.

Table of Contents

<i>Abstract</i>	<i>1</i>
<i>Acknowledgements</i>	<i>2</i>
<i>Table of Contents</i>	<i>4</i>
<i>List of Tables</i>	<i>8</i>
<i>List of Acronyms</i>	<i>9</i>
Chapter 1: Introduction	10
1.1 Background to the Study	10
1.2 Aims of the study	12
1.3 Rationale	12
1.4 Dissertation Outline	13
Chapter 2: Literature Review	14
2.1 Intelligibility, Comprehensibility and Accentedness	14
2.1.1 Definitions.....	14
2.1.2 Measurements.....	15
2.2 The Role of Native-Speaker Norms	17
2.3 Reverse Linguistic Stereotyping	20
2.3.1 In ENL Contexts.....	21
2.3.2 In English as a Foreign Language (EFL) Contexts.....	28
2.4 The NEST and NNEST Divide	31
2.5 Gaps in Research	33
Chapter 3: Method	35
3.1 Research Questions	35
3.2 Study Design	35
3.2.1 Quantitative Experimental Design.....	35
3.2.2 Matched-Guise Technique.....	36
3.2.3 Online Format.....	37
3.3 Ethical Considerations	37
3.4 Participants	38
3.4.1 Population.....	38
3.4.2 Recruitment.....	39
3.4.3 Participant Statistics.....	39
3.5 Pilot Study	40
3.6 Guises and Speech Sample	41
3.6.1 Speaker.....	41
3.6.2 Script.....	41
3.6.3 Recording and Editing.....	41
3.6.4 Guises.....	42

3.7 Variables and Measures	43
3.7.1 Language Attitudes.....	43
3.7.2 Comprehension	44
3.7.3 Intelligibility	44
3.7.4 Comprehensibility	46
3.7.5 Accentedness.....	47
3.7.6 Teaching quality.....	47
3.7.7 Listener Background Characteristics	47
3.8 Procedure.....	51
3.9 Limitations and COVID-19 Impact	52
3.9.1 Uneven Groups.....	52
3.9.2 Sampling.....	53
3.9.3 Sample Size.....	53
3.9.4 Online Format	53
Chapter 4: Results.....	55
4.1 Preliminary Analyses.....	55
4.1.1 Equality of Groups	55
4.1.3 Efficacy of Matched-Guise Technique	57
4.2 Relationships Between Variables	58
4.3 Effects of Speaker Ethnicity on Dependent Variables.....	59
4.3.1 Effect of Speaker Ethnicity on Social Attractiveness Ratings.....	60
4.3.2 Effect of Speaker Ethnicity on Superiority Ratings.....	60
4.3.3 Effect of Speaker Ethnicity on Comprehension	60
4.3.4 Effect of Speaker Ethnicity on Intelligibility	60
4.3.5 Effect of Speaker Ethnicity on Comprehensibility Ratings.....	61
4.3.6 Effect of Speaker Ethnicity on Accentedness Ratings	61
4.3.7 Effects of Speaker Ethnicity on Teaching Quality Ratings	62
4.4 Interaction Effects of Listener Background Characteristics and Ethnicity on Dependent Variables.....	62
4.4.1 Interaction Between Condition and Knowledge of L2s Other Than English	63
4.4.2 Interaction Between Condition and Exposure to NESs.....	63
4.4.3 Interaction Between Condition and Exposure to NNEs.....	64
4.5 Summary of Results.....	64
Chapter 5: Discussion	66
5.1 RQ1: Does Speaker Ethnicity Affect Listener Attitudes Towards the Speaker?.....	66
5.2 RQ2: Does Speaker Ethnicity Affect Listener Measures of Comprehension, Intelligibility, Comprehensibility, and Accentedness?	66
5.2.1 Comprehension and Intelligibility	66
5.2.2 Accentedness.....	69
5.2.3 Comprehensibility	71
5.3 RQ3: Does Speaker Ethnicity Affect Listener Judgements of Teaching Quality?	72
5.4 RQ4: Are Any of the Above Effects Mediated by Listener Background Characteristics? 73	
5.5 An Alternative to RLS.....	74
Chapter 6: Conclusion	75

6.1 Summary of Findings	75
6.2 Implications	76
6.2.1 Listening Training.....	76
6.2.2 Addressing Listener Expectations	76
6.2.3 Hiring Practices.....	77
6.2.4 Testing.....	78
6.3 Limitations and Suggestions for Improvement.....	78
6.3.1 Quantitative method	79
6.3.2 Matched-Guise Technique and Unidirectional Listening.....	79
6.3.3 Between-Subjects Design and Listener Biases.....	80
6.3.4 Proficiency Measure.....	81
6.3.5 Speech Sample	81
6.4 Further Recommendations for Research.....	82
6.4.1 Research in EFL Contexts.....	82
6.4.2 Future Intelligibility Studies	82
6.4.3 Expectation-Based Theories.....	82
6.4.4 Effects of Exposure to NESs and NNESs	83
References.....	84
List of Appendices	91
Appendix A: Confirmation of Ethical Approval.....	92
Appendix B: Participant Poster.....	93
Appendix C: Pilot Study Feedback Form	95
Appendix D: Speech Sample Script.....	97
Appendix E: Guise Photographs	98
Appendix F: Attitude Items	99
Appendix H: Cloze Tests	101
Appendix I: Page Displayed For Missing Cloze Test Answers	102
Appendix J: Comprehensibility Items.....	103
Appendix K: Accentedness Items.....	104
Appendix L: Teaching Quality Items.....	105
Appendix M: Items for Exposure N(N)ESs.....	106
Appendix N: Participant Information	108
Appendix O: Descriptive Statistics.....	113
Appendix P: Assumption Testing for Independent Samples T-Tests	116
Appendix Q: Independent Samples t-tests for Mean Dependent Variables.....	122
Appendix R: Independent Samples t-tests for Teaching Quality Items.....	126
Appendix S: Graphic Representations of Independent Samples T-Tests	133
Appendix T: Interaction Effects of Knowledge of L2s Other Than English and Condition on Dependent Variables	137

Appendix U: Interaction Effects of Exposure to NESs and Condition on Dependent Variables.. 149

Appendix V: Interaction Effects of Exposure to NNEs and Condition on Dependent Variables 161

List of Tables

Table 1: Participant Demographic Information

Table 2: Recoding of Original Exposure Variable Values

Table 3: Recoding of Continuous N(N)ES Scores Into Categorical Values

Table 4: Self-Reported Proficiency

Table 5: CEFR Conversion

Table 6: Correlations

List of Acronyms

This dissertation uses the following acronyms.

ANOVA	Analysis of Variance
EFL	English as a foreign language
ELF	English as a lingua franca
ELT	English language teaching
ENL	English as a native language
IAT	Implicit Associations Test
L1	First language
L2	Second language
MWU	Mann-Whitney U
NES	Native English speaker
NEST	Native English-speaking teacher
NNES	Non-native English speaker
NNEST	Non-native English-speaking teacher
RLS	Reverse linguistic stereotyping
RQ	Research question
TOEFL	Test of English as a Foreign Language
TOEIC	Test of English for International Communication

Chapter 1: Introduction

1.1 Background to the Study

The era of globalisation has seen profound and wide-reaching changes to the role, function, and ownership of the English language. Native English speakers (NESs) are now outnumbered by non-native English speakers (NNESs) (Crystal, 2006), many of whom will never have ‘the dubious good fortune’ of meeting a NES (Nelson, 1995: p.276), and many interactions in English now do not involve a NES at all (Galloway & Rose, 2015; Pickering, 2006). Similarly, the majority of English learners in the world learn from non-native English-speaking teachers (NNESTs) (Braine, 2005). As such, using NESs as models for L2 speakers has become less pertinent, with intelligibility emerging as a more relevant target for English learners (Jenkins, 2002).

Despite this situation, preferences for native English-speaking teachers (NESTs) and for particular native varieties of English remain in the field of English language teaching (ELT) and among learners themselves. The privileging of native varieties and simultaneous delegitimisation of non-native varieties has been termed *native-speakerism* (Holliday, 2006). This ideology is not confined to the traditionally native English-speaking Inner Circle countries in Kachru’s (e.g. 1992) three circles model of English, but ‘plays a widespread and complex iconic role outside as well as inside the English-speaking West’ (Holliday, 2006: p.385). This complexity results not only simply in a system where NESTs are preferred over NNESTs, but also in the belief that ‘We Are Two Different Species’, or the assumption that NESTs and NNESTs are fundamentally different and incomparable, and therefore have different strengths and weaknesses (Selvi, 2014: p.584).

In the case of South Korea (hereafter, *Korea*), recent research has shown teachers and students to display preferences for native varieties of English to varying degrees. Park (2012) notes that despite the increasing recognition that English is more likely to be used for

communication with other NNEs, native-speakerism, and the preference for American English specifically, ‘seems to be deep-rooted in Koreans’ minds and practiced widely in Korean society’ (p.319). In particular, native-speakerism is linked for many Korean English users to pronunciation and a belief that there is a ‘correct’ English accent. Bissett and Ma (2015) found that 44.4% of the Korean university students they surveyed strongly agreed that they wanted to sound like a NES, with approximately 73% preferring American English because, as one participant put it, ‘English users mostly speak with an American accent’ (p.9). Ahn's (2014) interviews with NESTs and Korean NNEs in Korea found a range of attitudes towards Korean English, from ‘strong resistance to a full acceptance of the legitimacy of [Korean English] as a developing variety of English’ (p.206), with one teacher naming intelligibility to Americans as the main arbiter of the acceptability and status of Korean English as a legitimate variety of English. Another Korean NNE explicitly linked ‘correct’ English to ‘[t]he English that the white Americans speak’ (Ahn, 2014: p.220), illustrating a perceived connection between native-speaker status, English legitimacy, and race.

According to Shuck (2006), this ‘ideology of nativeness’ (p.260) associates race, nationality, native-speaker status and accent to the extent that these categories become interchangeable. Over the last three decades, research has produced evidence that manipulation of one of these traits as an independent variable can affect listeners’ interpretation of the others. For example, Asian English speakers with stronger non-native accents have been found to be judged as being more Asian than speakers with more moderate non-native accents (e.g., Rubin & Smith, 1990), while NESs may be perceived as having a non-native accent if their ethnicity is implied to be non-Caucasian (e.g. Kang & Rubin, 2009; Rubin, 1992). All three of these studies also found that speakers judged to be NNEs and/or to have a non-native accent were also more likely to be judged as poorer quality teachers.

This relationship undoubtedly has a detrimental effect not only on the employment opportunities of NNESTs, but of non-Caucasian NESTs, who may be openly excluded from consideration for teaching positions in countries such as China, Japan and Korea, by job advertisements specifying that they will only consider Caucasian applicants, or find themselves more covertly discriminated against when prospective employers require photos alongside applications (Ruecker, 2011).

1.2 Aims of the study

This study investigates whether university students in Korea evaluate a teacher differently depending on whether they believe he is Asian or Caucasian. Specifically, the study aims to determine whether the ethnicity of a NEST has any effect on Korean L1 listeners' attitudes towards him, their comprehension of him, or their judgements of his intelligibility, comprehensibility, accentedness, and teaching quality. Also of interest is whether their judgements are mediated by any interaction between the speaker's ethnicity and listener background characteristics, namely proficiency in English, knowledge of any L2s other than English, and extent of exposure to both NESs and other NNESTs.

1.3 Rationale

Although research in Korea has shown, on the one hand, that young learners express preference for American NESTs (Butler, 2007) and, on the other, that university students believe Korean NNESTs and NESTs to be best suited for different aspects of language teaching (Chun, 2014), little research has explored how a teacher's ethnicity affects these attitudes, nor how ethnicity interacts with judgements of intelligibility, comprehensibility, and accentedness. This study aims to fill that gap by investigating the extent to which Korean university students' opinions and perceptions of an American NEST change depending on

whether they believe him to be Caucasian or Asian. The findings will provide a comparison to previous studies which have explored this question in English as a native language (ENL) contexts (e.g. Eisenclas & Michael, 2019; Kang & Rubin, 2009; Rubin, 1992; Rubin & Smith, 1990), as well as insights into current challenges that non-Caucasian NESTs and NNESTs may face in Korea.

1.4 Dissertation Outline

This dissertation is divided into the following chapters: Chapter 1 has introduced the background, aims and rationale of the study. Chapter 2 reviews the literature regarding intelligibility, comprehensibility and accentedness, the shift away from measuring these concepts based on native-speaker norms, and the studies into nonlanguage factors affecting listener measurements of the concepts which have informed the present study. Chapter 3 details and justifies the methods used in the study. Chapter 4 presents the quantitative results, which are discussed in further detail and related back to previous research in Chapter 5. The dissertation concludes with Chapter 6, which presents a summary of the findings, discusses the limitations of the research, and presents some implications and recommendations.

Chapter 2: Literature Review

2.1 Intelligibility, Comprehensibility and Accentedness

2.1.1 Definitions

Definitions of intelligibility and comprehensibility differ between studies and researchers, presenting difficulties when comparing findings across studies (Kang et al., 2018), but Munro and Derwing's (1995) definitions are perhaps the most common. They define *intelligibility* as 'the extent to which a speaker's message is actually understood by the listener' (p.76), while *comprehensibility* is the listener's judgement of how easy or difficult that message is to understand. Munro and Derwing differentiate accentedness from intelligibility and comprehensibility, equating *accentedness* to *non-native accent*, i.e. the extent to which a NNES's pronunciation differs from that of a NES. This study will also generally use the terms *accentedness* and *accent* in the same manner, although it is acknowledged that these terms are imprecise and often problematic. As Derwing and Munro (2009) point out, 'everyone has an accent' (p.476) and Lippi-Green (2012) clarifies that '*accent* can only be understood and defined if there is something to compare it with' (p.45). *Accentedness* should therefore be understood as a subjective, context-specific construct, measured against a baseline which varies from context to context and even listener to listener.

Intelligibility, comprehensibility and accentedness are related but different constructs, as demonstrated by Munro and Derwing (1995). In their study, 10 proficient NNESs (L1 Mandarin) and two NESs were recorded in an unprepared speaking task, from which short extracts were taken and played to 18 NES listeners. Intelligibility was measured by the listeners' transcriptions of each extract, while perceived comprehensibility and accentedness were each rated on a 9-point Likert scale. There was a positive correlation between comprehensibility (rated from high to low comprehensibility) and accentedness (rated from *no foreign accent* to *very strong foreign accent*) for all but one of the listeners, but the

strength of the correlation varied between listeners, from 0.41 to 0.82. Similarly, intelligibility (scored from low to high intelligibility) and comprehensibility were correlated, but the strength of the correlation varied among listeners, from -0.44 to -0.90. Intelligibility and accentedness were significantly correlated for only 28% of listeners, from -0.37 to -0.48. The researchers conclude that high intelligibility is not dependent on a native or near-native accent, nor does a foreign accent necessarily equate to low intelligibility. Furthermore, they suggest that the difference between intelligibility and comprehensibility may be attributable to processing effort: while two extracts may both be fully understood, one may require more effort to process.

2.1.2 Measurements

The above-mentioned difficulty in defining the constructs of intelligibility, comprehensibility and accentedness may be complicated by a lack of consistency in methods of measurement, which may contribute to contradictory research findings.

2.1.2.1 Measurements of Intelligibility. Listener transcriptions of short utterances are one of the most common means of measuring intelligibility (e.g. Barrass et al., 2020; Munro & Derwing, 1995; Zielinski, 2008). This method is less suitable for longer utterances, for which cloze exercises are often used instead. Some studies (e.g. Eisenclas & Michael, 2019; Kang & Rubin, 2009; Rubin, 1992) present these cloze exercises at the end of a several-minute-long speech sample, arguably also measuring listeners' memory capacity and ability to guess missing words from the surrounding context. This measurement may therefore be unsuitable for NNES listeners as it may be confounded by proficiency, especially for lower-proficiency listeners. Another method is listeners' answers to comprehension questions (e.g. Orikasa, 2016). The use of post-listening cloze tests and

comprehension questions reveals the overlap between comprehension and intelligibility in many studies, even though the former is more likely to target global understanding of a message, and the latter speech perception on a much more local, even phonemic, level. Finally, studies investigating intelligibility in interactions rather than unidirectional listening may use measures such as degree of successful task completion (e.g. Lindemann, 2002) or conversation analysis (e.g. O'Neal, 2015).

2.1.2.2 Measurements of Comprehensibility. Many studies (e.g. Derwing, Munro, & Thomson, 2008; Kennedy & Trofimovich, 2008; Sheppard et al., 2017) use a single-item measure of comprehensibility, which asks listeners to rate the ease or difficulty with which they understood the speaker. However, it is not clear that listeners asked to rate comprehensibility in this way can always differentiate between a speaker's comprehensibility and perceived proficiency. In a study exploring dynamic comprehensibility ratings, Nagle et al. (2019) provided listeners with computer software which allowed them to adjust their comprehensibility ratings while they listened to longer passages of speech. Listeners explained the reasons for their rating behaviour in stimulated recall interviews. The researchers found that listeners frequently downgraded the speaker's comprehensibility for errors that would not necessarily negatively impact understanding. This may suggest that listeners are not always able to differentiate between errors that affect comprehensibility and those that do not; that they may not be able to differentiate proficiency from comprehensibility; or that they may make rating decisions based on what they expect other listeners to understand rather than on what they themselves understand. The participants in Nagle et al.'s study also gave a single, global comprehensibility rating in addition to their dynamic ratings. The researchers found that moments of low comprehensibility, as indicated by the dynamic ratings, had a greater impact on global comprehensibility ratings than did

moments of high comprehensibility. In other words, listeners focused unevenly on instances where comprehensibility was reduced when determining the overall comprehensibility of an utterance, perhaps suggesting that these moments are overrepresented in global comprehensibility ratings at the expense of how well the overall message was understood. These findings may indicate that using several explicit items to measure comprehensibility may be necessary to obtain more valid measures.

2.1.2.3 Measurements of Accentedness. As with comprehensibility, measures of accentedness are usually based on the listener's subjective judgment, indicated by a single-item rating (e.g. Kang & Rubin, 2009; Munro & Derwing, 1995; Zheng & Samuel, 2017). This item usually asks the listener to rate the degree to which the speaker's accent is similar or different to a NES's accent. Some studies (e.g. Munro & Derwing, 1995) compare these listener ratings to expert phonetic transcriptions and pronunciation analysis of the given extracts, though arguably this too is ultimately subjective. However, inter-rater reliability on measurements of accentedness is usually high. According to Derwing and Munro (2009), 'listeners' judgements are the only meaningful window into accentedness and comprehensibility... [W]hat listeners perceive is ultimately what matters most' (p.478).

2.2 The Role of Native-Speaker Norms

Increasingly, the value of judging intelligibility, comprehensibility and accentedness according to native-speaker norms is being called into question. With the rise of globalisation, English has provided a linguistic connection between speakers of different languages and has become a lingua franca (Galloway & Rose, 2015). Interactions using English as a lingua franca (ELF) do not necessarily follow the norms and rules assumed in native Englishes, such

as native-speaker pronunciation or grammatical correctness, but may prioritise other aspects of communication, such as intelligibility and mutual understanding. Therefore, when discussing, researching, or teaching ELF, ‘models and practices that privilege native varieties of English are no longer serviceable’ (Pickering, 2006: pp.219-220). Furthermore, factors which affect intelligibility and comprehensibility for NNES listeners may differ from those which affect intelligibility and comprehensibility for NES listeners (Pickering, 2006). Research which investigates intelligibility and comprehensibility in these contexts is therefore needed for a fuller understanding of how these constructs affect communication outside the comparatively narrow sphere of NES-NES or NES-NNES interactions.

However, native-speakerism is still persistent in research and in the attitudes of English learners. Many studies still rely on native-speaker judgments of non-native speech to reach conclusions about intelligibility and comprehensibility, and to make recommendations for teaching (e.g. Kennedy & Trofimovich, 2008; Nagle et al., 2019; Zielinski, 2008). While such studies are of course relevant when considering NNESs’ intelligibility and comprehensibility to NESs specifically, findings must not then be held as universal or applied uncritically to NNES-NNES interactions. Indeed, Rajagopalan (2010) suggests that the enduring appeal to NESs’ judgements undermines the credibility of intelligibility studies as a whole, turning it into an area where native-speaker norms are quietly reasserted, despite an overall move away from this in the wider field. In intelligibility studies, ‘the figure of the native speaker creeps back in, only this time through the back door and that too most stealthily’ (Rajagopalan, 2010: p.468). This persistence of native-speaker norms is also felt by NNESs, who may focus on their intelligibility to an ‘imaginary NES hearer’ (Wang & Jenkins, 2016) rather than their actual NNES interlocutors.

There are several problems with judging NNESs’ speech by native-speaker norms. First, basing recommendations for improving intelligibility and comprehensibility on studies

which use native-speaker judgments of NNESs' speech means that learners who communicate primarily with other NNESs may be focusing their time, energy and effort on areas which will not improve their communicative abilities with their most common interlocutors. Such internalised native-speakerism may even hinder communication between NNESs. Second, as Rajagopalan (2010) suggests, focusing on what makes speech intelligible to NESs reinforces NESs' ownership of English. This can have detrimental effects on NNESs' identities and social power (Norton, 1997) and position NESs as gatekeepers of the language and of who is considered a legitimate English user (Widdowson, 1994). Finally, positioning NESs as the final arbiters of what is intelligible or not removes any share of the burden for successful communication from NESs and places it fully on NNESs. Lippi-Green (2012) contends that NESs frequently 'feel perfectly empowered to reject their responsibility, and to demand that a person with an accent carry the majority of the burden in the communicative act' in a way which they do not when they encounter other NESs with low intelligibility or comprehensibility (p.72).

In sum, native-speakerism presents intelligibility and comprehensibility as exclusively speaker characteristics, rather than as constructs that can be affected at least partially by listener characteristics and listener strategies, while also framing NNESs who do not adhere to native-speaker norms as deficient speakers of the language.

A recent example of how intelligibility studies can avoid judging NNESs by native-speaker norms is Barrass et al.'s (2020) study on the features of Korean L1 speakers' pronunciation which negatively affect their intelligibility and comprehensibility to Mandarin L1 listeners. Fourteen Korean L1 speakers each recorded 23 sentences which each contained one phonological feature predicted to cause pronunciation issues based on previous contrastive analyses of English and Korean. Intelligibility was measured by 65 Mandarin L1 listeners' transcriptions of the sentences. Comprehensibility was measured by the listeners'

subjective rating on a 7-point semantic differential scale of how difficult or easy they found each sentence and a longer unscripted utterance to understand. The study was able to identify which phonological features caused breakdowns in intelligibility and comprehensibility for the Mandarin L1 listeners, which features had the largest effect, and how they affected intelligibility and comprehensibility measures differently. While there was a moderate correlation between intelligibility and comprehensibility ratings, the same features did not necessarily affect intelligibility and comprehensibility to the same extent. Some phonological features predicted to cause intelligibility or comprehensibility issues did not do so, despite diverging from native-speaker pronunciation. For example, assimilation of [l] and [r] – a feature stereotypically associated with low intelligibility of Korean L1 speakers, as well as L1 speakers of other East Asian languages – was not shown to cause any intelligibility or comprehensibility breakdown for the Mandarin L1 listeners. This study therefore indicates that features of NNEs' pronunciation that diverge from native-speaker pronunciation do not always cause intelligibility or comprehensibility issues, depending on the language background of the listener. The researchers suggest that teaching should focus on features which *have* been shown to cause problems for the populations with whom the learners are most likely to communicate, rather than attempting to bring all pronunciation in line with native-speaker norms. This study also provides more evidence that intelligibility and comprehensibility ratings, i.e. how much participants actually understand and how difficult they find it to understand, do not always exactly align.

2.3 Reverse Linguistic Stereotyping

Intelligibility, comprehensibility and accentedness are not objective features of a speaker's speech. They are also constructed by the listener, and can be affected by listener characteristics such as language background, previous exposure to similar speech, English

proficiency, and willingness to negotiate intelligibility with the speaker. Listener attitudes, for example towards the accent, nationality, ethnicity or social group membership of the speaker, can affect interpretations of intelligibility, comprehensibility and accentedness. This in turn can affect how the listener judges the speaker's speech and the speaker themselves, and in some cases change how the listener interacts with the speaker.

Rubin and his colleagues (e.g. Kang & Rubin, 2009; Rubin, 1992; Rubin et al., 2016) have put forward the theory of *reverse linguistic stereotyping* (RLS) as a key factor in how the speech of NNESs or stigmatised groups is received by listeners. Whereas linguistic stereotyping is a process wherein listeners make judgements about a speaker based on their speech, RLS is a stereotyping process where a listener's beliefs about a speaker's group membership – social, ethnic, national, etc. – cause 'distorted perceptions' (Kang & Rubin, 2009: p.442) of the speaker's speech, including perceptions of accentedness and actual comprehension.

2.3.1 In ENL Contexts

RLS studies initially focused on NNES university instructors in the USA. In his seminal study, Rubin (1992) noted that the blame for intelligibility difficulties among NNES instructors is often placed solely on the instructor-speaker rather than the student-listeners. Rubin set out to investigate whether this blame is appropriately placed, using a matched-guise technique now common in RLS studies. The listeners were 62 North American undergraduate students, who listened to one of two audio lecture extracts delivered by an American NES speaker. The lecturer's ethnicity was implied by displaying a photograph of either a Caucasian woman or an Asian woman while the audio played. Immediately after listening, the listeners completed a cloze test of the lecture transcript, while semantic differential items were used to rate homophily (how similar the listeners judged the speaker

to be to themselves), ethnicity, accent, and teaching quality. Rubin found that listeners shown an Asian face alongside the audio lecture were more likely to rate the speaker as more accented, and were more likely to comprehend less of the lecture compared to listeners shown a Caucasian face, even though both groups of participants listened to the same lectures delivered by the same NES speaking with a standard American accent. A limitation of the study, as Rubin acknowledges, is that participants did not include NNEs and that, due to the region of the USA in which the study took place, the participants may have had very little exposure to either NNEs or Asians in general. Rubin therefore ends by suggesting that participants with greater exposure to more diverse groups may have produced different results.

This hypothesis was not entirely borne out by later research. Kang and Rubin (2009) found that the propensity to engage in RLS can be measured as a listener characteristic, but that it was not affected by exposure to NNEs. Like Rubin (1992), this study found that participants judged a speaker to be less comprehensible, more accented, and a worse teacher if they believed the speaker to be Asian rather than Caucasian. Furthermore, these ratings correlated with the listeners' propensity to engage in RLS. The 158 participants included both NESs (n = 102) and NNEs (n = 56), deliberately sampled from a group of volunteers to represent a wide range of linguistic backgrounds and exposure to NNEs. As in Rubin's (1992) study, the participants listened to lecture extracts delivered by the same NES with a standard American accent, whose ethnicity was implied to be either Asian or Caucasian by a photograph paired with the recording. In Kang and Rubin's (2009) study, however, all participants were presented with both the Asian and Caucasian guise in a within-subjects design, as opposed to Rubin's (1992) between-subjects design. The study used Zahn and Hopper's (1985) Speech Evaluation Instrument to measure participants' language attitudes towards the two guises in three dimensions: superiority, social attractiveness, and dynamism

(although dynamism was removed from analysis, since higher or lower dynamism ratings do not necessarily correlate with positive or negative attitudes towards the speaker, which was the focus of Kang and Rubin's study). Accentedness and teaching quality were rated on 7-point semantic differential scales, while intelligibility (or comprehension) was measured by a cloze test immediately after listening to each lecture. The researchers found that participants with a higher degree of superiority RLS were more likely to rate the Asian guise as more accented than the Caucasian guise. Those with higher social attractiveness RLS were more likely to perform more poorly on the cloze test for the Asian guise, suggesting that they found the Asian guise less intelligible. Participants with either type of RLS rated the Asian-guise more harshly in terms of teaching quality. These rating behaviours were not limited to the NES listeners, but were also evident for NNES listeners. The NES participants had higher listening comprehension for the Asian guise than NNES listeners, but not for the Caucasian guise, suggesting that NNES listeners were more likely to find someone they believed to be a fellow NNES to be even more difficult to understand. In terms of other listener characteristics, more formal study of linguistics or foreign languages (*linguistic sophistication*) correlated with better comprehension, and participants with English teaching experience were more likely to rate the 'Asian' teacher's teaching more highly. General contact with NNESs did not have a statistically significant effect, leading the authors to conclude that 'mere contact between groups does not guarantee reduction of stereotypes and prejudice' (Kang & Rubin, 2009: p.452). However, the participants were asked to report the number of hours per week they spent with NNESs in general, whereas the listening task and ratings specifically targeted an Asian speaker. It is possible that the measure of contact with NNESs was too general to show any significant relationship with participants' evaluations of an Asian speaker, and that significant results could have been found if they instead reported time spent with Asian people.

Yi et al. (2014) found that not only does listening to a foreign accent place heavier demands on neural processing of speech, but also that stronger listener bias against a speaker's ethnicity increases the processing demands when the listener sees a race-matched face. This study measured bias using an implicit associations test (IAT), which measures bias by calculating the difference in response time and accuracy when participants are asked to categorise congruous-according-to-bias images or words together versus incongruous-according-to-bias images or words. In Yi et al's (2014) study, the congruous category grouped Caucasian faces with American scenes, and Asian faces with foreign scenes, while the incongruous category grouped Caucasian faces with foreign scenes, and Asian faces with American scenes. Higher IAT scores indicated greater bias towards associating Asian faces with foreignness. Nineteen American NES participants without significant exposure to languages other than American English heard 80 sentences produced by four American NESs and four Korean L1 NNEs. The sentences were presented either without visual stimuli, or with the video of the speaker saying the sentences. This differed from the matched-guise studies discussed above, as in this case the participants saw the face of the actual speaker when they listened to NNEs speech in the audio-visual condition. The participants listened to the sentences in a functional magnetic resonance imaging scanner. The researchers found that hearing a foreign accent was associated with increased brain activity, suggesting greater processing demands, and that this activity was increased in the audio-visual condition for Asian NNEs. This effect was heightened in individuals with higher IAT scores. Limitations of this study include the small sample size and the within-subjects design: as all participants experienced both the audio-only and audio-visual condition, they may have been primed to attend more than usual to accent and ethnicity. Furthermore, these results cannot be generalised to NNEs listeners, as L2 listening may not be processed in the same way as L1 listening. Furthermore, the study was US-centric in that it operationalised bias towards

Asians as the extent to which Asian faces were or were not associated with American scenes. It is unclear whether the same results would be found in other contexts.

Possible evidence that Yi et al.'s (2014) findings are context-specific is a recent study by Eisenclas and Michael (2019), which suggests that socio-contextual factors may mitigate the effect of ethnicity on comprehensibility and accentedness judgements. This set of two matched-guise experiments was based on the first of Rubin's 1992 studies and used a similar methodology, although in this case the participants were 252 undergraduate students at an Australian university, split across the two identical experiments (Experiment 1: n=109; Experiment 2: n=143). Unlike Rubin's (1992) study, but similarly to Kang and Rubin's (2009) study, Eisenclas and Michael (2019) included both NES and NNES participants. The linguistic sophistication of Eisenclas and Michael's participants was relatively high, with 100% and 76.3% of participants in the two experiments respectively studying at least one foreign language, and 50.5% and 73.4% speaking another language fluently. While linguistic sophistication was not a factor reported or accounted for in Rubin's (1992) study, Kang and Rubin's (2009) study deliberately sampled participants from a range of linguistic backgrounds, and therefore with a wider range of linguistic sophistication. Eisenclas and Michael's (2019) data collection method differed from the earlier two studies: participants completed their task individually online, they answered only comprehension questions after the audio lecture rather than completing a cloze test, and they rated the speakers on an instrument measuring personality traits linked to competence, integrity and social attractiveness rather than homophily. The researchers found almost no statistically significant effects: although participants in Experiment 1 rated the Asian guise as being marginally less sincere than the Caucasian guise, there were no other differences between personality ratings, and there was no difference between the guises in terms of comprehension for either experiment. The researchers posit a number of explanations for why their findings differ from

Rubin's (1992) findings. For example, Rubin (1992) used only domestic students as participants, whereas Eisenchlas and Michael (2019) included international students. This does not necessarily explain why Eisenchlas and Michael's findings differed from Kang and Rubin's (2009), as the latter also included NNEs in their participants. Eisenchlas and Michael also point out that the proportion of international students is higher in Australia than in the USA and, although Rubin (1992) did not report the number of international staff working at his university, Eisenchlas and Michael's study took place in a university with a high number of international staff. The researchers therefore hypothesise that their context may have been more diverse, with participants regularly exposed to positive interactions with NNEs. Furthermore, although Rubin gave no information on his participants' linguistic backgrounds, Eisenchlas and Michael's participants showed a high level of multilingualism and linguistic sophistication. This theory is supported by Kang and Rubin's (2009) finding that linguistic sophistication was positively correlated with comprehension of the Asian guise. Finally, the effect of time may also contribute to Eisenchlas and Michael's findings: multiculturalism increased greatly over the 27 years since the publication of Rubin's (1992) study, and even in the 10 years since Kang and Rubin's (2009) study. It is possible that more increased multiculturalism in society in general in that time has somewhat mitigated the effects found in both earlier studies, despite Kang and Rubin's (2009) finding that individual exposure to NNEs did not have a statistically significant effect.

Another example of how listener bias-related studies may have different results in different contexts is Babel and Russell's (2015) study, which found that listeners in Canada respond differently to speakers depending on their ethnicity, but that this is not necessarily linked to biases. The researchers played a total of 120 sentences recorded by six White Canadian and six Chinese Canadian NNEs to 40 NNE listeners. Half of each speaker's sentences were presented in an audio-only condition, and half in an audio-visual condition

where the audio was accompanied by a photograph of the speaker's face. Participants transcribed each sentence to measure intelligibility, and rated each speaker's accentedness on a 9-point Likert scale. Listeners' implicit biases were operationalised using an IAT which measured listeners' association between commonly Asian or White surnames and positive or negative words. Explicit biases were measured by asking participants to rate their agreement with a series of 10 stereotypes about Asian Canadians and White Canadians on a 7-point Likert scale. The researchers found that Chinese Canadians were less intelligible overall, and even less intelligible in the audio-visual condition. White Canadians were rated as less accented in the audio-visual condition than the audio-only condition, while Chinese Canadians were rated as more accented in the audio-visual condition than the audio-only condition. There was no association between listeners' IAT scores or explicit bias scores and changes in intelligibility or accentedness across conditions, suggesting that this effect was not due to biases against Asian Canadians. However, the bias tests were administered immediately after the listening task, which could have affected participants' performance. For example, hearing a series of Chinese Canadian NESs may have weakened implicit or explicit biases against Asian Canadians. Participants may also not have been truthful on the explicit bias test, as racial bias is a sensitive subject and, by this point in the procedure, after completing the listening task and IAT, many participants may have correctly guessed the aims of the study. Furthermore, associating Chinese Canadian NESs with NNES speech is itself a racial bias, though perhaps one not targeted by the bias tests used in this study.

The findings of the studies discussed so far in this section present potentially worrying implications for Asian English speakers, both NESs and NNESs, in ENL contexts. For Asian English speakers working as teachers in these contexts, both their speech and teaching quality may be judged more harshly, regardless of their actual ability in these areas,

simply because of their ethnicity. This effect may also be true for other non-Caucasian NESs and NNEs.

2.3.2 In English as a Foreign Language (EFL) Contexts

Kang and Rubin's (2009) finding that NNEs listeners also engage in RLS when listening to (who they believe to be) other NNEs suggest that non-Caucasians may also face this bias when working in EFL contexts. For example, NNEs instructors teaching in English as a medium of instruction contexts may be judged more harshly than their NES colleagues by their NNEs students. Similarly, lower regard of a NNEs's English and teaching ability may be disadvantageous to NNEsTs.

Recently, some studies have investigated NNEs evaluations of NNEs speech in EFL contexts. Hendriks et al. (2021) found that NNEs are especially critical of other NNEs with whom they share an L1. This study used an unstated number of speakers to record the same fragment of a lecture in English, resulting in listening stimuli with three accents: moderate Dutch, slight Dutch, and British English. There were 522 participants, categorised into three groups: 189 Dutch L1 NNEs, 175 NNEs with L1s other than Dutch, and 158 NESs. Each listener heard only one speaker. Intelligibility was measured by cloze tests where listeners transcribed eight missing words in four sentences from the speech sample, although it is not clear whether this was administered during or after listening, or whether the speech sample was presented as a whole or whether the four sentences were presented separately. Among other ratings, accentedness was measured using two 7-point Likert scale items; comprehensibility was measured using six 7-point Likert scale items; attitudes towards the speaker, divided into status, competence and likeability, were measured using eighteen 7-point Likert scale items; and teaching quality was measured using five 7-point Likert scale items. The researchers found that all three accents were equally intelligible, but that the

moderately-accented Dutch speakers were evaluated less positively than slightly-accented speakers and native speakers by NNES listeners in general, and most harshly by the Dutch L1 listeners. Specifically, the Dutch L1 listeners rated the moderately-accented Dutch speakers significantly lower on measures of comprehensibility, competence, likeability, status and teaching quality than the other speakers. There were no significant differences between the ratings of the three accents by NES listeners. The researchers conclude that negative evaluations of NNES speakers by NNES listeners may be amplified in the case of listeners who share an L1 by listeners' 'vicarious shame' about their own English accents (Hendriks et al., 2021: p.20).

Whereas Hendriks et al. (2021) found that NNES listeners' evaluations of actual NNES accents were more critical than of NES accents, but that intelligibility was not significantly different, Hu and Su's (2015) 105 Cantonese L1 university students at a Chinese university found an American NES less intelligible when they believed her to be a fellow Cantonese L1 speaker than when they believed her to be a NES. This between-subjects matched-guise study did not imply the speaker's ethnicity using photographs, unlike the earlier matched-guise studies discussed above. Instead, participants were informed before listening that the speaker was from either the USA or China, that she was applying for an English-teaching position at their university, and that their task was to test her suitability. Thus, this study explicitly identified the Cantonese guise as a NNES, and also raised the stakes for the listeners by positioning their participation and performance as partly responsible for the employment of someone who could, as far as they knew, be their future teacher. This may have affected results. In this study, the only dependent variable was participants' performance on a number of comprehension tests. On all but one measure, participants who were informed that the speaker was American scored significantly more highly than those who were informed that she was Chinese. While this could indicate that

participants suffered comprehension loss when they believed they were listening to a NNES, it could also indicate that participants preferred their potential future teacher to be a NES and therefore exerted more effort in a task that they believed could affect the likelihood of her employment in their institution.

Buckingham's (2014) 373 participants (presumably all Omani Arabic L1 speakers, although this is not clearly stated) were also told that the speakers in the within-subjects matched-guise study were applying for teaching positions at their university in Oman. They heard a total of 10 short speech samples (57 to 127 seconds), two recorded by each of five NNESs (nationality and L1s: Pakistan, Urdu; Iran, Azeri/Farsi; Kenya, Luo/Swahili; Syria, Arabic; and Oman, Arabic), but were told that they were listening to 10 different speakers. One recording from each speaker was presented with the speaker's real nationality, while the second recording from each speaker was presented with a fictitious nationality. Four of the fictitious guises were said to be NESTs from England, the USA and Canada, and one a NNEST from the Philippines. Participants completed a questionnaire gathering their background information as well as their attitudes towards NESTs and their pronunciation goals. They then listened to the speakers and rated each guise on their English ability and pronunciation, status, social values, and suitability and attractiveness as a teacher. The pronunciation of three of the four NEST guises were rated more highly than their NNEST matched-guise counterparts, and all of the NES guises were rated higher overall than their NNEST counterparts. However, the NES guises were not all more highly rated than all the NNEST guises: after the two UK NES guises, the next most highly-rated guises were the two Arab guises. Buckingham therefore suggests that the participants viewed the Arab teachers 'as models for their own learning path' (2014: p.67). Notably, the Pakistani speaker received the lowest overall ratings when his real nationality was presented, and the highest overall ratings when he was presented as a UK NEST. Buckingham posits that Pakistani English

speakers are regarded by Omani students as having particularly low social prestige. The stark difference in this speaker's ratings may therefore demonstrate the significant role that listener attitudes towards a speaker's ethnicity may have on their evaluations of the speaker as a teacher of English. However, the data collection instrument which gathered these ratings was short, with only nine items in total, and only one to three items per construct. The constructs are therefore fairly bluntly measured.

To summarise, the research on listener attitudes towards non-Caucasian English speakers has had mixed findings, demonstrating various levels of bias towards non-Caucasians and NNEs, their pronunciation, and their suitability as teachers. These differences could be due to the different ways in which the relevant constructs are measured, or to the highly context-specific nature of biases, attitudes and their ramifications, or both.

2.4 The NEST and NNE Divide

Some of the research above indicates that English L2 learners in some contexts hold different attitudes towards NESTs and NNEs. However, these attitudes do not always correlate to overall positive or overall negative, or an overall preference for NESTs over NNEs or vice versa. Rather, learners, administrators and researchers often propose that NESTs and NNEs have different strengths – a view which is evident in contexts where NESTs tend to be used for teaching speaking, writing and listening, while NNEs teach grammar and reading (Selvi, 2014).

This division of labour is common in the Korean context, and research has demonstrated that Korean learners view the strengths of their NESTs and NNEs differently. For example, Chun (2014) investigated Korean university students' beliefs about NESTs and Korean NNEs. A total of 156 participants completed a questionnaire about their attitudes towards NESTs and Korean NNEs, including their perceived attributes,

strengths and weaknesses. Responses from participants with no experience learning with NESTs were removed, leaving 125 for analysis. Chun found that participants had no overall preference for NESTs or Korean NNESTs, but perceived them to have different strengths and weaknesses. NESTs were believed to have ‘superior linguistic competence’ and ‘status’, while Korean NNESTs were believed to be better at guiding students ‘with the psychological aspects of language learning’ and to be more sensitive to students’ needs (Chun, 2014: p.574). NESTs were perceived to use a greater range of teaching methods, but were also more likely to be seen as unprepared, possibly due to a perception among participants of NESTs being more casual than Korean NNESTs. Korean NNESTs were believed to be more suitable for beginner-level students, as NESTs’ perceived inability to speak the students’ L1 was seen as a frustrating communicative barrier. Chun concludes that the study shows that NESTs and Korean NNESTs have different teaching styles, but that neither is overall better than the other. However, this is arguably not what her results show – rather, they show that students *believe* that NESTs and Korean NNESTs have different teaching styles. These self-reported beliefs may come from experience (e.g. Korean NNESTs really are better at teaching grammar than NESTs), or they may come from institutional limitations (e.g. institutions assign Korean NNESTs to teach grammar and NESTs to teach speaking), or they may come from stereotyping. Chun points out that, when asked to ascribe attributes to NESTs or Korean NNESTs or both, individual participants had a greater number of responses to NESTs overall. This may suggest that the participants were more likely to stereotype NESTs by assigning attributes to them as a group, possibly caused by their greater exposure to Korean NNESTs than NESTs. As Chun also acknowledges, participants’ perceptions of NESTs may be due to ‘novelty effects’ (2014: p.577). Although not mentioned in Chun’s study, intelligibility, comprehensibility and accentedness may also play into these beliefs. Extrapolating from the findings of Rubin (1992) and Kang and Rubin (2009), learner attitudes towards NESTs and

NNESTs may be intertwined with their beliefs about intelligibility, comprehensibility and accentedness. In other words, if learners believe that NNEs are less intelligible, less comprehensible and more accented, this may affect their attitudes towards the teaching abilities of NNEs, which may in turn further affect their judgements of intelligibility, comprehensibility, and accentedness.

2.5 Gaps in Research

To summarise the research discussed above, studies have indicated that listener perceptions of and attitudes towards speaker ethnicity and native-speaker status can affect intelligibility scores as well as judgements of comprehensibility, accentedness, and teaching quality. While earlier studies in this field have taken place in an ENL context and have shown the disadvantages that NNEs or non-Caucasian English speakers face in ENL contexts, more recent studies have examined the effects of NNE listener attitudes towards the above factors in EFL contexts. How these disadvantages extend to NNEs among other NNEs or to what extent, how attitudes may differ in EFL contexts, and whether negative attitudes towards certain varieties of English have different consequences in EFL nations than in ENL nations seems to vary depending on country and context. A gap in the research is the extent to which Asian English speakers are disadvantaged (or otherwise) in Korea, particularly in the role of teachers of English. Furthermore, although research has suggested that English learners in Korea hold differing beliefs about the strengths and weaknesses of NESTs and NNEs, it is not clear how far these beliefs can be ascribed to actual experience, and how much to stereotype.

This study therefore aimed to investigate whether Korean university students hold biases which would affect how intelligible, comprehensible and accented a speaker they

believed to be a fellow Korean English speaker was, and whether their evaluations of that speaker's ability as a teacher would change depending on his perceived ethnicity.

Chapter 3: Method

This chapter discusses and justifies the methods used in this study, including the study design, participants, materials, variables and measures, procedure, and limitations. Data analysis is discussed in the following chapter. In short, this study consists of a between-subjects experiment in which participants evaluated a speech sample recorded by a NES on measures of social attractiveness, superiority, comprehensibility, accentedness and teaching quality, as well as completing two tasks to measure comprehension and intelligibility. To determine whether speaker ethnicity changed listeners' evaluative or listening behaviour, the participants were divided into two groups. In one group, the speaker was implied to be Caucasian, while in the other he was implied to be Asian.

3.1 Research Questions

To achieve the aims detailed in Section 2.5, this study set out to answer the following research questions:

- RQ1 Does speaker ethnicity affect listener attitudes towards the speaker?
- RQ2 Does speaker ethnicity affect listener measures of intelligibility, comprehensibility, and accentedness?
- RQ3 Does speaker ethnicity affect listener judgements of teaching quality?
- RQ4 Are any of the above effects mediated by listener background characteristics?

3.2 Study Design

3.2.1 *Quantitative Experimental Design*

A quantitative design was chosen as it was considered most appropriate to fulfil the study's aim of establishing whether phenomena previously explored mainly in ENL contexts were applicable to the EFL context of Korea. To maximise the possibility for comparison

with previous RLS research begun by Rubin and colleagues, particularly Rubin (1992) and Kang and Rubin (2009), the present study closely followed their design and data collection methods, employing an experimental listening task and questionnaire. The study followed Rubin (1992) and Eisenclas and Michael (2019) in using a between-subjects design with participants randomly allocated to two experimental conditions. The decision to use a between-subjects design, rather than the within-subjects design used by Kang and Rubin (2009), was made because the latter was predicted to be too long for the online format of the present study, which risks a higher rate of dropout and non-response than paper surveys (Cohen et al., 2018).

3.2.2 Matched-Guise Technique

Like other RLS studies, the present study employed a matched-guise technique. This technique, originally developed by Lambert and colleagues (e.g. Lambert, 1967; Lambert et al., 1960) controls for speaker variables such as voice quality and speaking style. In the original incarnation of matched-guise experiments, bilingual or bidialectal speakers are used to produce speech samples in each language or dialect. In intelligibility studies, especially those investigating RLS, the same speech sample from a single speaker is used (i.e. not in different languages or dialects), while different nonlanguage characteristics, primarily ethnicity or nationality, are ascribed to the speaker in each experimental group (e.g. Eisenclas & Michael, 2019; Hu & Su, 2015; Kang & Rubin, 2009; Rubin, 1992). This is usually accomplished by displaying photographs and personal information, such as name or nationality. In this way, any differences found in the way that listeners react to the speaker can be confidently attributed to the listeners' attitudes towards the targeted nonlanguage characteristics of the speaker and each listener's propensity to RLS (Rubin, 2012). In the present study, both groups heard the same speech sample produced by a NES, but the

ethnicity of the speaker was implied to be Caucasian in the *Caucasian guise* condition, and Asian in the *Asian guise* condition. Ethnicity was implied using photographs and pseudonyms. More details about the speaker, speech sample and guises are given in Section 3.6.

3.2.3 Online Format

The experimental listening task and questionnaire were administered online using Qualtrics (<http://www.qualtrics.com>). Internet-based instruments are now well-established in the field (Dörnyei, 2007) and present a number of advantages. According to Cohen et al. (2018), the online format allows for wider dissemination and therefore access to a larger sample. It also provides convenience for participants, which can encourage more responses. Furthermore, administering the listening task offline would have been a time-consuming process which would have limited participation according to the researcher's logistical ability to schedule listening sessions with participants. Internet-based data collection also facilitates full anonymity for participants, which encourages greater participation as well as honesty, particularly when dealing with sensitive subjects – in this case, rating another person's speech (Cohen et al., 2018). Anonymity for participants was also required by the institution from which participants were originally gathered (see Section 3.3 for more details). Finally, the online format ensured safety for all involved as well as ongoing feasibility during the uncertainty of the COVID-19 pandemic (see Section 3.9 for a discussion of the pandemic's impact on this research).

3.3 Ethical Considerations

All participants were over the age of 18 and volunteered to take part. Participants were initially gathered from the university where the researcher taught, which made full anonymity for participants a condition of research. Participants were informed on the first

page of the survey that their responses would be entirely anonymous. As they were university students invited to take part by their teachers, they may have felt pressured to take part. To ensure that participation was completely voluntary, participants were informed that anonymity included no collection of IP addresses and that their instructors would not know whether they took part. The last page of the survey informed participants of a final opportunity to opt out of the study by closing the browser window rather than clicking the *submit* button. Responses from participants who did not submit were automatically deleted.

Ethical approval for this research was sought from and granted by the Central University Research Ethics Committee (Appendix A).

3.4 Participants

3.4.1 Population

One of the study's aims was to investigate the effect that a speaker's ethnicity might have on students' judgements of their teaching ability (RQ3). People who have already spent time out of education may be more likely to be influenced by experiences they have had since graduating, so their responses may no longer accurately reflect a student's perspective.

Therefore, the participant population was limited to current undergraduate students. As the focus was on the Korean context, international students, who may have significantly different educational and language-learning experiences, were also excluded. The target population was thus undergraduate students over the age of 18, who spoke Korean as an L1. In Korea, English has been a compulsory subject from elementary school since 1997, while English listening tests have been part of the national college entrance exam since 1991 (Choi, 2021).

It is thus a reasonable assumption that most current undergraduate students educated in Korea have received several years of English instruction, including in listening. The population was therefore not limited to those majoring in English, but included students of any subject.

3.4.2 Recruitment

Participants were recruited initially from one low-tier university in South Korea at which the researcher taught. A participant poster (see Appendix B) was uploaded to individual class homepages on the university's learner management system. The researcher drew students' attention to this poster in lecture videos throughout the semester, from October to December 2020, and requested that colleagues in the same department do likewise. However, as initial uptake was limited, the participant pool was expanded in November 2020 to include three other universities, where potential participants with the same inclusion criteria were contacted by teachers of English known to the researcher. More participants were recruited in April 2021 in the same way.

3.4.3 Participant Statistics

Data from 75 Korean L1 undergraduate students were collected. One participant was removed due to not completing any of the listening task. Demographic information of the remaining 74 participants is displayed in Table 1, presented for the whole sample as well as divided for the two experimental conditions. Participants were randomly allocated to one of the two conditions by the Qualtrics software. Although the software was set to do so evenly, the resulting final groups were slightly imbalanced with 33 participants (44.6%) assigned to the Asian guise condition and 41 (55.4%) assigned to the Caucasian guise condition. However, according to Pituch and Stevens (2016), groups can be considered approximately equal if largest/smallest < 1.5, which is the case in this dataset ($41 / 33 = 1.24$).

Table 1*Participant Demographic Information*

Group	<i>n</i>	Age				Sex	
		M	SD	Minimum	Maximum	Male	Female
Asian	33	21.67	2.16	18	26	13 (39.4%)	20 (60.6%)
Caucasian	41	21.76	2.095	18	28	14 (34.1%)	27 (65.9%)
Total	74	21.72	2.11	18	28	27 (36.5%)	47 (63.5%)

3.5 Pilot Study

A small pilot study was conducted to check that the Qualtrics software worked across different devices and to ensure that all instructions and procedures were clear. Participants with the same inclusion criteria as the main study were recruited on a voluntary basis from a class taught by the researcher in August 2020. This class was part of a non-credit programme preparing students at the end of their undergraduate studies for internships abroad. As a result, the participants were slightly older, aged 22 to 26, with a mean age of 24 ($SD = 1.29$). Seven students (three male) completed the listening task and survey, and were asked to complete a feedback form (Appendix C). Based on the pilot study responses and feedback, some small procedural changes were made to the online instrument, which are noted below where relevant.

3.6 Guises and Speech Sample

3.6.1 Speaker

The speaker was a professional NES voice actor and former English teacher from the USA in his early 30s. He spoke with a neutral, standard American accent.

3.6.2 Script

To target participants' reactions to an English language teacher, the present study used a short lecture explaining a language feature. To make sure that the topic was not too technical or difficult to follow, but also less likely to be familiar to listeners such that background knowledge could confound results, it was decided to describe an idiom. The idiom *burn your bridges* was chosen, as it was considered fairly easy to explain but unlikely to be previously known by the initial target participants. All pilot test participants indicated that they did not know the idiom. A script was written in which the meaning of the idiom was explained and three examples of use given (see Appendix D). Cobb's (n.d.) VocabProfiler revealed that 97.8% (n = 227) of words in the script were in the most frequent 2000 word families (K1 and K2 lists) according to the British National Corpus. The word *ex-girlfriend* was counted by the programme as two words, appearing in the seventh most frequent 1000 words (K7 list) and off-list respectively, but the researcher felt, based on personal experience with Korean L1 English learners, that the word *ex-girlfriend* would likely be known by all participants.

3.6.3 Recording and Editing

The speech sample was recorded using professional recording equipment at a local public radio station. Extracts used for intelligibility testing (see Section 3.6.3) were cut using

GarageBand (GarageBand '11, Version 6.0.5 (428.5)). All files were saved and uploaded to Qualtrics in mp3 format.

The full speech sample was 92 seconds long. This was considerably shorter than the speech samples used in previous studies (e.g. 4 minutes and 20 seconds in Eisenchlas and Michael's [2019] study; 4 minutes in Kang and Rubin's [2009] study). The reason for the shorter length was that online data collection instruments run a higher risk of participant dropout, especially if they are 'long or too complicated' (Cohen et al., 2018: p.363). Efforts were therefore made to reduce the length of time it took participants to complete the listening tasks and questionnaire. Furthermore, lecture length has been shown to affect learners' listening comprehension: for example, Carrell et al. (2004) found that comprehension scores were higher when lecture length was shorter (2.5 minutes as opposed to 5 minutes), especially among higher proficiency learners. The speech sample was therefore kept short to limit the difficulty of the task and avoid a floor effect.

3.6.4 Guises

Speaker ethnicity was implied using photographs and names (Appendix E). In the Asian guise condition, a picture of a Korean man captioned with the name *Lee Sung Min* was displayed while the speech sample played. In the Caucasian guise condition, a picture of a Caucasian man captioned with the name *Christopher Johnson* was displayed while the speech sample played. The photographs were taken by the researcher. The two men were chosen to be as similar as possible. Both were in their late 30s, with short dark hair worn in a similar style, and both wore a light shirt with a dark-coloured necktie. The photographs were taken against a white background from the shoulders up. The subjects were asked to look into the camera with a closed-mouth smile. Names were assembled from lists of the most popular names given to boys born in 1980 and the most common family names in each country.

3.7 Variables and Measures

3.7.1 Language Attitudes

Language attitudes were taken from Zahn and Hopper's (1985) Speech Evaluation Instrument (see Section 2.3.1). Dynamism items were removed to preserve only the items relevant to the research questions, leaving two attitude factors of *social attractiveness* and *superiority*. Some items were also removed from these factors because of the potential for them to be confusing when translated into Korean (e.g. *unkind-kind* was removed as there is not a strong difference in meaning between this and *unfriendly-friendly* in Korean) or because the concepts targeted were unlikely to result in differences depending on perceived ethnicity in the context of the study (e.g. *poor-rich* was removed, as it was believed that perceived ethnicity would not be a salient factor in this judgement when, under both conditions, participants believed they were listening to a university professor). Some remaining items were rephrased for clarity or to better fit the context (e.g. *uneducated-educated* was changed to *low level of education-high level of education*, as it was deemed unlikely that participants would believe a university teacher of any ethnicity to be uneducated). In total, seven items were averaged for the superiority factor, which had a Cronbach's alpha of .851, and seven for the social attractiveness factor, with a Cronbach's alpha of .936. Three items measuring superiority (*disfluent-fluent*, *unclear-clear* and *inexperienced-experienced*) were also used as measures of accentedness, comprehensibility and teaching quality respectively. A single item asked participants to indicate whether they believed the speaker to be Asian or Caucasian. All items were rated on a 7-point semantic differential scale. All attitude items can be found in Appendix F.

3.7.2 *Comprehension*

The variable labelled *comprehension* in Rubin and colleagues' RLS studies (e.g. Kang & Rubin, 2009; Rubin, 1992) was measured via cloze tests of full transcripts administered after listening to a lecture. However, as discussed in Section 2.1.2.1, this is closer to measures of intelligibility used in other studies (e.g. Munro & Derwing, 1995), but with the potentially confounding influences of memory and, in the case of NNES listeners, proficiency. Operationalising comprehension in this way was therefore unsuitable for the present study: it was predicted that the participants, as NNESs in an EFL context, were likely to be of generally lower proficiency than those in Rubin and colleagues' RLS studies, who were NESs or NNESs in a native-speaking context. A cloze test of the full speech sample may have been too difficult for many, placing heavy demands on working memory and processing skills, and potentially creating a floor effect. Instead, comprehension and intelligibility were measured separately. The comprehension variable was intended to target participants' understanding of the gist of the speech sample, and was measured by a single question. After listening, participants were asked to choose which of four situations best exemplified a situation to which the idiom presented in the lecture could be applied (Appendix G). The result was a nominal value of *correct* or *incorrect*.

3.7.3 *Intelligibility*

The intelligibility variable targeted participants' ability to process local level, low-context details. Participants completed five cloze tests in which they typed in one missing word from each of five short (5 to 7 seconds) extracts from the lecture (Appendix H). The extracts and missing words were chosen as words which were likely known to all participants, but could not be easily guessed from context without listening to the audio. Cobb's (n.d.) VocabProfiler showed that four of the target words appeared in the British National Corpus's

most frequent 1000 words (K1 list), and the remaining target word appeared in the second most frequent 1000 words (K2 list).

Each of the cloze tests was scored as *correct* or *incorrect*. Criteria for scoring the answers were discussed with the researcher's supervisor and a doctoral student in intelligibility, who offered advice about the advantages and disadvantages of a number of approaches to the scoring system. Based on the guidance of these expert advisers, the researcher first trialled a relatively lenient scoring system. In addition to correctly-spelled answers, answers allowed as correct in this system were:

1. answers with incorrect transcription of vowels, which when transliterated from the Roman to Korean alphabet could be written using the same letter;
2. morphological errors which were semantically and syntactically possible in the context;
3. spelling errors which could result in the same phonemic realisation;
4. homophones.

Scoring criteria and participant answers were then sent to four pilot raters, three of whom were teachers of English (two NESTs and one Korean L1 NNEST) and one of whom was a former interpreter (L1 Italian, L2s English, Korean and Mandarin Chinese). However, the pilot raters remarked that the scoring criteria were far too lenient, and resulted in almost all answers being permissible. The researcher subsequently found that the leniency of this scoring system resulted in a ceiling effect when results were analysed. Therefore, the researcher reverted to the harshest scoring system suggested by the expert advisers, which allowed only:

1. correctly-spelled answers;
2. spelling errors which could result in the same phonemic realisation;
3. homophones.

The final intelligibility score used for analysis was the sum of correct answers (i.e. 0 to 5). The pilot test revealed that it was necessary to differentiate between participants who did not answer the questions due to technical difficulties and those who did not answer due to linguistic limitations. An additional page was therefore displayed if the participant did not answer a cloze test question, which allowed them to indicate why they had not answered (Appendix I). If a participant indicated that they did not answer the question because of technical difficulties, their total intelligibility score was excluded from analysis, as their failure to answer the question was not related to intelligibility. If a participant indicated that their missing response was for any other non-technical reason, their response was coded as *incorrect* and included in the final intelligibility score.

3.7.4 Comprehensibility

Considering the findings of Nagle et al. (2019) that single global comprehensibility ratings may not be sufficient, comprehensibility was measured on the same five 7-point semantic differential items used in Kang (2010) and Kang et al. (2015). Kang et al. (2015) reported an internal consistency Cronbach's alpha of between .93 and .94 for this measure. In the present study, it had a Cronbach's alpha of .943. However, analysis showed that the Cronbach's alpha would rise to .96 without the item *unclear-clear*, possibly because this item may have been interpreted by participants as referring to the overall explanation of the idiom presented in the speech sample rather than to the speech itself. The item was therefore subsequently removed from analysis. All items used in the comprehensibility measure can be found in Appendix J.

3.7.5 Accentedness

Accentedness was measured by five 7-point semantic differential items from Kang and Rubin's (2009) adaptation of Zahn and Hopper's (1985) Speech Evaluation Instrument, with changes of expression (e.g. rephrasing *foreign* to *non-native*) made to better fit the context of an EFL country. The Cronbach's alpha for this measure was .904. All items measuring accentedness can be found in Appendix K.

3.7.6 Teaching quality

Teaching quality was measured on ten 7-point semantic differential items. Some items were assembled from Kang and Rubin's (2009) adaptation of Zahn and Hopper's (1985) Speech Evaluation Instrument, for example, *informal-formal* and *boring-interesting*. Additionally, attributes on which Chun (2014) found Korean university students believed NETs and NNESTs to differ were included, for example, *unqualified-qualified* and *unprepared-prepared*. These items were combined for a single teaching quality score and were also considered separately. However, *informal-formal* was removed from the aggregate score, as the two poles do not necessarily fit a continuum of poor to good teaching. The Cronbach's alpha for the aggregate score was slightly higher with this item removed, increasing from .921 to .924, maintaining an excellent fit of the adapted measure. All teaching quality items can be found in Appendix L.

3.7.7 Listener Background Characteristics

A background questionnaire was used to gather information about the participants, including demographic information such as age and sex (presented in Table 1, above). The questionnaire also gathered the following information for the purpose of answering RQ4.

3.7.7.1 Proficiency. It was not possible to include a test of proficiency in the study, as this would have added significant time to the online task, nor was it expected that all participants would have taken the same external language test. Therefore, participants were asked to report their English proficiency as *beginner*, *intermediate* or *advanced*. In order to test the reliability of this self-reported measure, participants were also asked to provide their most recent score, if they had one, on two standardised English tests common in Korea: the Test of English for International Communication (TOEIC) and the Test of English as a Foreign Language (TOEFL). It was hoped that enough participants would report such scores to enable the reliability of the self-reported measure to be tested. If the reported scores correlated well with participants' self-reported proficiency, it could be cautiously assumed that all participants were generally able to accurately judge their own proficiency. The results of this analysis are presented in Section 4.1.2.

3.7.7.2 Exposure to NESs and NNEs. Eight multiple choice questions targeted the amount of exposure participants had to NESs and NNEs and their experience using the English language. Separate questions were asked about NESs and NNEs. The questions asked for these variables are presented in Appendix M. Although the original intention was to consider these as separate variables to answer RQ4, when these variables were used in two-way, between-groups analyses of variance (ANOVAs) during the analysis phase of the research, the number of cases in some groupings were very low (for example, there were only three cases in the grouping *Caucasian condition, medium frequency of interaction with NESs*). Therefore, the initial exposure variables were combined into two composite variables: exposure to NESs and exposure to NNEs. To calculate the composite variables, the values of each individual variable were relabelled as 1 (*low*), 2 (*medium*) or 3 (*high*). Variables which originally had more than three values were collapsed (see Table 2). The new numerical

values of the variables related to NESs were then added together to create a score with a range of four to 12. Finally, this intermediary continuous variable was collapsed into a final categorical variable with values of *low*, *medium* and *high*, representing overall exposure to NESs. The process was repeated to create a composite variable for exposure to NNESs (see Table 3). It is acknowledged that this scoring system is somewhat arbitrary, especially as categorical values were used to create a continuous composite value. Results using these variables are therefore intended to be treated with caution.

3.7.7.3 Knowledge of L2s Other Than English. Participants were asked to indicate whether they spoke any languages other than Korean and English. This resulted in a nominal value of *yes* or *no*.

Table 2*Recoding of Original Exposure Variable Values*

Variable	Original values	Recoded values
Frequency of interaction with N(N)ESs	Yearly or less often	1 (low)
	Once every few months	2 (medium)
	Monthly	
	Weekly	3 (high)
	Daily	
Length of English classes with N(N)EST	One year or less	1 (low)
	One to five years	2 (medium)
	More than five years	3 (high)
Length of time spent living / traveling in N(N)ES countries (other than Korea)	One month or less	1 (low)
	One month to a year	2 (medium)
	More than one year	3 (high)
Frequency of speaking English in N(N)ES countries (other than Korea)	Never / I've never visited a N(N)ES country (other than Korea)	1 (low)
	Once or twice	2 (medium)
	Some of the time	
	All or most of the time	3 (high)

Table 3*Recoding of Continuous N(N)ES Scores Into Categorical Values*

Sum of Recoded N(N)ES Exposure Scores	Corresponding Categorical Exposure to N(N)ESs Values
4-6	Low
7-9	Medium
10-12	High

3.8 Procedure

After reading the participant information (Appendix N), consenting to take part and confirming that they were over 18, participants were presented with a page providing instructions for the first task. They were asked to test their device's audio before advancing to the next page, using an embedded audio player which played a short extract of piano music. On the following page, the full speech sample played automatically while the photograph and name for the relevant condition were displayed. In order to standardise test conditions, participants were unable to pause, rewind or repeat the audio. The *next page* button was not displayed until after 93 seconds had elapsed to prevent participants from skipping this page. The page auto-advanced after 100 seconds and participants were unable to navigate back to the speech sample without fully exiting and restarting the questionnaire.

Immediately after listening, participants answered the global comprehension question. They then completed the speaker ratings. Items measuring accentedness and comprehensibility were displayed together. Teaching quality was then rated, followed by attitudes. Items were programmed to display in a random order in each section to avoid any order effect.

After the rating procedure, participants were given instructions for the cloze tests. They were asked again to check their audio and were informed of the time limit for each question. The five cloze questions were then presented on separate pages with the relevant photograph and name for each condition displayed. As the extracts were very short and easily missed if participants were not ready, participants were given control of the audio and the ability to pause or replay each extract. In order to standardise test conditions for all participants as much as possible, the survey was programmed to display each cloze test page for 20 seconds only before auto-advancing to the next question. If participants did not enter

an answer, they were prompted to state why. Lastly, the participants answered questions about their language background.

The questionnaire was presented in Korean as the default language to ensure students had access to all instructions in their L1, but they had an option to switch to English. This included the multiple choice answers to the global comprehension question. Only the extracts used in the cloze test were not available in Korean. A professional Korean L1 translator was hired to translate the survey, and the translation was checked by the researcher, an L2 speaker of Korean. The questionnaire was completed in Korean by 73 participants (97.3%). The full procedure took participants five to ten minutes to complete.

3.9 Limitations and COVID-19 Impact

3.9.1 Uneven Groups

There was a slightly imbalanced number of participants in the two conditions. Due to the anonymity and consent requirements of the initial university from which participants were gathered, incomplete responses were automatically deleted. It was therefore not possible to see how many responses were started and then abandoned in each condition, or whether the lower number of completed responses in the Asian guise condition was due to chance or an aspect of the condition that encouraged dropout. This imbalance could be because the software allowed one week from the survey start time for each response to be completed before deletion. Therefore, even if conditions were equally assigned during this time, the number of incomplete responses from each condition which were subsequently deleted by the software may have been unequal. This outcome could have been avoided or minimised by setting a shorter response time window. The shortest time allowed by Qualtrics is one hour, which would have been sufficient for the purposes of this survey.

3.9.2 Sampling

Unlike in Kang and Rubin's (2009) study, there was no scope in the present study to deliberately sample participants for a range of background characteristics. Furthermore, institutional rules about participant anonymity at the university where the researcher was employed did not allow for purposive sampling. Therefore, there may be some self-selection bias (Dörnyei, 2007).

3.9.3 Sample Size

With only 74 complete responses, the sample size was smaller than expected. This may have negatively affected the power to reveal small effects in analysis. A larger sample size could be achieved by administering the tasks and survey in person during class time, but this was prevented in the present study by institutional rules regarding full anonymity. Alternatively, time during class could have been provided for students to complete the online task and survey, but in-person, synchronous classes were cancelled due to the COVID-19 pandemic. The pandemic may also have reduced prospective participants' willingness to take part. By the time the first responses were gathered in October 2020, most universities in South Korea, including the four targeted in this study, had provided almost entirely online learning for more than seven months. Prospective participants may therefore have been fatigued with online education and less willing to volunteer their time for another online activity. Anecdotal comparisons with previous sample sizes achieved in online research carried out by the researcher's colleagues with the same population suggests this was the case.

3.9.4 Online Format

The online format presents a number of other limitations, as identified by Cohen et al. (2018), including lack of control over experimental conditions, no way to prevent multiple

submissions, and no supervision of participants to ensure the tasks were completed alone. In particular, obscuring the matched-guise technique from the participants was difficult with an online, asynchronous procedure. Participants who restarted the survey may have been randomly assigned to the other condition, or participants may have compared their condition with their classmates. As detailed above, steps were taken to assume as much experimental control as possible, such as setting time limits for cloze tests and removing listener control for the full speech sample. Finally, technological issues such as audio failure or device incompatibility may have restricted participants' performance on listening tasks, contributed to abandoned responses, or affected participants' rating behaviour.

Chapter 4: Results

This section reports the results of the online experimental task and questionnaire. First, the results of preliminary analyses to establish the equality of the experimental groups, the reliability of the self-reported proficiency measure and the efficacy of the matched-guise technique are presented. Next, the main research questions are addressed. All data were analysed in SPSS (IBM SPSS Statistics, Version 27). Descriptive statistics are presented in Appendix O.

4.1 Preliminary Analyses

4.1.1 Equality of Groups

Levene's test indicated that the variances for age were equal, $F(1, 72) = 0.46, p = .499$. A chi-square test for independence showed no significant association between condition and self-reported English proficiency, $\chi^2(2, n = 74) = 1.60, p = .449$, Cramer's $V = .15$. Similarly, chi-square tests for independence showed no significant association between condition and exposure to NESs, $\chi^2(2, n = 73) = 1.66, p = .436$, Cramer's $V = .15$, nor between condition and exposure to NNESs, $\chi^2(2, n = 74) = 0.74, p = .691$, Cramer's $V = .10$.

4.1.2 Reliability of Self-Reported Proficiency

All 74 participants self-reported their English proficiency, 26 reported a TOEIC score, and seven reported a TOEFL score. Additionally, one participant wrote in an Oral Proficiency Interview Computer (OPIc) test score. For analysis, the test scores were converted to three categories (A, B and C, based on the Central European Framework of Reference for Languages [Council of Europe, 2001]), using a conversion table provided by

the European Consortium for Accreditation

(http://ecahe.eu/w/index.php/English_language_test_equivalency_table).

Tables 4 and 5 show the frequencies of self-reported proficiency and test score conversions respectively. No participants who reported test scores fell into the A category, with scores instead skewed heavily towards the higher categories. In contrast, the self-reported proficiency results show a more normal distribution. As the Spearman rank correlation coefficient confirmed that there was no significant relationship between self-reported proficiency and test scores ($r_s = .22, p = .25$), it was concluded that there was not enough evidence of the reliability of participants' self-reported English proficiency. Therefore, further analyses using the proficiency data were not carried out, as the reliability of results could not be confidently assumed.

Table 4

Self-Reported Proficiency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Beginner	20	27	27	27
	Intermediate	37	50	50	77
	Advanced	17	23	23	100
	Total	74	100	100	

Table 5*CEFR Conversion*

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	B	9	12.2	31	31
	C	20	27	69	100
	Total	29	39.2	100	100
Missing	System	45	60.8		
Total		74	100		

4.1.3 Efficacy of Matched-Guise Technique

In order to establish whether the matched-guise technique was successful, i.e. whether participants believed that the person in the displayed photograph was the speaker they were hearing, an independent-samples t-test was conducted to compare ethnicity ratings by participants in the two conditions. There was a significant difference for participants in the Asian condition ($M = 4.88$, $SD = 1.73$) and the Caucasian condition ($M = 2.29$, $SD = 1.10$; $t(51.9) = 7.47$, $p < .001$, two-tailed). Levene's test indicated unequal variances ($F = 12.84$, $p < .001$), so degrees of freedom were adjusted from 72 to 51.9. The effect size was large ($d = 1.83$). However, the distribution of data was not normal in either condition. Although t-tests are considered robust to violations of the assumption of normal distribution (Laerd Statistics, 2015), a Mann-Whitney U (MWU) test was conducted to be conservative (see Section 4.3 for justification of this decision). This test similarly revealed a significant difference with a large effect size between ethnicity ratings in the Asian condition ($Md = 4$, $n = 33$) and the Caucasian condition ($Md = 2$, $n = 41$), $U = 150$, $z = -5.83$, $p < .001$, $r = -.68$.

These results indicate that participants in the Caucasian condition were much more likely to perceive the speaker as being Caucasian, while those in the Asian condition were much more likely to perceive the speaker as being Asian. Therefore, the matched-guise technique was successful. It can be assumed that participants in the two conditions believed that they were listening to speakers of the ethnicities implied by the photographs displayed.

4.2 Relationships Between Variables

The relationships between the seven dependent variables (social attractiveness, superiority, comprehension, intelligibility, comprehensibility, accentedness, and teaching quality) as well as the ethnicity rating, before division into the two conditions, were explored using the Pearson product-moment correlation coefficient. Table 6 shows the results of this analysis. Large effect sizes ($r > .50$) are highlighted in black. Medium effect sizes ($r > .30$) are highlighted in dark grey. Small effect sizes ($r > .10$) are highlighted in light grey.

Of particular note, participants who rated the speaker highly on superiority also rated him highly on comprehensibility ($r = .53, p < .01$), accentedness ($r = .66, p < .01$) and teaching quality ($r = .73, p < .01$). It is perhaps not surprising that participants who evaluated the speaker highly for teaching quality also found him more comprehensible ($r = .67, p < .01$). There were also strong associations between teaching quality on the one hand, and social attractiveness ($r = .74, p < .01$) and superiority ($r = .73, p < .01$) on the other. Accentedness had a medium-strength correlation with social attractiveness ($r = .47, p < .01$) and with teaching quality ($r = .44, p < .01$). Perceived ethnicity had no significant relationships with any variable except, notably, intelligibility, with which it was negatively correlated with a medium effect size ($r = -.42, p < .01$). This indicates that participants who perceived the speaker to be more Asian had lower scores on the intelligibility measures.

Table 6*Correlations*

Variable	1	2	3	4	5	6	7	8
1 Attractiveness								
2 Superiority	.76**							
3 Comprehension	.20	.22						
4 Intelligibility	-.02	.08	.34**					
5 Comprehensibility	.45**	.53**	.56**	.31**				
6 Accentedness	.47**	.66**	.16	.25*	.33**			
7 Teaching quality	.74**	.73**	.37**	.11	.67**	.44**		
8 Ethnicity	-.08	-.12	-.16	-.42**	-.14	-.22	-.07	

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

* Correlation is significant at the 0.05 level (2-tailed).

4.3 Effects of Speaker Ethnicity on Dependent Variables

Independent-samples t-tests were conducted to compare the mean ratings on each dependent variable in the two conditions. Assumptions were tested in all cases (Appendix P). Although some of the dependent variables were not normally distributed, the independent-samples t-test was considered robust enough to still provide reliable results (e.g. Pallant, 2016; Laerd Statistics, 2015). However, to be conservative, MWU tests were also conducted where assumptions were violated. In all cases, the non-parametric tests returned the same results as the parametric equivalent in terms of significance (or lack thereof), but in some cases where significance was reached, the effect sizes differed. In the following section, the MWU results are reported only for significant results. Statistics for the independent-samples t-tests are presented in Appendix Q. Results of the t-tests are presented graphically in Appendix S.

4.3.1 Effect of Speaker Ethnicity on Social Attractiveness Ratings

An independent-samples t-test was conducted to compare the mean social attractiveness ratings of participants in the two conditions. There was no significant difference for participants in the Asian condition ($M = 5.42, SD = 1.06$) and the Caucasian condition ($M = 5.48, SD = 0.99; t(72) = -0.24, p = .811$, two-tailed), indicating that speaker ethnicity did not affect listeners' ratings of the speaker's social attractiveness.

4.3.2 Effect of Speaker Ethnicity on Superiority Ratings

An independent-samples t-test was conducted to compare the mean superiority ratings of participants in the two conditions. There was no significant difference for participants in the Asian condition ($M = 5.80, SD = 0.86$) and the Caucasian condition ($M = 5.75, SD = 0.79; t(72) = 0.23, p = .820$, two-tailed). Speaker ethnicity was not shown to affect listeners' superiority ratings.

4.3.3 Effect of Speaker Ethnicity on Comprehension

A chi-square test for independence (with Yates' Continuity Correction) was conducted between speaker ethnicity and comprehension. All expected cell frequencies were greater than five. The test indicated no significant association between condition and comprehension, $\chi^2(1, n = 74) = .22, p = .640, \phi = .08$.

4.3.4 Effect of Speaker Ethnicity on Intelligibility

An independent-samples t-test was conducted to compare intelligibility scores of participants in the two conditions. For the intelligibility variable, cases were removed if

participants indicated that they had not answered a question due to audio problems. Therefore, the number of cases in this analysis was smaller than in other analyses (Asian: $n = 32$; Caucasian: $n = 39$). The t-test revealed a significant difference between participants in the Asian condition ($M = 3.84$, $SD = 1.22$) and the Caucasian condition ($M = 4.46$, $SD = 0.76$; $t(49.55) = -2.50$, $p = .016$, two-tailed). Levene's test indicated unequal variances ($F = 4.80$, $p = .032$), so degrees of freedom were adjusted from 69 to 49.55. The effect size was medium ($d = -.62$). A MWU test also revealed a significant difference between intelligibility scores in the Asian condition ($Md = 4.00$, $n = 32$) and the Caucasian condition ($Md = 5.00$, $n = 39$), $U = 808.50$, $z = 2.32$, $p = .021$, but the effect size was small ($r = .28$). Both tests showed that participants who saw a Caucasian face had higher overall intelligibility scores than participants who saw an Asian face.

4.3.5 Effect of Speaker Ethnicity on Comprehensibility Ratings

An independent-samples t-test was conducted to compare comprehensibility ratings of participants in the two conditions. There was no significant difference for participants in the Asian condition ($M = 5.42$, $SD = 1.48$) and the Caucasian condition ($M = 5.44$, $SD = 1.81$; $t(72) = -0.06$, $p = .955$, two-tailed).

4.3.6 Effect of Speaker Ethnicity on Accentedness Ratings

An independent-samples t-test was conducted to compare the accentedness ratings of participants in the two conditions. This revealed a significant difference for participants in the Asian condition ($M = 6.15$, $SD = 0.93$) and the Caucasian condition ($M = 6.65$, $SD = 0.58$; $t(51.40) = -2.72$, $p = .009$, two-tailed). Levene's test indicated unequal variances ($F = 10.51$, $p = .002$), so degrees of freedom were adjusted from 72 to 51.40. The effect size was medium ($d = -0.67$). This indicates that participants who saw a Caucasian face rated the speaker as

having a more native-like accent than participants who saw an Asian face. A MWU test also showed a significant difference between accentedness ratings in the Asian condition ($Md = 6.40$, $n = 33$) and the Caucasian condition ($Md = 7.00$, $n = 41$), $U = 889.50$, $z = 2.49$, $p = .013$, but with a small effect size, bordering on medium ($r = .29$).

4.3.7 Effects of Speaker Ethnicity on Teaching Quality Ratings

An independent-samples t-test was conducted to compare the teaching quality ratings of participants in the two conditions. There was no significant difference for participants in the Asian condition ($M = 5.65$, $SD = 0.90$) and the Caucasian condition ($M = 5.63$, $SD = 1.09$; $t(72) = 0.08$, $p = .938$, two-tailed).

In order to examine whether participants judged the speaker's teaching quality differently depending on ethnicity along the lines found by Chun (2014), independent-samples t-tests were also run on each separate item that made up the teaching quality composite variable. No results were significant, either before or after the application of a Bonferroni correction, indicating that participants did not judge the speaker's teaching abilities differently depending on whether they thought he was Asian or Caucasian. Statistics for these independent-samples t-tests are presented in Appendix R.

4.4 Interaction Effects of Listener Background Characteristics and Ethnicity on Dependent Variables

As discussed in Section 4.1.2, confidence in the reliability of participants' self-reported English proficiency was too low for any investigation into the impact of English proficiency on any of the dependent variables. The remaining listener background characteristics which were of interest were those regarding experience of L2s other than English, and exposure to NESs and NNESSs. To determine whether any of the above effects

were mediated by such exposure, a number of two-way, between-groups ANOVAs were conducted. As acknowledged in Section 3.7.7.2, scoring for the final exposure variables is somewhat arbitrary, and the limited sample size still leads to smaller than ideal numbers of cases in some of the groupings in the following ANOVA analyses. The results should therefore be treated with caution, as validity is questionable and the possibility of Type II errors high. They are presented for transparency nonetheless.

4.4.1 Interaction Between Condition and Knowledge of L2s Other Than English

No statistically significant interaction effects of condition and knowledge of L2s other than English were found on any of the dependent variables, indicating that knowledge of other L2s did not mediate how listeners responded to the speaker depending on ethnicity. No main effects were found for knowledge of L2s other than English. Statistics for these analyses can be found in Appendix T.

4.4.2 Interaction Between Condition and Exposure to NESs

No statistically significant interaction effects of condition and exposure to NESs were found on any of the dependent variables. In other words, exposure to NESs was not found to affect participants' ratings of the speaker differently depending on ethnicity. Statistics for these analyses can be found in Appendix U.

Some main effects were found for exposure to NESs. As these effects were not related to speaker ethnicity, they are not the focus of this study, but are reported here for transparency. Exposure to NESs was found to affect superiority ratings, $F(2, 67) = 3.46$, $p = .037$. Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the low exposure to NESs group ($M = 6.21$, $SD = 0.62$) differed significantly from the medium exposure group ($M = 5.60$, $SD = 0.81$). The effect size was medium (partial eta squared = .09).

Exposure to NESs also affected comprehensibility ratings, $F(2, 67) = 3.19, p = .048$, with a medium effect size (partial eta squared = .09). Post-hoc comparisons using the Tukey HSD test indicated that the mean score for the medium exposure to NESs group ($M = 5.12, SD = 1.66$) differed significantly from the mean score for the high exposure to NESs group ($M = 6.38, SD = 1.13$).

Some main effects were also found for condition. There was a statistically significant main effect of condition on intelligibility, $F(1, 64) = 8.12, p = .006$, with a medium effect size (partial eta squared = .113). This aligns with the finding of the independent-samples t-test discussed in Section 4.3.4. Unlike the independent-samples t-test reported in Section 4.3.6, no main effect for condition on accentedness was found, $F(1, 67) = 4.15, p = .046$, as the alpha value was set at .01 (Pallant, 2016) due to Levene's test indicating that the variance of accentedness ratings across the groups in the ANOVA was unequal ($F = 2.74, p = .026$).

4.4.3 Interaction Between Condition and Exposure to NNEs

There were also no statistically significant interaction effects of condition and exposure to NNEs on any of the dependent variables, suggesting that listeners' exposure to NNEs did not cause any difference in how they responded to the speaker depending on ethnicity. No main effects were found for exposure to NNEs. Statistics for these analyses can be found in Appendix V.

4.5 Summary of Results

To summarise, it was found that:

1. speaker ethnicity did not affect listener ratings of speaker superiority or social attractiveness, but both these attitude variables were associated with ratings of comprehensibility, accentedness, and teaching quality;

2. speaker ethnicity did not affect listeners' overall comprehension of a message;
3. listeners performed less well on measures of intelligibility if the speaker was implied to be Asian than if the speaker was implied to be Caucasian;
4. listeners rated the speaker to have a less native-like accent if the speaker was implied to be Asian than if he was implied to be Caucasian;
5. speaker ethnicity did not affect listeners' evaluations of the speaker's teaching quality, but listeners who rated the speaker as having a less native-like accent also tended to give lower teaching quality ratings;
6. listeners' knowledge of L2s other than English did not mediate any effects of ethnicity on the dependent variables;
7. the extent of listeners' exposure to either NESs or NNESs also did not mediate any effects of ethnicity.

Chapter 5: Discussion

In this chapter, findings are discussed in relation to each of the research questions, and compared to previous findings in the literature discussed in Chapter 2. Finally, another theory is outlined as an alternative to RLS to explain the present study's findings.

5.1 RQ1: Does Speaker Ethnicity Affect Listener Attitudes Towards the Speaker?

Listener attitudes towards the speaker did not differ depending on whether the listeners saw an Asian or a Caucasian face, nor was there any correlation between ethnicity ratings and social attractiveness or superiority ratings. It did not appear, therefore, that the listeners considered ethnicity as a factor in determining social attractiveness or superiority. Eisenclas and Michael (2019) similarly found no significant effects of speaker ethnicity on listener ratings of the speaker's personal characteristics, with the exception of a marginally significant ($p = .048$) finding that listeners judged the Asian guise to be slightly less sincere, although their participants rated the speaker's personality traits on different items. Eisenclas and Michael's study and the present study suggest that listeners in general do not base character judgements of a speaker on the speaker's ethnicity. The present study's findings differed from those of Chun (2014), who found that Korean university students rated NESTs as having superior status to Korean NNESTs. However, see Section 5.3 for a discussion of potential reasons why Chun's findings were not borne out by the present study.

5.2 RQ2: Does Speaker Ethnicity Affect Listener Measures of Comprehension, Intelligibility, Comprehensibility, and Accentedness?

5.2.1 Comprehension and Intelligibility

Listeners who saw a Caucasian face showed higher intelligibility scores than those who saw an Asian face, when intelligibility was measured by the listeners' transcription of a

single word from short extracts of speech. Similarly, when participants were analysed as a single group, there was a correlation between ethnicity ratings and intelligibility, with higher intelligibility associated with the speaker being perceived as more Caucasian. However, listeners' overall comprehension of a longer extract of speech, as measured by their ability to correctly identify a situation to which the idiom explained in the speech could be applied, was not affected by speaker ethnicity. Ethnicity ratings for the entire sample also did not correlate with comprehension.

The findings regarding intelligibility are comparable to the findings of Rubin (1992) and Kang and Rubin (2009) regarding their comprehension measure. As discussed in Section 3.7.2, the measure of comprehension used in both these studies was more comparable to the present study's intelligibility measure, though it also shares aspects of the present study's comprehension measure. In Rubin's (1992) first study, participants who saw an Asian face while listening to a NEST understood less of the speech than those who saw a Caucasian face. In Kang and Rubin's (2009) within-subjects design, participants with biases against Asian English speakers also understood less when they saw an Asian face than when they saw a Caucasian face. Eisenchlas and Michael's (2019) comprehension measure was closer in concept to the present study's comprehension measure, though in their case participants answered six comprehension questions rather than one. Eisenchlas and Michael also found no effect of implied speaker ethnicity on comprehension. None of the three previous studies discussed here measured both intelligibility and comprehension as operationalised in the present study. The present study may therefore provide some confirmation and bridging between Rubin (1992) and Kang and Rubin's (2009) findings on the one hand, and Eisenchlas and Michael's (2019) findings on the other. The cumulative evidence suggests that listener biases against Asian English speakers may manifest in intelligibility loss at a local level, but may not impede overall global understanding of a message.

It is possible that listeners expect, and indeed may experience, listening difficulties when they believe a speaker to be Asian, but that these difficulties are only significant enough to reduce performance when the demand level of the task is higher. The belief that they are listening to an Asian English speaker may lead listeners to feel more unsure about what they are hearing when their attention is focused on low-context, specific details, with the more difficult cloze test transcription task more affected by listeners' biases about their comprehension abilities (or the speaker's communication abilities). This theory is supported by Kennedy and Trofimovich's (2008) finding that, when listening to NNESs, intelligibility scores are higher when more semantic context is provided.

This effect may be explained by Yi et al.'s (2014) finding that hearing a Korean NNES accent placed heavier demands on the neural processing of NES listeners than when they heard an American NES accent, and that this effect was heightened when the listeners saw an Asian face alongside Korean-accented speech, especially when listeners were found to have a stronger bias towards associating Asian faces with foreignness. While Yi et al. did not investigate whether speaker ethnicity alone, i.e. decoupled from a non-native accent, affects neural processing demands, it can be hypothesised that listeners may experience increased effort when they believe they are listening to a NNES based on the speaker's ethnicity, regardless of the speaker's actual accent. In the present study, the greater processing capacity required may not have been significant enough to affect the outcome of the less cognitively demanding task of global comprehension, but may have led to reduced performance in the more demanding task of low-context cloze test transcription.

Alternatively, the different findings for comprehension and intelligibility in this study may indicate a ceiling effect for comprehension caused by participants' previous knowledge of the idiom, its meaning, and its possible applications. Although this was explored in the pilot study, which revealed that none of the pilot study participants had previous knowledge

of the idiom, the subsequent expansion of the participant pool may have significantly raised the average proficiency of participants beyond that of the pilot study participants and of the original institution from which main study participants were intended to be gathered. The higher relative positions of the latterly-included universities on university ranking lists such as the *Times Higher Education Asia University Rankings* (2020) suggest that this may be the case. A comprehension question that could only be answered by listening to the information in the audio lecture may have yielded a different result.

5.2.2 *Accentedness*

Speaker ethnicity was found to affect accentedness ratings. Participants who saw an Asian face were more likely to rate the speaker as sounding less native-like. This supports the findings of Rubin (1992) and Babel and Russell (2015), who both found that speakers of Asian ethnicity were judged to have more non-native accents than Caucasian speakers. Buckingham (2014) similarly found that speakers were rated as having better pronunciation when they were presented as NESTs than when they were presented as NNESTs.

At least two interpretations of this relationship between ethnicity and accentedness are possible. The first is that listeners who believe a speaker to be Asian may be more likely to ‘hear’ more non-native accented speech, regardless of the actual native-speaker status of the speaker. The second is that some listeners expect a speaker they believe to be Asian to have a less native-like accent and rate them accordingly, regardless of what they actually hear. This may be because their expectations override what they hear (i.e., in this case, they were aware of hearing a native accent but their rating behaviour was nevertheless influenced by their expectations of hearing a non-native accent), or because they are not good judges of accent in the first place, and instead rely on their expectations when directed to make judgements (i.e., in this case, they were unable to confidently discern whether the speaker was a NES or NNEST).

from the audio alone, and therefore used his implied ethnicity to make their rating decisions). That listeners are not necessarily able to discriminate between accents is supported by Rubin and Smith's (1990) finding that NES listeners were unable to distinguish in their accentedness ratings between moderately- and highly-accented Chinese English.

According to a study by Zheng and Samuel (2017), which found that differences in accentedness judgements in matched-guise studies can be ascribed to decision-level effects, the hypothesis that listeners interpret the speech differently rather than perceive it differently depending on ethnicity is more likely to be correct. In a series of six experiments, Zheng and Samuel were able to show that the effects of the speaker's implied ethnicity on accentedness ratings could be mitigated by adjusting visual presentations of the speakers and other experimental conditions. The researchers concluded that listeners may judge speech apparently coming from an Asian speaker as more accented because of their interpretation of the input, rather than online perception of the input itself.

Kang and Rubin (2009) found a connection between accentedness ratings and attitudes, where listeners with a higher propensity to superiority RLS judged the Asian guise to sound less native-like. Although the present study did not explore RLS as a listener background variable, there was a large correlation between accentedness and both superiority and social attractiveness ratings, indicating that listeners who rated the speaker highly on the attitude measures were also likely to rate him as having a native-like accent. However, as discussed in Section 5.1, speaker guise did not affect attitude ratings, nor was there any correlation between attitude and ethnicity ratings as a whole. The nature of the relationship between accentedness and attitude ratings is therefore unclear. Possibly, accentedness is a more significant determiner of attitude than ethnicity – believing a speaker to have a non-native accent may reduce some listeners' opinion of that speaker's superiority and social attractiveness, regardless of ethnicity. In the present study's context, this may be linked to the

‘vicarious shame’ that Hendriks et al. (2021: p.20) propose as a source of NNESs’ negative evaluations of the speech of NNESs with whom they share an L1. Overall, the findings related to accentedness suggest that listeners’ conceptualisation of ethnicity and native-speaker status are closely linked, as proposed by Shuck (2006).

5.2.3 Comprehensibility

Listeners who saw an Asian face did not rate the speaker as less comprehensible than those who saw a Caucasian face. This is an interesting finding when considered in light of the findings that speaker ethnicity *did* affect intelligibility scores and accentedness ratings. These results indicate that, although seeing an Asian face may lead to lower intelligibility at a local, low-context level for some listeners, this does not significantly affect their conscious, subjective beliefs about how easily they can understand the speaker overall. This also supports the interpretation that no difference in comprehension scores was found between the two conditions because perceived ethnicity does not significantly affect global or gist understanding of a message. The present study differed from previous matched-guise RLS studies in including a measure of subjective comprehensibility, so results regarding this construct cannot be directly compared. However, Hendriks et al. (2021) found that Dutch L1 NNES listeners rated other moderately-accented Dutch L1 NNESs as less comprehensible than did non-Dutch L1 NNES listeners or NES listeners, but that there was no difference in ratings for slightly-accented Dutch L1 NNESs. The present study supports these findings, suggesting that the ‘vicarious shame’ (2021: p.20) proposed by Hendriks et al. as the reason for harsher ratings of speakers who share an L1 only affects comprehensibility ratings when actual accentedness is higher.

5.3 RQ3: Does Speaker Ethnicity Affect Listener Judgements of Teaching Quality?

Listener judgements of teaching quality were not affected by whether they saw an Asian or a Caucasian face, nor was there any relationship between ethnicity ratings and teaching quality. However, teaching quality correlated with superiority and social attractiveness ratings, and with accentedness (listeners who judged the speaker to sound more native-like also rated them as a better teacher). These findings reflect Rubin's (1992) first study's finding that teaching qualification ratings did not have a significant relationship with the speaker's implied ethnicity, and his second study's findings that stronger foreign accent ratings correlated with lower ratings of teacher effectiveness. These results suggest that ethnicity itself is not a factor in listeners' teaching ratings, but their beliefs about the speaker's native-speaker status are. Again, this may not be the case on an individual level – Kang and Rubin (2009) found that individuals with a higher propensity to either type of RLS were more likely to rate the Asian guise as being a poorer teacher.

The participants in the present study did not demonstrate that they viewed either guise as having the perceived strengths and weaknesses found by Chun (2014) in her study into Korean university students' beliefs about NESTs and NNESTs. A major difference between Chun's study and this study is that Chun did not ask her participants to rate a specific teacher, but rather NESTs and Korean NNESTs in general. Essentially, Chun's participants were prompted to stereotype. The results of the present study suggest that students are capable of evaluating teachers individually, without judging them through the lens of ethnicity-based stereotypes. This may not, however, indicate that students do not revert to stereotypes when thinking about teachers of different ethnicities in the abstract. More research is necessary to explore whether Korean university students still hold stereotypical beliefs about NESTs and Korean NNESTs, when these stereotypes come into play, and what the effects may be on both students and teachers.

5.4 RQ4: Are Any of the Above Effects Mediated by Listener Background

Characteristics?

Listener background characteristics were not found to affect the relationship between speaker ethnicity and any of the dependent variables. As the participants' self-reported English proficiency was not found to be a reliable measure of proficiency, only knowledge of other L2s other than English and the composite variables of exposure to NESs and NNESs were considered in the analysis. It is likely that the small sample size in this study may have increased the risk of Type II errors. The results of these analyses should therefore be treated with extreme caution.

An absence of effect of exposure to NNESs on ratings depending on ethnicity would be in line with Kang and Rubin's (2009) finding that contact with NNESs did not affect teaching quality ratings, accent standardness ratings, or listening comprehension. If reliable, the present study's results may suggest that while English learners' contact with other NNESs does not reduce stereotypes, contact with NESs may not increase them. It may be the case that any associations between ethnicity, native-speaker status, and the variables examined in this study are entrenched on a deeper level than could be affected by the amount of exposure the participants in this study had to NESs or NNESs. The present study's findings that listeners' knowledge of other L2s also did not affect ratings of the speaker contradicts Kang and Rubin's (2009) finding that listeners' linguistic sophistication correlated with better comprehension. However, Eisenclas and Michael (2019) also found no effect for knowledge of other languages among their participants. The small sample size of the present study and the binary yes/no nature of this particular variable prevents any reliable contribution to or clarification of these previous studies' contradictory findings.

5.5 An Alternative to RLS

A related concept is proposed by some researchers as an alternative to RLS and bias-based theories, which would also explain the findings in the present study, but based on different psychological mechanisms. Variously described as an *exemplar* model (McGowan, 2015), a *usage-based* model (Gnevsheva, 2018), and an *expectation-based* model (Hanulíková, 2018), the theory proposes that word recognition is primed by a listener's expectations about a speaker's pronunciation. This has a facilitating effect on comprehension when those expectations are met, but an inhibiting effect when they are not. In the case of ethnicity, listeners' comprehension may be negatively affected if they expect to hear non-native speech from an Asian speaker, but instead hear native speech, and vice versa. For example, McGowan (2015) found that listeners were more accurate in transcribing Chinese-accented speech in noise when they saw a photograph of an Asian face than when they saw a photograph of a Caucasian face. Similarly, ratings of accentedness may be exaggerated when the speaker's ethnicity does not match listeners' expectations, as found by Gnevsheva (2018), in whose study Caucasian German L1 NNESs were rated by NES listeners as significantly more accented in an audio-visual condition than in an audio-only condition, while there was no significant difference in accentedness ratings across conditions for Korean L1 NNESs.

This expectation theory would explain the present study's findings that listeners found the Asian guise less intelligible and more accented as an effect of hearing native-accented speech when they expected non-native-accented speech. According to this explanation, if the speech sample had been produced by a NNES, listeners may have found the Caucasian guise less intelligible.

Chapter 6: Conclusion

This chapter first summarises the aims and findings of the study. Implications are then discussed, followed by limitations and suggestions for improvement. Finally, additional recommendations for future research are proposed.

6.1 Summary of Findings

The study aimed to investigate whether NNES Korean university students' judgements of a speaker were affected by the speaker's implied ethnicity. Specifically, their attitudes towards the speaker's personal characteristics, their ability to understand him, their judgements of his intelligibility, comprehensibility and accentedness, and how they rated him as a teacher were measured in a between-subjects experiment with a matched-guise technique. The study found that the students did not have differing attitudes towards the same NES speaker depending on whether they believed he was Asian or Caucasian. They did not base teaching quality judgements on his perceived ethnicity, nor did they believe him to be less comprehensible if they believed him to be Asian. Their comprehension of his overall message was also not affected by his implied ethnicity. This much is positive news for Asian English speakers and teachers, as it suggests that Korean university students' listening and evaluating behaviours are not affected by stereotypes in these areas. However, the participants did judge the speaker to sound less native-like when his ethnicity was implied to be Asian. Furthermore, participants who saw an Asian face performed worse on a low-context cloze test transcription than those who saw a Caucasian face, suggesting that they found the Asian guise less intelligible. In general, participants who rated the speaker as having a less native-like accent also gave him lower teaching quality ratings, and rated him lower on measures of social attractiveness and superiority.

6.2 Implications

6.2.1 Listening Training

Many studies in the field of intelligibility (e.g. Barrass et al., 2020; Kennedy & Trofimovich, 2008; Zielinski, 2008) make pedagogical recommendations for pronunciation teaching, based on the idea that intelligibility issues can be addressed by further effort on the part of NNEs and their teachers to improve NNEs' pronunciation. However, this study adds to the body of work that demonstrates that intelligibility is not only a speaker characteristic, and that some intelligibility issues arise from listener characteristics that are beyond the control of the speaker. This is not to suggest that pronunciation training does not have its place in the EFL classroom, but that it must be more widely recognised by teachers, English users, learners, and researchers, that the communication burden does not rest solely on the speaker and their ability to make themselves understood, but also on the listener and their ability to understand. As such, listening training for NNEs should take into account the fact that their future interlocutors will likely include other NNEs with accents that differ from NESs. While Barrass et al. (2020) propose pronunciation teaching for speakers that targets phonological features that will increase intelligibility for NNEs' most common NNEs interlocutors, it may also be valuable for listening training to target phonological features that will help NNEs listeners to understand their most common NNEs interlocutors.

6.2.2 Addressing Listener Expectations

Teachers can also encourage recognition among learners that ethnicity is not a marker of native-speaker status. Addressing often misleading expectations about what accent a speaker will use may improve communication for both listener and speaker. One way to address such expectations may be through the use of social interventions. However, interventions must be carefully constructed, or they may run the risk of being

counterproductive. Rubin's (1992) third study reported a pilot intervention in which NES undergraduate students were assigned to evaluate and give feedback to NNES Asian instructors after observing two of their classes. Pre- and post-tests revealed that there was no significant change in the NES participants' response to NNES instructors. Rubin reports that anecdotal evidence suggests some participants became more critical of NNES instructors during the intervention. While Rubin blames the lack of success on the intervention being too short, it is also possible that the intervention cast the undergraduate students in a position of authority over their instructors by virtue of their ethnicity and native-speaker status, which could have exacerbated existing biases against the professional competence and social status of the NNES Asian instructors. As the present study found that lower superiority ratings were associated with lower teaching quality, comprehensibility and accentedness ratings, an intervention may be more successful if it focuses on raising the perceived superiority of Asian instructors of English. To target listener expectations regarding the association between ethnicity, native-speaker status and accentedness, interventions should emphasise the equal status both of non-Caucasian NESs with Caucasian NESs and of NNESs with NESs. Exposure to and normalisation of different varieties of English, both native and non-native, and a diversity of English speakers, may help to weaken listener biases.

6.2.3 Hiring Practices

While this study did not show that Korean listeners judged Asian English speakers to be less effective teachers or to have less positive attitudes towards Asian English speakers, it did reveal that Asian English speakers were judged as having less native-like accents, even when the speaker was in fact a NES. This may negatively affect employment opportunities for both Asian NESTs and NNESTs in Korea if institutions place value on native-like accents, which is possible considering the preference demonstrated by some learners for native-

accented teachers (e.g. Butler, 2007). Blind hiring practices, where the ethnicity of applicants is obscured from hiring personnel, could ensure that non-Caucasian applicants are treated fairly when it comes to judgements regarding the suitability of their accents. A more challenging step is to promote the legitimacy of non-Caucasian English speakers, both NESs and NNEs, as well as non-native accents and different varieties of English. Reaching such a goal may require change not only on an individual and institutional level, but also deeper, societal change in the way that ownership of English is viewed and in the relative value placed on non-native and different native varieties of English. More immediately, the association between judgements of accentedness and teaching quality in this study should prompt institutions to consider more carefully any reliance on student evaluations of their teaching staff. Although this study does not provide evidence for the direction of this relationship, it should be considered that lower teaching ratings of non-Caucasian English teachers may be based on judgements of accent rather than professional skills.

6.2.4 Testing

The tendency to judge an Asian speaker's pronunciation as less native-like may also have implications for language testing, particularly where test-takers' pronunciation is incorporated into the result. Research into whether this is the case and, if so, the ramifications for non-Caucasian test-takers is therefore important.

6.3 Limitations and Suggestions for Improvement

This section discusses limitations and suggestions for improvement for future research in addition to those already discussed in Section 3.9.

6.3.1 Quantitative method

Quantitative methods are useful to determine the existence of relationships and effects, but as Dörnyei (2007) points out, '[s]imilar scores can result from quite different underlying processes, and quantitative methods are generally not very sensitive to uncovering the reasons for particular observations or the dynamics underlying the examined situation or phenomenon' (p.35). Undoubtedly, qualitative data would provide valuable additional insights into this study's topic. For future research, further data gathering in the form of interviews, focus groups, or the inclusion of open-ended questions in online questionnaires could shed light on what considerations contribute to participants' rating behaviour, and further explore any associations they make between ethnicity and native-speaker status.

6.3.2 Matched-Guise Technique and Unidirectional Listening

The matched-guise technique can be criticised as artificial, and is 'usually denuded of gestural communication' (Rubin, 1992: p.528). This criticism of artificiality can be made of any unidirectional, audio-only listening exercise, but in intelligibility studies in particular, removing both visual stimuli and the possibility of interaction between the speaker and listener, one or both of which would be present in many real-life situations, removes two vital components of communication which could significantly mitigate or alter the results of this study. For example, O'Neal (2015) demonstrated that intelligibility problems can be resolved when NNES interlocutors use communication strategies to negotiate pronunciation. Meanwhile, Lindemann (2002) found that NESs with high levels of bias against an interlocutor's ethnicity may use communication strategies which withhold necessary feedback about their partner's intelligibility, leaving the latter unaware that they have not been understood and therefore removing the possibility of further accommodation. Future

research could investigate any mitigating effect of interaction on the dependent variables explored in the present study.

Furthermore, the criticism levelled at Chun (2014) in Section 5.3 – that asking students to rate teachers in the abstract along ethnic lines encourages stereotyping behaviour that may not be evident when rating specific teachers – could be made of the present study. It could be argued that asking participants to rate the speaker on such a large range of items based on only a short extract of speech could also have encouraged them in some cases to draw on ethnicity-based stereotypes, if they felt they had insufficient exposure to the speaker to rate him more objectively (e.g. Robinson, 1972). In other words, the artificial nature of the matched-guise technique and the accompanying rating tasks may have prompted participant rating behaviour which would not have otherwise occurred. Providing a more comprehensive introduction to the speaker through multiple speech samples may mitigate this potentially confounding effect.

6.3.3 Between-Subjects Design and Listener Biases

Kang and Rubin (2009) demonstrated that propensity to RLS is a listener characteristic which can predict individual listeners' comprehension, accentedness ratings and teaching quality ratings of a non-Caucasian speaker. The between-subjects design of the present study was unable to measure this variable, which could have provided more nuance to the findings. If the within-subjects design necessary for this measurement is unfeasible, use of an IAT, such as those used by Babel and Russell (2015) and Yi et al. (2014), could enable alternative validated measurements of listener biases.

6.3.4 Proficiency Measure

RQ4 targeted the mediating effect of listener characteristics on the dependent variables. Listener proficiency was of particular interest, but preliminary analyses revealed that the self-reported proficiency measure was too unreliable for further analysis. Not enough participants reported TOEFL or TOEIC scores to use these validated tests as a proficiency measure. The addition of a proficiency test within this study was unfeasible, but would be useful in a future study if possible. A middle ground, requiring less resources than a proficiency test but potentially more accurate than a single self-reported level, would be a self-assessment based on ‘can do’ statements, such as the Common European Framework of Reference for languages (Council of Europe, 2001).

6.3.5 Speech Sample

As discussed in Section 5.2.1, the idiom presented in the speech sample may have been known to participants, confounding the comprehension measure with listeners’ previous knowledge. Future research should consider choosing comprehension questions which rely on understanding of the speech sample rather than background knowledge. A ceiling effect in all variables may also have been created by the use of a NES. Some recent studies (e.g. Hanulíková, 2018; McGowan, 2015) have demonstrated that presenting speech samples in noise can lead to more significant results by increasing the difficulty of the listening task. Future research could investigate the effect of adverse listening conditions on NNES evaluations of speakers.

6.4 Further Recommendations for Research

6.4.1 Research in EFL Contexts

This study demonstrated that Korean L1 NNEs in Korea displayed some of the same biases related to speaker ethnicity as NESs and NNEs in native-speaking contexts, but that they differed in other areas. However, NNEs' evaluations of speech may be highly context-dependent. Further research in different EFL contexts is necessary to investigate how generalisable these findings are.

6.4.2 Future Intelligibility Studies

Future research in the field of intelligibility should recognise listener reactions to speaker ethnicity as a potential confounding variable, especially in measures of intelligibility and accentedness. Measures of listener bias such as an IAT may be used to indicate how much variance can be ascribed to nonlanguage factors.

6.4.3 Expectation-Based Theories

This dissertation has explained the results of the study with theories regarding RLS and listener biases. However, Section 5.5 briefly introduced the alternative of expectation-based theories, which could also account for the findings here. Whether listeners are influenced by RLS or by a mismatch of expectations regarding speaker ethnicity and speaker accent warrants further research. A replication of the present study with non-native accented speech rather than native-accented speech could illuminate which competing theory is a more likely explanation. If RLS theories are correct, in such a study, results would be the same: the Caucasian guise would emerge as more intelligible and less accented than the Asian guise. If expectation-based theories are correct, the Caucasian guise would emerge as less intelligible and more accented than the Asian guise, as this pairing would violate listeners' expectations.

6.4.4 Effects of Exposure to NESs and NNEs

Finally, although the main effects of exposure, unrelated to ethnicity, are not the focus of this study, it is interesting to note that exposure to NESs affected participants' comprehensibility and superiority ratings, but that exposure to NNEs did not have any effect. It is difficult to draw conclusions regarding non-significant findings with such a small sample size and variables of questionable validity, but further research into how exposure to various English speakers affects listener attitudes and listening behaviour may be fruitful.

References

- Ahn, H. (2014). Teachers' attitudes towards Korean English in South Korea. *World Englishes*, 33(2), 195–222. <https://doi.org/10.1111/WENG.12081>
- Babel, M., & Russell, J. (2015). Expectations and speech intelligibility. *The Journal of the Acoustical Society of America*, 137(5), 2823–2833. <https://doi.org/10.1121/1.4919317>
- Barrass, J. P., Briggs Baffoe-Djan, J., Rose, H., & Boggs, J. A. (2020). Intelligibility and comprehensibility of Korean English speakers' phonological features in lingua franca listening contexts. *Journal of Asia TEFL*, 17(1), 124–142. <https://doi.org/10.18823/asiatefl.2016.13.1.1.i>
- Bissett, J., & Ma, J. H. (2015). Attitudes of Korean University Students towards English Pronunciation. *Foreign Language Education Research*, 18, 1–15.
- Braine, G. (2005). *Teaching English to the world [electronic resource]: history, curriculum, and practice*. L. Erlbaum Associates.
- Buckingham, L. (2014). Attitudes to English teachers' accents in the Gulf. *International Journal of Applied Linguistics*, 24(1), 50–73. <https://doi.org/10.1111/IJAL.12058>
- Butler, Y. G. (2007). How are nonnative-English-speaking teachers perceived by young learners? *TESOL Quarterly*, 41(4), 731–755. <https://doi.org/10.1002/j.1545-7249.2007.tb00101.x>
- Carrell, P. L., Dunkel, P. A., & Mollaun, P. (2004). The effects of notetaking, lecture length, and topic on a computer-based test of ESL listening comprehension. *Applied Language Learning*, 14(1), 83.
- Choi, T.-H. (2021). English fever: educational policies in globalised Korea, 1981–2018. *History of Education*. <https://doi.org/10.1080/0046760X.2020.1858192>
- Chun, S. Y. (2014). EFL learners' beliefs about native and non-native English-speaking teachers: Perceived strengths, weaknesses, and preferences. *Journal of Multilingual and*

Multicultural Development, 35(6), 563–579.

<https://doi.org/10.1080/01434632.2014.889141>

Cobb, T. Compleat VP. [Computer programme]. Accessed 4th August 2021 at

<https://www.lex tutor.ca/vp/comp/>

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.).
Routledge.

Council of Europe. (2001). *Common European Framework of Reference for Languages:
Learning, Teaching, Assessment*. Cambridge University Press.

<https://doi.org/10.1142/S0219498810003951>

Crystal, D. (2006). English worldwide. In D. Denison & R. Hogg (Eds.), *A History of the
English Language* (pp. 420–439). Cambridge University Press. [https://doi.org/DOI:
10.1017/CBO9780511791154.010](https://doi.org/DOI:10.1017/CBO9780511791154.010)

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

<https://doi.org/10.1017/S026144480800551X>

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL
learners' fluency and comprehensibility development. *Applied Linguistics*, 29(3), 359–
380. <https://doi.org/10.1093/applin/amm041>

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and
mixed methodologies*. Oxford University Press.

Eisenclas, S. A., & Michael, R. B. (2019). What's in a face? The impact of nonlinguistic
'ethnic' facial features on accent perception. *Journal of Multilingual and Multicultural
Development*, 40(10), 879–891. <https://doi.org/10.1080/01434632.2019.1587445>

Galloway, N., & Rose, H. (2015). *Introducing Global Englishes*. Routledge.

Gnevsheva, K. (2018). The expectation mismatch effect in accentedness perception of Asian

and Caucasian non-native speakers of English. *Linguistics*, 56(3), 581–598.

<https://doi.org/10.1515/ling-2018-0006>

Hanulíková, A. (2018). The effect of perceived ethnicity on spoken text comprehension under clear and adverse listening conditions. *Linguistics Vanguard*, 4(1), 1–9.

<https://doi.org/10.1515/lingvan-2017-0029>

Hendriks, B., Meurs, F. Van, & Usmany, N. (2021). The effects of lecturers' non-native accent strength in English on intelligibility and attitudinal evaluations by native and non-native English students. *Language Teaching Research*, 1–30.

<https://doi.org/10.1177/1362168820983145>

Holliday, A. (2006). Native-speakerism. *ELT Journal*, 60(4), 385–387.

<https://doi.org/10.1093/elt/ccl030>

Hu, G., & Su, J. (2015). The effect of native/non-native information on non-native listeners' comprehension. *Language Awareness*, 24(3), 273–281.

<https://doi.org/10.1080/09658416.2015.1077853>

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an International Language. *Applied Linguistics*, 23(1), 83–103.

<https://doi.org/10.1093/applin/23.1.83>

Kachru, B. B. (1992). World Englishes: Approaches, issues and resources. *Language Teaching*, 25(1), 1–14. <https://doi.org/10.1017/S0261444800006583>

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301–315.

<https://doi.org/10.1016/j.system.2010.01.005>

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*,

28(4), 441–456. <https://doi.org/10.1177/0261927X09341950>

- Kang, O., Rubin, D. L., & Lindemann, S. (2015). Mitigating U.S. undergraduates' attitudes toward international teaching assistants. *TESOL Quarterly*, *49*(4), 681–706.
[https://doi.org/10.1016/S0924-8099\(05\)80003-4](https://doi.org/10.1016/S0924-8099(05)80003-4)
- Kang, O., Thomson, R. I., & Moran, M. (2018). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning*, *68*(1), 115–146. <https://doi.org/10.1111/lang.12270>
- Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, *64*(3), 459–489. <https://doi.org/10.3138/cmlr.64.3.459>
- Laerd Statistics. (2015). Independent-samples t-test using SPSS Statistics. *Statistical Tutorials and Software Guides*. Retrieved from <https://statistics.laerd.com/>
- Lambert, W. E. (1967). A social psychology of bilingualism. *Journal of Social Issues*, *23*(2), 91–109. <https://doi.org/10.1111/j.1540-4560.1967.tb00578.x>
- Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *J Abnorm Soc Psychol*, *60*(1), 44–51.
<https://doi.org/10.1037/h0044430>
- Lindemann, S. (2002). Listening with an attitude: A model of native-speaker comprehension of non-native speakers in the United States. *Language in Society*, *31*(3), 419–441.
<https://doi.org/10.1017/S0047404502020286>
- Lippi-Green, R. (2012). *English with an accent: language, ideology, and discrimination in the United States* (2nd ed.). Routledge.
- McGowan, K. B. (2015). Social Expectation Improves Speech Perception in Noise. *Language and Speech*, *58*(4), 502–521. <https://doi.org/10.1177/0023830914565191>
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97.

<https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>

- Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*, *41*(4), 647–672. <https://doi.org/10.1017/S0272263119000044>
- Nelson, C. L. (1995). Intelligibility and world Englishes in the classroom. *World Englishes*, *14*(2), 273–279. <https://doi.org/10.1111/j.1467-971X.1995.tb00356.x>
- Norton, B. (1997). Language, identity, and the ownership of English. *TESOL Quarterly*, *31*(3), 409. <https://doi.org/10.2307/3587831>
- O’Neal, G. (2015). Interactional intelligibility: The relationship between consonant modification and pronunciation intelligibility in English as a Lingua Franca in Japan. *Asian Englishes*, *17*(3), 222–239. <https://doi.org/10.1080/13488678.2015.1041871>
- Orikasa, M. (2016). The intelligibility of varieties of English in Japan. *World Englishes*, *35*(3), 355–371. <https://doi.org/10.1111/weng.12209>
- Pallant, J. (2016). *SPSS survival manual* (6th ed.). Open University Press.
- Park, J. K. (2012). Korean college students’ perception of the pronunciation of English as an International Language. *English 21*, *25*(3), 319–341.
- Pickering, L. (2006). Current research on intelligibility in English as a lingua franca. *Annual Review of Applied Linguistics*, *26*, 219–233. <https://doi.org/10.1017/S0267190506000110>
- Pituch, K. A., & Stevens, J. (2016). *Applied multivariate statistics for the social sciences : analyses with SAS and IBM’s SPSS* (6th ed.). Routledge.
- Rajagopalan, K. (2010). The soft ideological underbelly of the notion of intelligibility in discussions about “World Englishes.” *Applied Linguistics*, *31*(3), 465–470. <https://doi.org/10.1093/applin/amq014>
- Robinson, W. P. (1972). *Language and social behaviour*. Penguin.

- Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education, 33*(4), 511–531.
<https://doi.org/10.1007/BF00973770>
- Rubin, D. L. (2012). The power of prejudice in accent perception: Reverse linguistic stereotyping and its impact on listener judgments and decisions. *Proceedings of the 3rd Annual Pronunciation in Second Language Learning and Teaching Conference*, 11–18.
<https://doi.org/10.13140/RG.2.1.1465.4485>
- Rubin, D. L., Coles, V. B., & Barnett, J. T. (2016). Linguistic stereotyping in older adults' perceptions of health care aides. *Health Communication, 31*(7), 911–916.
<https://doi.org/10.1080/10410236.2015.1007549>
- Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations, 14*(3), 337–353.
[https://doi.org/10.1016/0147-1767\(90\)90019-S](https://doi.org/10.1016/0147-1767(90)90019-S)
- Ruecker, T. (2011). Challenging the native and nonnative English speaker hierarchy in ELT: New directions from race theory. *Critical Inquiry in Language Studies, 8*(4), 400–422.
<https://doi.org/10.1080/15427587.2011.615709>
- Selvi, A. F. (2014). Myths and misconceptions about nonnative English speakers in the TESOL (NNEST) movement. *TESOL Journal, 5*(3), 573–611.
<https://doi.org/10.1002/tesj.158>
- Sheppard, B. E., Elliott, N. C., & Baese-Berk, M. M. (2017). Comprehensibility and intelligibility of international student speech: Comparing perceptions of university EAP instructors and content faculty. *Journal of English for Academic Purposes, 26*, 42–51.
<https://doi.org/10.1016/j.jeap.2017.01.006>
- Shuck, G. (2006). Racializing the nonnative English speaker. *Journal of Language, Identity,*

- and Education*, 5(4), 259–276. https://doi.org/10.1207/s15327701jlie0504_1
- Times Higher Education (2020). *Asia University Rankings 2020*. Retrieved from https://www.timeshighereducation.com/world-university-rankings/2020/regional-ranking#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats
- Wang, Y., & Jenkins, J. (2016). “Nativeness” and intelligibility: Impacts of intercultural experience through English as a lingua franca on Chinese speakers’ language attitudes. *Chinese Journal of Applied Linguistics*, 39(1), 38–58. <https://doi.org/10.1515/cjal-2016-0003>
- Widdowson, H. G. (1994). The ownership of English. *TESOL Quarterly*, 28(2), 377–389. <https://doi.org/10.2307/3587438>
- Yi, H. G., Smiljanic, R., & Chandrasekaran, B. (2014). The neural processing of foreign-accented speech and its relationship to listener bias. *Frontiers in Human Neuroscience*, 8, 1–12. <https://doi.org/10.3389/fnhum.2014.00768>
- Zahn, C. J., & Hopper, R. (1985). Measuring language attitudes: The speech evaluation instrument. *Journal of Language and Social Psychology*, 4(2), 113–123. <https://doi.org/10.1177/0261927X8500400203>
- Zheng, Y., & Samuel, A. G. (2017). Does seeing an Asian face make speech sound more accented? *Attention, Perception, and Psychophysics*, 79(6), 1841–1859. <https://doi.org/10.3758/s13414-017-1329-2>
- Zielinski, B. W. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, 36(1), 69–84. <https://doi.org/10.1016/j.system.2007.11.004>

List of Appendices

Appendix A: Confirmation of Ethical Approval

Appendix B: Participant Poster

Appendix C: Pilot Study Feedback Form

Appendix D: Speech Sample Script

Appendix E: Guise Photographs

Appendix F: Language Attitude Items

Appendix G: Comprehension Question

Appendix H: Cloze Tests

Appendix I: Page Displayed For Missing Cloze Test Answers

Appendix J: Comprehensibility Items

Appendix K: Accentedness Items

Appendix L: Teaching Quality Items

Appendix M: Items for Exposure N(N)ESs

Appendix N: Participant Information

Appendix O: Descriptive Statistics

Appendix P: Assumption Testing for Independent-Samples T-Tests

Appendix Q: Independent-Samples t-tests for Mean Dependent Variables

Appendix R: Independent-Samples t-tests for Teaching Quality Items

Appendix S: Graphic Representations of Independent-Samples T-Tests

Appendix T: Interaction Effects of Knowledge of L2s Other Than English and Condition on
Dependent Variables

Appendix U: Interaction Effects of Exposure to NESs and Condition on Dependent Variables

Appendix V: Interaction Effects of Exposure to NNESs and Condition on Dependent
Variables

Appendix A: Confirmation of Ethical Approval

Joshua McGrane

22 July 2020 at 01:51



Re: CUREC for ██████████ MSc ALLT

[Details](#)

To: ██████████ Cc: Education Research Office, ██████████

Security: Signed (Joshua McGrane)



Dear ██████████

Title: "Listener attitudes towards speaker intelligibility, comprehensibility, and teaching quality"

The above application (reference number ED-CIA-20-217) has been considered on behalf of the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to DREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly, approval has been granted.

If your research involves participants whose ability to give free and informed consent is in question (this includes those under 18 and vulnerable adults), then it is advisable to read the following NSPCC professional reporting requirements for cases of suspected abuse http://www.nspcc.org.uk/inform/research/questions/reporting_child_abuse_wda74908.html

Should there be any subsequent changes to the project which raise ethical issues not covered in the original application you should submit details to research.office@education.ox.ac.uk for consideration.

Good luck with your research study.

Yours sincerely,
Joshua McGrane
Member of DREC

Appendix B: Participant Poster

Department of Education
15 Norham Gardens
Oxford, OX2 6PY
+44 (0) 1865 274024



[REDACTED]

화자의 이해명료성, 이해가능성 및 수업의 질에 대한 청자의 태도

옥스퍼드 대학교 중앙연구윤리위원회 승인 조회번호: ED-CIA-20-217

연구에 참여할 지원자 모집: 온라인으로 영어 듣기 문제를 풀고 설문지 작성하기

[REDACTED]

본인 [REDACTED]은 한국 대학생들을 대상으로 영어 발화의 명료성, 이해가능성 및 영어 수업의 질을 판단하는데 영향을 미치는 요소가 무엇인지 파악하기 위한 연구를 진행하고 있습니다.

온라인상으로 짧은 듣기 문제를 풀고 설문지를 작성해 주실 18 세 이상의 지원자를 모집하고 있습니다. 지원자는 현재 학부생이어야 합니다.

5-10 분 정도면 완료할 수 있는 간단한 문항들이며 2020 년 9 월 30 일에서 2021 년 4 월 30 일 사이에 실시됩니다. 참여자들은 영어로 된 짧은 음성 발체문을 듣고 그에 대한 이해도 문제를 풀 다음 오디오에 등장한 화자들에 관한 질문 및 자신의 배경을 묻는 질문에 답하시게 됩니다. 참여자의 신상을 판별할 수 있는 내용은 일절 포함되어 있지 않습니다.

참여를 원하거나 더 많은 정보를 보고 싶다면 QR 코드를 스캔하거나 위의 링크를 클릭해 주십시오. 참여는 전적으로 여러분의 선택에 달렸고 참여시 익명이 보장됩니다.

감사합니다.

Department of Education

15 Norham Gardens

Oxford, OX2 6PY

+44 (0) 1865 274024



[REDACTED]

**Listener attitudes towards speaker intelligibility,
comprehensibility, and teaching quality**
CUREC Approval Reference: ED-CIA-20-217

**VOLUNTEERS NEEDED FOR RESEARCH: ONLINE LISTENING
TASK AND SURVEY**

[REDACTED]

I, [REDACTED], am conducting a study investigating factors which influence Korean university students' judgements of the intelligibility, comprehensibility, and teaching quality of English teachers.

I am looking for volunteers, aged 18 and over, to complete a short online listening task and questionnaire. You should be an undergraduate student.

The task is a short online study which takes about 5 to 10 minutes of your time, between 2020/9/30 and 2021/4/30. You will listen to some short audio extracts in English, answer a few comprehension questions, and then fill in a questionnaire giving your opinions about the speakers you heard and some non-identifying background information about yourself.

If you want to participate or would like more information, please scan the QR code or go to the link provided above. There is no obligation to take part and your participation will be completely anonymous.

Thank you!

Appendix C: Pilot Study Feedback Form

설문 피드백

[REDACTED]

여러분의 설문 참여 목적은 설문지의 이상 여부를 확인하기 위한 것이었습니다. 여러분의 자료는 분석 자료로 쓰이지 않을 것이고 연구의 최종 결과물에 포함되지 않을 것이며 출판되지 않을 것입니다.

온라인으로 설문 참여 후 이 인쇄 버전을 봐 주십시오. 헛갈리거나 명확하지 못했거나 답하기 어려웠던 부분에 동그라미 표시를 해 주십시오. 질문 옆에 어떤 부분이 헛갈리거나 명확하지 못했거나 답하기 어려웠는지를 기재해 주시기 바랍니다. 한국어나 영어로 작성하시면 됩니다.

마지막으로 다음의 질문에 간략하게 한국어 또는 영어로 답해 주십시오.

1. 설문을 완료하는데 대략 얼마나 걸렸습니까? ____ 분
2. 오디오에서 들었던 속어를 이전에 들어본 적이 있습니까? 네 / 아니요
 - 그렇다면 그 뜻을 이미 알고 있었습니까? 네 / 아니요
3. 연습문제나 설문조사에 답하는데 기술적 어려움이 있었습니까?
네 / 아니요

4. 연습문제나 설문조사에 답하는데 있어 앞서 언급하지 않은 다른 어려움이 있었습니까? 네 / 아니요

5. 남기고 싶은 의견이 있습니까? 네 / 아니요

Survey Feedback

[REDACTED]

You have been asked to take this survey so that the researcher (REDACTED) can see if there are any problems with the survey. Your data will *not* be used for analysis, will not be part of the final results, and will not be published.

After you take the survey online, please look at this printed version. Circle any questions, tasks, or information that were confusing, unclear, or difficult to answer. Please write next to the question what you found confusing, unclear, or difficult to answer. You can write in Korean or English.

Finally, please briefly answer the following questions in Korean or English:

1. Approximately how long did it take to complete the survey? _____ minutes

2. Had you ever heard the idiom presented in the audio before? Yes / No

- **If yes, did you already know what it meant? Yes / No**

3. Did you have any technical difficulties with the task or survey? Yes / No

4. Did you have any other difficulties with the task or survey that you did not already mention? Yes / No

5. Do you have any other comments? Yes / No

Appendix D: Speech Sample Script

Burn your bridges.

This idiom means that you do something that makes it impossible for you to return to your previous situation. If you burn your bridges, you have done something that's not easily undone. Imagine you are on a journey from place A to place B, and you cross a bridge. If you burn that bridge behind you, how can you get back to place A if you change your mind? You can't - or at least, it will be very difficult. We often use 'burn your bridges' when talking about relationships with other people. Let's look at some examples:

My friend is having trouble finding a new job, but he can't go back to his old job - he really burned his bridges when he quit, by yelling at his old boss.

In this example, the friend made it impossible for himself to return to his old job, because he had a big argument with his old boss, so his old boss won't want to hire him again. Here's the next example:

I was careful not to burn my bridges when I broke up with my ex-girlfriend, in case I change my mind.

Here, I'm saying that I wasn't mean or unkind to my ex-girlfriend when we broke up, so if I decided I wanted to date her again there's a chance she will say 'yes'.

Appendix E: Guise Photographs

Asian Guise



Lee Sung Min

Caucasian Guise



Christopher Johnson

Appendix F: Attitude Items

Social Attractiveness

Hostile	-	Good-natured
Cold	-	Warm
Unlikeable	-	Likeable
Unpleasant	-	Pleasant
Inconsiderate	-	Considerate
Dishonest	-	Honest
Unfriendly	-	Friendly

Superiority

Low level of education	-	High level of education
Unintelligent	-	Intelligent
Disorganized	-	Organized
Disadvantaged	-	Advantaged
Disfluent	-	Fluent *
Unclear	-	Clear **
Inexperienced	-	Experienced ***

Ethnicity

Caucasian	-	Asian
-----------	---	-------

* Also used in measure of accentedness

** Also used in measure of comprehensibility

*** Also used in measure of teaching quality

Appendix G: Comprehension Question

In what situation could you use the English idiom presented in the lesson you just heard?

8. After I argued with my friend, I posted all the secrets I know about him publicly on social media.
9. The route across the bridge is the shortest, but there's a lot of traffic, so I think it's quicker to take a different road.
10. My friends all tell me that I'm too critical of myself, so I'm trying not to feel too bad about failing my math test.
11. I wanted to take surfing lessons, but by the time I decided to sign up all the classes were full for this summer.

금방 강의에서 들었던 속어를 사용하기에 알맞은 상황은 무엇입니까?

1. 친구와 다툰 후 내가 알고 있는 친구의 모든 비밀을 공개적으로 SNS 에 게시했다
2. 다리를 건너가는 길이 거리상으로는 가장 짧지만 그 길은 차가 너무 많이 막혀서 다른 길로 가는 것이 더 빠를 것이라 생각한다.
3. 내 친구들이 나는 너무 내자신에 대해 비판적이라고 말한다. 그래서 수학 시험을 망친 것에 대해 너무 자책하지 않으려고 노력하고 있다.
4. 서핑 수업을 듣고 싶었으나 등록하려고 마음을 먹었을 때는 이미 이번 여름 수강 인원이 다 차 있었다.

Appendix H: Cloze Tests

1. you have done something that's not easily _____

(Correct answer: *undone*)

2. and you _____ a bridge

(Correct answer: *cross*)

3. he really burned his bridges when he _____

(Correct answer: *quit*)

4. so his old boss won't want to _____ him again

(Correct answer: *hire*)

5. so if I decided I wanted to _____ her again

(Correct answer: *date*)

Appendix I: Page Displayed For Missing Cloze Test Answers

Please indicate why you did not answer the question on the previous page.

1. The audio did not work
2. I didn't know the answer
3. I ran out of time
4. Other _____

앞 페이지의 질문에 답하지 않은 이유를 선택해 주세요.

1. 오디오가 나오지 않았다
2. 답을 모른다
3. 시간이 부족했다
4. 기타 _____

Appendix J: Comprehensibility Items

Incomprehensible	-	Highly comprehensible
Unclear	-	Clear *
Made it very difficult to grasp the meaning	-	Made it very easy to grasp the meaning
Hard to understand	-	Easy to understand
Required lots of effort to understand	-	Required no effort to understand

* Also used in measure of superiority

Appendix K: Accentedness Items

Has non-native-like intonation	-	Has native-like intonation
Speaks with a non-native English	-	Speaks with a native English accent accent
Disfluent	-	Fluent *
Sounds like a non-native English speaker	-	Sounds like a native English speaker
Speaks with acceptable	-	Speaks with unacceptable pronunciation pronunciation

* Also used in measure of superiority

Appendix L: Teaching Quality Items

Does not motivate students to learn	-	Motivates students to learn
Ineffective teacher	-	Effective teacher
Unhelpful teacher	-	Helpful teacher
Boring	-	Interesting
Difficult for students to interact with	-	Easy for students to interact with
Unqualified	-	Qualified
Unprepared	-	Prepared
Unapproachable	-	Approachable
Informal	-	Formal
Inexperienced	-	Experienced *

* Also used in measure of superiority

Appendix M: Items for Exposure N(N)ESs

Questions about NESs and NNEs were presented separately for a total of eight questions. Square brackets indicate phrasing for items relating to exposure to NNEs.

1. How often do you interact with [non-]native English speakers in English?
 - A. Daily
 - B. Weekly
 - C. Monthly
 - D. Once every few months
 - E. Yearly or less often

2. In your life, how long have you had English classes with a [non-]native English speaking teacher in total?
 - A. One year or less
 - B. One to five years
 - C. More than five years

3. In your life, how long have you spent traveling or living in an [non-]English-speaking country [other than Korea] in total?
 - A. One month or less
 - B. One month to a year
 - C. More than one year

4. When you visited an [non-]English-speaking country [other than Korea], how often did you speak English there?
 - A. All or most of the time
 - B. Some of the time
 - C. Once or twice
 - D. Never / I've never visited an [non-]English-speaking country [other than Korea]

1. 얼마나 자주 영어 [비]원어민과 영어로 의사소통을 합니까?
 - A. 매일
 - B. 일주일에 한 번
 - C. 한 달에 한 번
 - D. 몇 달에 한 번
 - E. 일 년에 한 번 또는 그 이하

2. 평생동안 [비]원어민 영어 선생님과 수업한 기간이 얼마나 됩니까?
 - A. 1 년 또는 그 이하
 - B. 1-5 년
 - C. 5 년 이상

3. [한국을 제외하고 비]영어권 국가에 여행을 가거나 거주했던 총 기간이 얼마나 됩니까?
 - A. 1 개월 이하
 - B. 1 개월 - 1 년
 - C. 1 년 이상

4. [한국을 제외한 비]영어권 국가에 방문했을 때 영어를 얼마나 사용했나요?
 - A. 항상 또는 거의
 - B. 가끔
 - C. 한두 번
 - D. 한 번도 쓰지 않음 / [한국을 제외한 비]영어권 국가 방문한 적 없음

Appendix N: Participant Information

Listener attitudes towards speaker intelligibility, comprehensibility, and teaching quality

CUREC Approval Reference: ED-CIA-20-217

General Information

The aim of this study is to investigate factors which influence Korean university students' judgements of the intelligibility, comprehensibility, and teaching quality of English teachers.

We appreciate your interest in participating in this listening exercise and questionnaire. You have been invited to participate as you are over 18 years of age, speak Korean as your first language, and are currently an undergraduate student. Please read through this information before agreeing to participate by selecting the 'Yes, I agree to take part' box below.

You may ask any questions before deciding to take part by contacting the researcher (details below).

I, [REDACTED], a Masters student at the Department of Education, University of Oxford, am investigating what makes speech in English intelligible and comprehensible to Korean students, as well as what factors may influence teaching evaluations.

You will listen to two short extracts of speech in English and complete a simple comprehension exercise. Then you will complete a questionnaire rating the speakers you heard, as well as giving some non-identifying background information about yourself and your experiences with English. This should take about 5 to 10 minutes. All questions, except

the listening exercise, will be provided in both Korean and English. Your answers will be completely anonymous as they contain no identifying or personal information. None of your teachers or your university's administrative staff will know if you took part or not.

Do I have to take part?

No, your participation is entirely voluntary. If you do decide to take part, you may withdraw for any reason at any point during the questionnaire before submitting your answers by exiting the questionnaire or closing the browser before pressing the 'Submit' button on the final page.

How will my data be used?

Your answers will be completely anonymous, as we will not ask you for any personal or identifying information. The researchers, your teachers, and your university's administrative staff will not be able to know if you personally took part or not.

Your data will be stored in a password-protected file and may be used in academic publications. Your IP address will not be stored. All questions are optional. Research data will be stored for a minimum of three years after publication or public release.

Who will have access to my data?

No personal data will be collected in this research. Only the researcher will have access to your anonymous answers to the online questionnaire.

Responsible members of the University of Oxford and funders may be given access to this data for monitoring and/or audit of the study to ensure we are complying with guidelines, or as otherwise required by law.

This questionnaire is for an MSc project. The Principal Researcher is [REDACTED], who is attached to the Department of Education at the University of Oxford. This project is being completed under the supervision of [REDACTED].

This project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (reference number: ED-CIA-20-217).

Who do I contact if I have a concern about the study or I wish to complain?

If you have a concern about any aspect of this study, please speak to [REDACTED] ([REDACTED]) or [REDACTED] supervisor [REDACTED] ([REDACTED]) and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with.

If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

Dr Liam Gearon (Chair of the Departmental Research Ethics Committee)

liam.gearon@education.ox.ac.uk

+44 (0)1865 274047

화자의 이해명료성, 이해가능성 및 수업의 질에 대한 청자의 태도

옥스퍼드 대학교 중앙연구윤리위원회 (CUREC, Central University Research Ethics

Committee) 승인 조회번호: ED-CIA-20-217

개요

이 연구의 목적은 한국의 대학생들이 영어 강사들의 이해명료성, 이해가능성 및 수업의 질을 판단하는 데 있어 영향을 미치는 요소들이 무엇인지 조사하는 것입니다. 이 청취 연습과 설문지 작성에 응해 주셔서 감사합니다. 여러분은 18 세 이상의 성인으로서 한국어를 모국어로 하며 학부생으로서 이 설문지에 참여하도록 요청받았습니다. ‘네, 참여에 동의합니다’를 선택하기 전 아래의 글을 먼저 잘 읽어 보시기 바랍니다.

설문에 응하기 전 질문이 있다면 담당 연구자에게 무엇이든 물어보십시오(하단에 연락처 기재되어 있음).

본인([REDACTED])은 현재 옥스퍼드 대학교 교육학과의 석사과정을 밟고 있으며 한국 학생들에게 있어 영어 발화를 명료하고 이해하기 쉽게 만드는 요소는 무엇인지, 또한 강사 평가에 영향을 주는 요소는 무엇인지에 대해 연구하고 있습니다.

여러분은 영어로 된 몇 개의 짧은 발췌문을 듣고 그에 대한 이해도 연습 문항에 답을 하게 될 것입니다. 그 후 여러분이 들었던 화자에 대해 평가하는 설문지를 작성할 것이며 개인 정보가 배제된 본인 소개 및 영어와 관련된 본인의 경험을 작성하게 됩니다. 이 과정은 총 5~10 분 정도 소요될 예정입니다. 듣기 연습문제를 제외한 모든 질문은 영어와 한글 두 언어로 제공될 것입니다. 제출하신 답안지는 개인의 신상을 판별할 수 있는 어떠한 정보도 담고 있지 않으므로 완전히 익명입니다. 교수나 행정직원을 포함한 어느 누구도 여러분의 설문 참여여부를 알 수 없습니다.

꼭 참여해야 하나요?

그렇지 않습니다. 참여는 전적으로 여러분의 선택에 달려있습니다. 참여하기로 결정한 후 어떤 이유에서든 중도에 그만두기를 원한다면 언제든지 브라우저를 닫고 설문조사를 그만두실 수 있습니다.

제 자료가 어떻게 쓰이나요?

이 설문지에는 어떠한 개인정보도 포함되어 있지 않으므로 여러분의 대답은 완전히

익명입니다. 연구자들, 여러분의 교수님들 및 행정직원들을 포함한 누구도 여러분의 설문 참석 여부를 알 수 없습니다.

여러분의 자료는 비밀번호로 보호된 파일에 저장될 것이고 학술적 출판 용도로 쓰일 수 있습니다. 여러분의 IP 주소는 저장되지 않습니다. 모든 질문에 다 답을 해야만 하는 것은 아닙니다. 연구 자료는 공개된 후 최소 3 년간 저장됩니다.

제 자료를 누가 볼 수 있나요?

이 연구는 어떠한 개인 정보도 수집하지 않습니다. 여러분의 익명 온라인 답변을 볼 수 있는 사람은 오직 이 연구의 담당자 한 명뿐입니다. 규정과 법률 준수 여부를 감시/ 감사하기 위해 옥스퍼드 대학교의 책임자나 자금 지원자들에게 자료에 대한 접근 권한이 주어질 수도 있습니다.

이 설문지는 석사과정 프로젝트를 목적으로 만들어졌습니다. 이 연구의 담당자는 옥스퍼드 대학교 교육학과의 [REDACTED]입니다. 이 연구는 [REDACTED]의 지휘하에 진행되고 있습니다.

이 프로젝트는 옥스퍼드대학 중앙연구윤리위원회(CUREC, Central University Research Ethics Committee 승인 조회번호: ED-CIA-20-217)로부터 검토 및 윤리 승인을 받았습니다.

이 연구에 관한 우려사항이나 불만이 있는 경우 누구에게 문의해야 하나요?

이 연구에 관해 우려사항이 있는 경우 [REDACTED]([REDACTED])이나 [REDACTED]([REDACTED])에게 연락 주시면 성심 성의껏 답변해 드리겠습니다. 여러분의 문의는 평일 기준 10 일 이내로 문제 해결책을 포함한 답변을 드리겠습니다.

답변을 받은 이후로도 계속해서 불만이 지속된다면 옥스퍼드 대학교 연구윤리 위원회의 회장에게 연락 주시면 신속히 해결할 방법을 모색하겠습니다.

Dr Liam Gearon (연구윤리 위원회 회장)

liam.gearon@education.ox.ac.uk

+44 (0)1865 274047

Appendix O: Descriptive Statistics

Table O1

Continuous Dependent Variables

		N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
Condition		Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Asian	Attractiveness	33	4.00	7.00	5.42	1.06	0.02	0.41	-1.44	0.80
	Superiority	33	4.00	7.00	5.80	0.86	-0.33	0.41	-0.75	0.80
	Intelligibility	32	0.00	5.00	3.84	1.22	-1.16	0.41	1.58	0.81
	Comprehensibility	33	2.25	7.00	5.42	1.48	-0.61	0.41	-0.83	0.80
	Accentedness	33	4.00	7.00	6.15	0.93	-0.78	0.41	-0.66	0.80
	Teaching quality	33	4.00	7.00	5.65	0.90	-0.28	0.41	-0.78	0.80
Caucasian	Attractiveness	41	4.00	7.00	5.48	0.99	0.23	0.37	-1.15	0.72
	Superiority	41	4.00	7.00	5.75	0.79	-0.32	0.37	-0.62	0.72
	Intelligibility	39	3.00	5.00	4.46	0.76	-1.02	0.38	-0.43	0.74
	Comprehensibility	41	1.00	7.00	5.44	1.81	-1.05	0.37	0.21	0.72
	Accentedness	41	5.00	7.00	6.65	0.58	-1.71	0.37	1.97	0.72
	Teaching quality	41	2.78	7.00	5.63	1.09	-0.44	0.37	-0.15	0.72

Sample	Attractiveness	74	4.00	7.00	5.45	1.01	0.12	0.28	-1.27	0.55
	Superiority	74	4.00	7.00	5.77	0.82	-0.32	0.28	-0.71	0.55
	Intelligibility	71	0.00	5.00	4.18	1.03	-1.42	0.28	2.56	0.56
	Comprehensibility	74	1.00	7.00	5.43	1.66	-0.91	0.28	-0.02	0.55
	Accentedness	74	4.00	7.00	6.42	0.79	-1.30	0.28	0.63	0.55
	Teaching quality	74	2.78	7.00	5.64	1.00	-0.39	0.28	-0.28	0.55

Table O2*Comprehension and Background Characteristics*

Variable	Value	Asian		Caucasian		Sample	
		<i>n</i>	Percent	<i>n</i>	Percent	<i>n</i>	Percent
Comprehension	Incorrect	18	54.5	19	46.3	37	50.0
	Correct	15	45.5	22	53.7	37	50.0
	Missing	0	0.0	0	0.0	0	0.0
Self-Reported English Proficiency	Beginner	11	33.3	9	22.0	30	27.0
	Intermediate	14	42.4	23	56.1	37	50.0
	Advanced	8	24.2	9	22.0	17	23.0
	Missing	0	0.0	0	0.0	0	0.0
Other L2s	Yes	9	27.3	9	22.0	18	24.3
	No	24	72.7	31	75.6	55	75.3
	Missing	0	0.0	1	2.4	1	1.4
Exposure to NESSs	Low	7	21.2	8	19.5	15	20.3
	Medium	21	63.6	21	51.2	42	56.8
	High	5	15.2	11	26.8	16	21.6
	Missing	0	0.0	1	2.4	1	1.4
Exposure to NNEs	Low	9	27.3	10	24.4	19	25.7
	Medium	20	60.6	23	56.1	43	58.1
	High	4	12.1	8	19.5	12	16.2
	Missing	0	0.0	0	0	0	0.0

Appendix P: Assumption Testing for Independent Samples T-Tests

1. Ethnicity

There were no outliers, as assessed by boxplot inspection.

Ethnicity scores for the two conditions were not normally distributed, as assessed by Shapiro-Wilk's test ($p < .001$ in both conditions).

2. Social Attractiveness

There were no outliers, as assessed by boxplot inspection.

Social attractiveness scores for the two conditions were not normally distributed, as assessed by Shapiro-Wilk's test (Asian condition: $p = .007$; Caucasian condition: $p = .010$).

3. Superiority

There were no outliers, as assessed by boxplot inspection.

Superiority scores for the two conditions were normally distributed, as assessed by Shapiro-Wilk's test (Asian condition: $p = .101$; Caucasian condition: $p = .253$).

4. Intelligibility

There were no outliers, as assessed by boxplot inspection.

Intelligibility scores for the two conditions were not normally distributed, as assessed by Shapiro-Wilk's test (both conditions: $p < .001$).

5. Comprehensibility

There were no outliers, as assessed by boxplot inspection.

Comprehensibility scores for the two conditions were not normally distributed, as assessed by Shapiro-Wilk's test (Asian condition: $p = .003$; Caucasian condition: $p < .001$).

5. Accentedness

Five outliers were found in the Caucasian condition, as assessed by boxplot inspection.

However, removal of these outliers resulted in more outliers being revealed.

As the outliers did not affect the finding of a significant result, the decision was ultimately made not to remove any of them. The tests reported in section 4.3.6 are reported without removal of outliers.

Accentedness scores for the two conditions were not normally distributed, as assessed by Shapiro-Wilk's test (both conditions: $p < .001$).

6. Teaching Quality

There were no outliers, as assessed by boxplot inspection.

Shapiro-Wilk's test revealed that teaching quality scores were normally distributed for the Asian condition ($p = .143$), but not for the Caucasian condition ($p = .015$).

Appendix Q: Independent Samples t-tests for Mean Dependent Variables

Table Q1

Independent Samples t-tests

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2- tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Ethnicity	Equal variances assumed	12.84	< .001	7.82	72	< .001	2.59	0.33	1.93	3.25
	Equal variances not assumed			7.47	51.90	< .001	2.59	0.35	1.89	3.28
Attractiveness	Equal variances assumed	0.70	.405	-0.24	72	.811	-0.06	0.24	-0.53	0.42
	Equal variances not assumed			-0.24	66.38	.812	-0.06	0.24	-0.54	0.42
Superiority	Equal variances assumed	0.67	.417	0.23	72	.820	0.04	0.19	-0.34	0.43
	Equal variances not assumed			0.23	65.82	.822	0.04	0.19	-0.34	0.43

Intelligibility	Equal variances assumed	4.80	.032	-2.61	69	.011	-0.62	0.24	-1.09	-0.15
	Equal variances not assumed			-2.50	49.55	.016	-0.62	0.25	-1.12	-0.12
Comprehensibility	Equal variances assumed	1.17	.283	-0.06	72	.955	-0.02	0.39	-0.80	0.76
	Equal variances not assumed			-0.06	71.98	.954	-0.02	0.38	-0.79	0.74
Accentedness	Equal variances assumed	10.51	.002	-2.86	72	.006	-0.50	0.18	-0.85	-0.15
	Equal variances not assumed			-2.72	51.40	.009	-0.50	0.18	-0.87	-0.13
Teaching quality	Equal variances assumed	0.76	.386	0.08	72	.938	0.02	0.24	-0.45	0.49
	Equal variances not assumed			0.08	71.91	.937	0.02	0.23	-0.44	0.48

Table Q2*Independent Samples Effect Sizes*

		Standardizer a	Point Estimate	95% Confidence Interval	
				Lower	Upper
Ethnicity	Cohen's d	1.41	1.83	1.28	2.37
	Hedges' correction	1.43	1.81	1.26	2.35
	Glass's delta	1.10	2.35	1.65	3.03
Attractiveness	Cohen's d	1.02	-0.06	-0.52	0.40
	Hedges' correction	1.03	-0.06	-0.51	0.40
	Glass's delta	0.99	-0.06	-0.52	0.40
Superiority	Cohen's d	0.82	0.05	-0.41	0.51
	Hedges' correction	0.83	0.05	-0.40	0.51
	Glass's delta	0.79	0.06	-0.40	0.51
Intelligibility	Cohen's d	0.99	-0.62	-1.10	-0.14
	Hedges' correction	1.00	-0.62	-1.09	-0.14
	Glass's delta	0.76	-0.82	-1.32	-0.31
Comprehensibility	Cohen's d	1.68	-0.01	-0.47	0.45
	Hedges' correction	1.69	-0.01	-0.47	0.44
	Glass's delta	1.81	-0.01	-0.47	0.45
Accentedness	Cohen's d	0.75	-0.67	-1.14	-0.20
	Hedges' correction	0.76	-0.66	-1.13	-0.19
	Glass's delta	0.58	-0.87	-1.36	-0.37
Teaching quality	Cohen's d	1.01	0.02	-0.44	0.48
	Hedges' correction	1.02	0.02	-0.44	0.47
	Glass's delta	1.09	0.02	-0.44	0.48

a The denominator used in estimating the effect sizes.

Cohen's d uses the pooled standard deviation.

Hedges' correction uses the pooled standard deviation, plus a correction factor.

Glass's delta uses the sample standard deviation of the control group.

Appendix R: Independent Samples t-tests for Teaching Quality Items

Table R1

Group Statistics for Teaching Quality Items

	Condition	N	Mean	Std. Deviation	Std. Error Mean
Motivates students to learn	Asian	33	4.97	1.40	0.24
	Caucasian	41	5.27	1.50	0.23
Effective teacher	Asian	33	5.76	0.97	0.17
	Caucasian	41	5.8	1.19	0.19
Helpful teacher	Asian	33	5.82	1.01	0.18
	Caucasian	41	5.95	0.95	0.15
Interesting	Asian	33	5.36	1.34	0.23
	Caucasian	41	5.39	1.50	0.23
Easy to interact with	Asian	33	5.48	1.42	0.25
	Caucasian	41	5.07	1.79	0.28
Qualified	Asian	33	6.15	0.97	0.17
	Caucasian	41	6.02	1.11	0.17
Prepared	Asian	33	6.09	1.01	0.18
	Caucasian	41	5.95	1.14	0.18
Approachable	Asian	33	5.55	1.39	0.24
	Caucasian	41	5.34	1.49	0.23
Formal	Asian	33	5.7	1.10	0.19
	Caucasian	41	5.63	1.18	0.18
Experienced	Asian	33	5.67	1.14	0.20
	Caucasian	41	5.88	1.12	0.18

Table R2*Independent Samples t-tests*

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Motivates students to learn	Equal variances assumed	0.20	.657	-0.88	72	.384	-0.30	0.34	-0.98	0.38
	Equal variances not assumed			-0.88	70.35	.381	-0.30	0.34	-0.97	0.38
Effective teacher	Equal variances assumed	0.96	.331	-0.19	72	.854	-0.05	0.26	-0.56	0.46
	Equal variances not assumed			-0.19	71.98	.851	-0.05	0.25	-0.55	0.45
Helpful teacher	Equal variances assumed	0.42	.519	-0.58	72	.562	-0.13	0.23	-0.59	0.32
	Equal variances not assumed			-0.58	66.50	.566	-0.13	0.23	-0.59	0.33

Interesting	Equal variances assumed	0.07	.793	-0.08	72	.937	-0.03	0.34	-0.69	0.64
	Equal variances not assumed			-0.08	71.13	.936	-0.03	0.33	-0.69	0.63
Easy to interact with	Equal variances assumed	2.01	.160	1.08	72	.286	0.41	0.38	-0.35	1.18
	Equal variances not assumed			1.10	71.98	.274	0.41	0.37	-0.33	1.16
Qualified	Equal variances assumed	0.06	.808	0.52	72	.606	0.13	0.25	-0.36	0.62
	Equal variances not assumed			0.53	71.41	.601	0.13	0.24	-0.36	0.61
Prepared	Equal variances assumed	0.23	.630	0.55	72	.583	0.14	0.25	-0.37	0.65
	Equal variances not assumed			0.56	71.269	.579	0.14	0.25	-0.36	0.64
Approachable	Equal variances assumed	0.05	.829	0.60	72	.549	0.20	0.34	-0.47	0.88
	Equal variances not assumed			0.61	70.38	.546	0.20	0.34	-0.47	0.88
Formal	Equal variances assumed	0.24	.629	0.24	72	.815	0.06	0.27	-0.47	0.60

	Equal variances not assumed			0.24	70.31	.814	0.06	0.27	-0.47	0.59
Experienced	Equal variances assumed	0.42	.519	-0.80	72	.426	-0.21	0.26	-0.74	0.32
	Equal variances not assumed			-0.80	68.29	.427	-0.21	0.26	-0.74	0.32

Table R3*Independent Samples Effect Sizes*

		Standardizer a	Point Estimate	95% Confidence Interval	
				Lower	Upper
Motivates students to learn	Cohen's d	1.46	-0.21	-0.66	0.26
	Hedges' correction	1.47	-0.20	-0.66	0.25
	Glass's delta	1.50	-0.20	-0.66	0.26
Effective teacher	Cohen's d	1.10	-0.04	-0.50	0.42
	Hedges' correction	1.11	-0.04	-0.50	0.41
	Glass's delta	1.19	-0.04	-0.50	0.42
Helpful teacher	Cohen's d	0.98	-0.14	-0.60	0.32
	Hedges' correction	0.99	-0.14	-0.59	0.32
	Glass's delta	0.95	-0.14	-0.60	0.32
Interesting	Cohen's d	1.43	-0.02	-0.48	0.44
	Hedges' correction	1.45	-0.02	-0.47	0.44
	Glass's delta	1.50	-0.02	-0.48	0.44
Easy to interact with	Cohen's d	1.64	0.25	-0.21	0.71
	Hedges' correction	1.66	0.25	-0.21	0.70

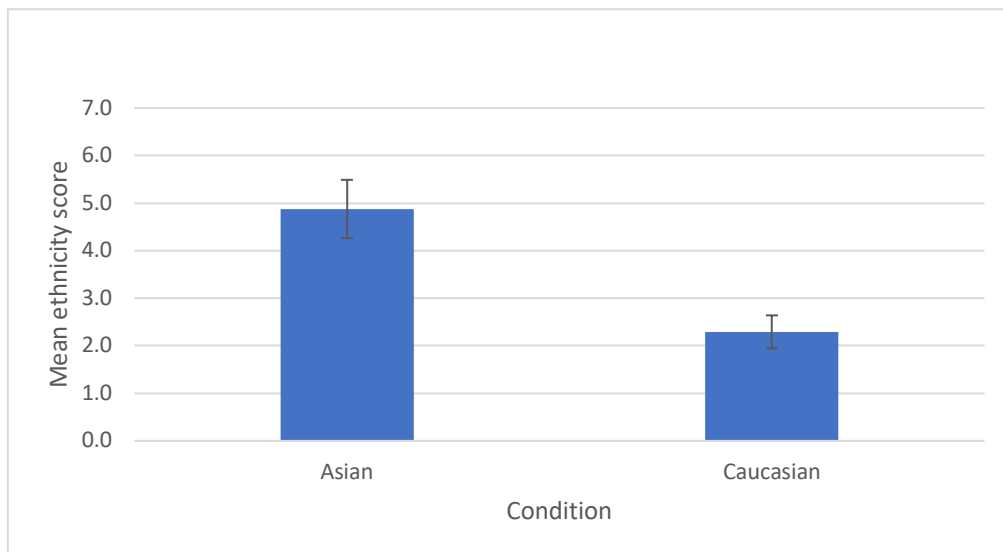
	Glass's delta	1.79	0.23	-0.23	0.69
Qualified	Cohen's d	1.05	0.12	-0.34	0.58
	Hedges' correction	1.06	0.12	-0.33	0.57
	Glass's delta	1.11	0.12	-0.35	0.57
Prepared	Cohen's d	1.08	0.13	-0.33	0.59
	Hedges' correction	1.10	0.13	-0.33	0.58
	Glass's delta	1.14	0.12	-0.34	0.58
Approachable	Cohen's d	1.45	0.14	-0.32	0.60
	Hedges' correction	1.47	0.14	-0.32	0.59
	Glass's delta	1.49	0.14	-0.32	0.60
Formal	Cohen's d	1.15	0.06	-0.40	0.51
	Hedges' correction	1.16	0.05	-0.40	0.51
	Glass's delta	1.18	0.05	-0.41	0.51
Experienced	Cohen's d	1.13	-0.19	-0.65	0.27
	Hedges' correction	1.14	-0.19	-0.64	0.27
	Glass's delta	1.12	-0.19	-0.65	0.27

a The denominator used in estimating the effect sizes.

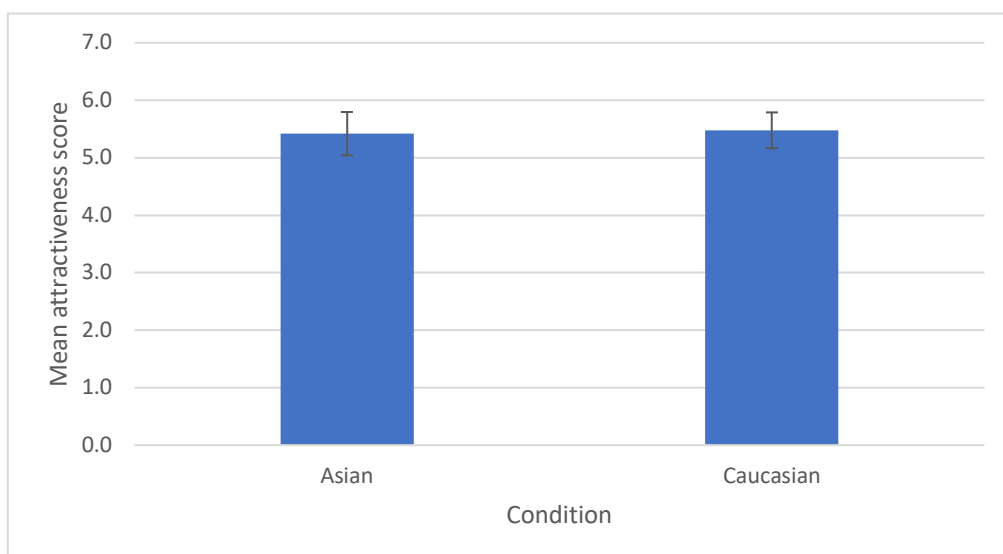
Cohen's d uses the pooled standard deviation.

Hedges' correction uses the pooled standard deviation, plus a correction factor.

Glass's delta uses the sample standard deviation of the control group.

Appendix S: Graphic Representations of Independent Samples T-Tests**Figure S1***Mean Ethnicity Score by Condition*

Note: Low = Caucasian. Error bars: 95% CI.

Figure S2*Mean Social Attractiveness Score by Condition*

Note: Error bars: 95% CI.

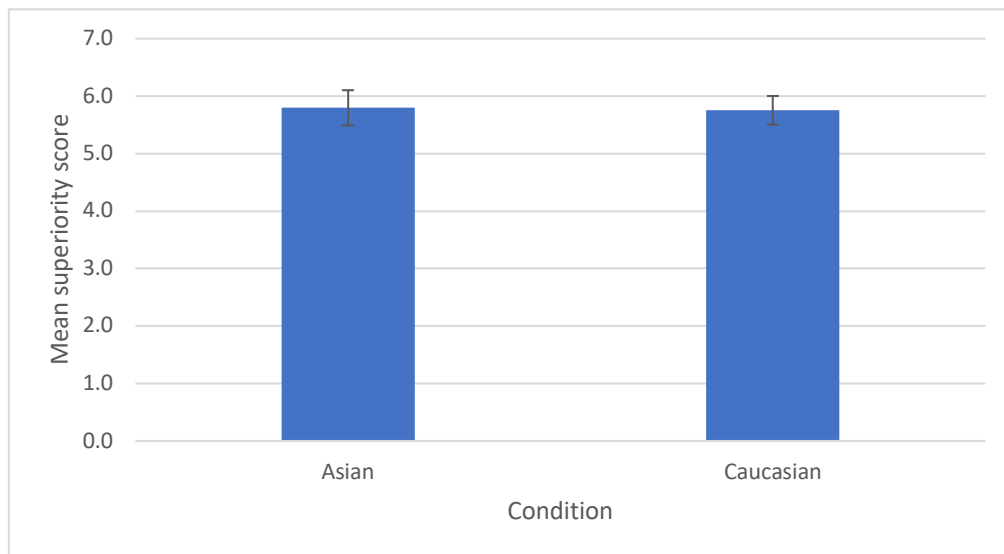
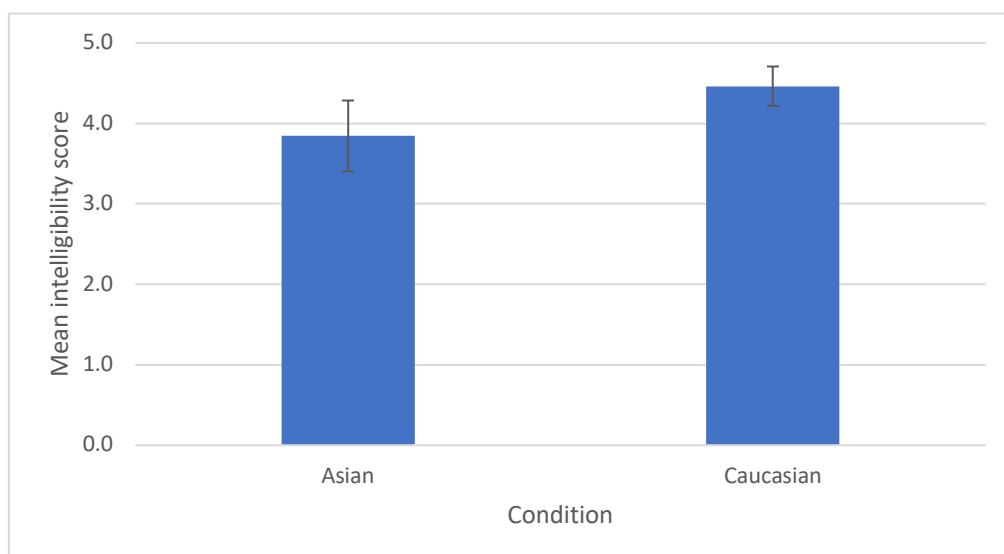
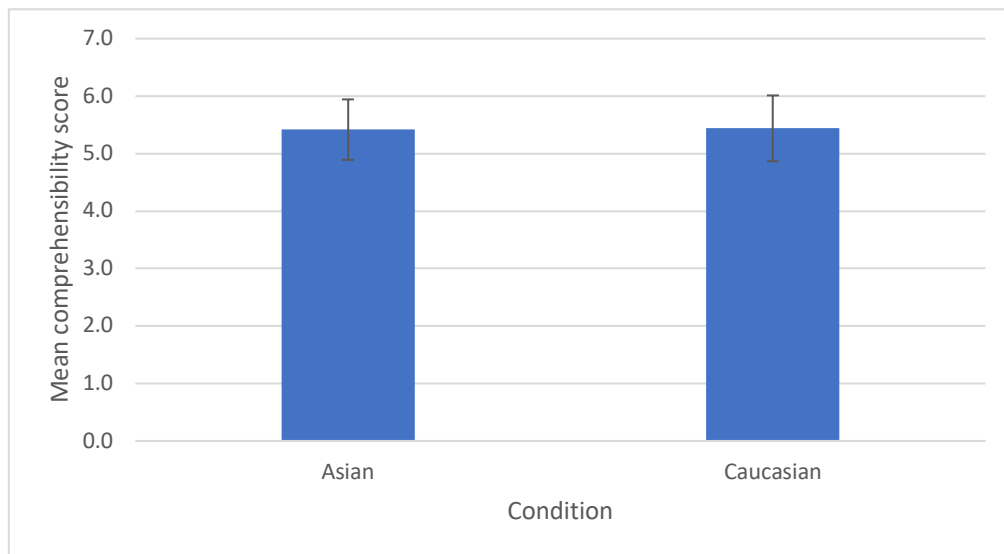
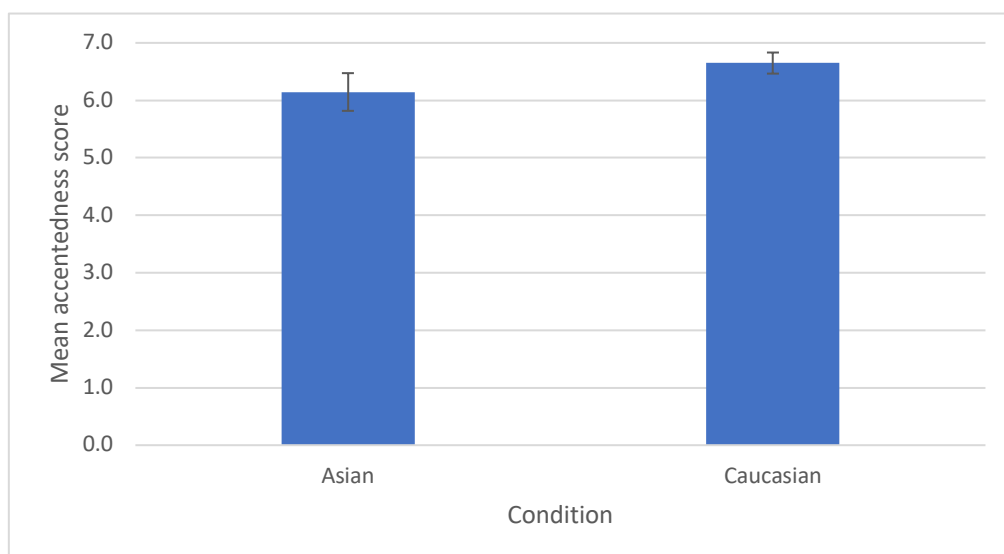
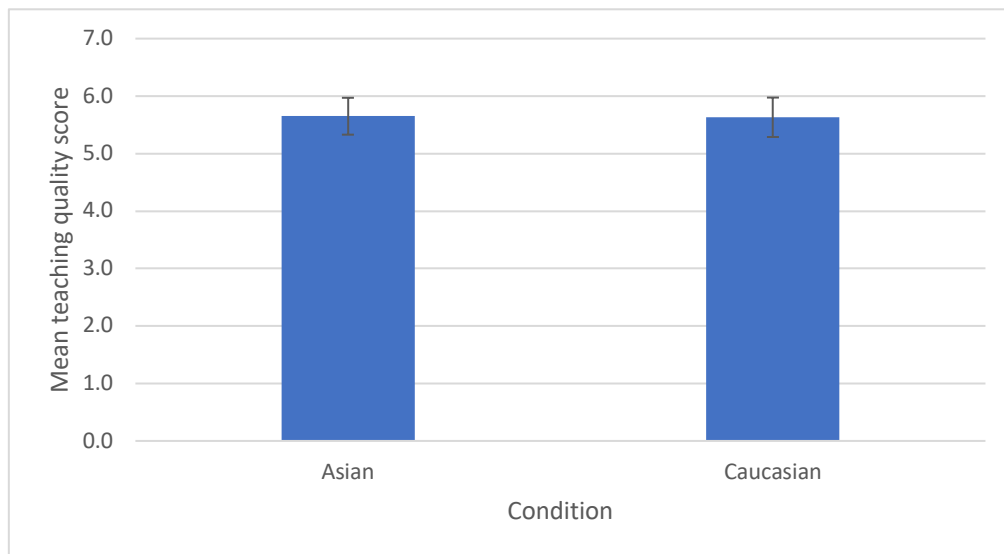
Figure S3*Mean Superiority Score by Condition**Note: Error bars: 95% CI.***Figure S4***Mean Intelligibility Score by Condition**Note: Error bars: 95% CI.*

Figure S5*Mean Comprehensibility Score by Condition*

Note: Error bars: 95% CI.

Figure S6*Mean Accentedness Scores by Condition*

Note: Low = non-native accent. Error bars: 95% CI.

Figure S7*Mean Teaching Quality Scores by Condition*

Note: Error bars: 95% CI.

**Appendix T: Interaction Effects of Knowledge of L2s Other Than English and
Condition on Dependent Variables**

Table T1

Descriptive Statistics for Social Attractiveness

Condition	Knowledge of L2s Other Than English	Mean	Std. Deviation	N
Asian	Yes	5.24	1.22	9
	No	5.49	1.01	24
	Total	5.42	1.06	33
Caucasian	Yes	5.73	1.14	9
	No	5.37	0.94	31
	Total	5.45	0.99	40
Total	Yes	5.48	1.18	18
	No	5.42	0.97	55
	Total	5.44	1.02	73

Table T2*ANOVA Summary Table for Social Attractiveness*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1.32a	3	0.44	0.42	0.742	0.02
Intercept	1609.11	1	1609.11	1523.89	0	0.96
Condition	0.48	1	0.48	0.46	0.502	0.01
Knowledge of L2s	0.04	1	0.04	0.04	0.849	0.00
Condition * Knowledge of L2s	1.24	1	1.24	1.18	0.282	0.02
Error	72.86	69	1.06			
Total	2233.20	73				
Corrected Total	74.18	72				

a R Squared = .018 (Adjusted R Squared = -.025)

Table T3*Descriptive Statistics for Superiority*

Condition	Knowledge of L2s		Mean	Std. Deviation	N
	Other Than English				
Asian	Yes		5.78	1.00	9
	No		5.80	0.83	24
	Total		5.80	0.86	33
Caucasian	Yes		5.94	0.85	9
	No		5.67	0.77	31
	Total		5.73	0.79	40
Total	Yes		5.86	0.90	18
	No		5.73	0.79	55
	Total		5.76	0.82	73

Table T4*ANOVA Summary Table for Superiority*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.59a	3	0.20	0.29	0.835	0.01
Intercept	1815.29	1	1815.29	2642.24	0	0.98
Condition	0.00	1	0.00	0.00	0.959	0.00
Knowledge of L2s	0.20	1	0.20	0.29	0.593	0.00
Condition * Knowledge of L2s	0.29	1	0.29	0.43	0.517	0.01
Error	47.41	69	0.69			
Total	2469.37	73				
Corrected Total	48.00	72				

a R Squared = .012 (Adjusted R Squared = -.031)

Table T5*Descriptive Statistics for Intelligibility*

Condition	Knowledge of L2s Other Than English	Mean	Std. Deviation	N
Asian	Yes	3.67	1.23	9
	No	3.91	1.24	23
	Total	3.84	1.22	32
Caucasian	Yes	4.56	0.73	9
	No	4.41	0.78	29
	Total	4.45	0.76	38
Total	Yes	4.11	1.08	18
	No	4.19	1.03	52
	Total	4.17	1.04	70

Table T6*ANOVA Summary Table for Intelligibility*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	6.86a	3	2.29	2.25	0.091	0.09
Intercept	912.35	1	912.35	897.62	0	0.93
Condition	6.43	1	6.43	6.33	0.014	0.09
Knowledge of L2s	0.04	1	0.04	0.04	0.85	0.00
Condition * Knowledge of L2s	0.50	1	0.50	0.49	0.485	0.01
Error	67.08	66	1.02			
Total	1292.00	70				
Corrected Total	73.94	69				

a R Squared = .093 (Adjusted R Squared = .052)

Table T7*Descriptive Statistics for Comprehensibility*

Condition	Knowledge of L2s	Mean	Std. Deviation	N
	Other Than English			
Asian	Yes	5.22	1.43	9
	No	5.49	1.53	24
	Total	5.42	1.48	33
Caucasian	Yes	5.97	2.06	9
	No	5.23	1.75	31
	Total	5.40	1.82	40
Total	Yes	5.60	1.77	18
	No	5.35	1.64	55
	Total	5.41	1.67	73

Table T8*ANOVA Summary Table for Comprehensibility*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	4.28a	3	1.43	0.50	0.681	0.02
Intercept	1622.15	1	1622.15	572.78	0	0.89
Condition	0.83	1	0.83	0.29	0.591	0.00
Knowledge of L2s	0.75	1	0.75	0.26	0.609	0.00
Condition * Knowledge of L2s	3.42	1	3.42	1.21	0.276	0.02
Error	195.41	69	2.83			
Total	2334.31	73				
Corrected Total	199.69	72				

a R Squared = .021 (Adjusted R Squared = -.021)

Table T9*Descriptive Statistics for Accentedness*

Condition	Knowledge of L2s	Mean	Std. Deviation	N
	Other Than English			
Asian	Yes	6.11	1.11	9
	No	6.16	0.87	24
	Total	6.15	0.93	33
Caucasian	Yes	6.89	0.18	9
	No	6.57	0.64	31
	Total	6.64	0.59	40
Total	Yes	6.50	0.87	18
	No	6.39	0.77	55
	Total	6.42	0.79	73

Table T10*ANOVA Summary Table for Accentedness*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	5.16a	3	1.72	2.96	0.038	0.11
Intercept	2234.81	1	2234.81	3852.74	0	0.98
Condition	4.76	1	4.76	8.21	0.006	0.11
Knowledge of L2s	0.25	1	0.25	0.44	0.511	0.01
Condition * Knowledge of L2s	0.46	1	0.46	0.79	0.377	0.01
Error	40.02	69	0.58			
Total	3050.64	73				
Corrected Total	45.18	72				

a R Squared = .114 (Adjusted R Squared = .076)

Table T11*Descriptive Statistics for Teaching Quality*

Condition	Knowledge of L2s	Mean	Std. Deviation	N
	Other Than English			
Asian	Yes	5.53	0.93	9
	No	5.69	0.91	24
	Total	5.65	0.90	33
Caucasian	Yes	5.75	1.48	9
	No	5.55	0.96	31
	Total	5.60	1.08	40
Total	Yes	5.64	1.20	18
	No	5.61	0.93	55
	Total	5.62	1.00	73

Table T12*ANOVA Summary Table for Teaching Quality*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	.51a	3	0.17	0.16	0.92	0.01
Intercept	1714.07	1	1714.07	1664.18	0	0.96
Condition	0.02	1	0.02	0.02	0.886	0.00
Knowledge of L2s	0.01	1	0.01	0.01	0.946	0.00
Condition * Knowledge of L2s	0.45	1	0.45	0.44	0.511	0.01
Error	71.07	69	1.03			
Total	2378.06	73				
Corrected Total	71.58	72				

a R Squared = .007 (Adjusted R Squared = -.036)

Appendix U: Interaction Effects of Exposure to NESs and Condition on Dependent Variables

Table U1

Descriptive Statistics for Social Attractiveness

Condition	Exposure to NESs	Mean	Std. Deviation	N
Asian	Low	5.53	1.09	7
	Medium	5.26	1.05	21
	High	5.94	1.10	5
	Total	5.42	1.06	33
Caucasian	Low	5.86	0.66	8
	Medium	5.28	1.03	21
	High	5.56	1.13	11
	Total	5.47	1.00	40
Total	Low	5.70	0.87	15
	Medium	5.27	1.03	42
	High	5.68	1.10	16
	Total	5.45	1.02	73

Table U2*ANOVA Summary Table for Social Attractiveness*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	4.10a	5	0.82	0.77	0.572	0.06
Intercept	1708.44	1	1708.44	1612.78	0	0.96
Condition	0.00	1	0.00	0.00	0.964	0.00
Exposure to NESs	3.51	2	1.76	1.66	0.198	0.05
Condition * Exposure to NESs	0.91	2	0.46	0.43	0.653	0.01
Error	70.97	67	1.06			
Total	2241.88	73				
Corrected Total	75.07	72				

a R Squared = .055 (Adjusted R Squared = -.016)

Table U3*Descriptive Statistics for Superiority*

Condition	Exposure to NESs	Mean	Std. Deviation	N
Asian	Low	6.18	0.82	7
	Medium	5.56	0.84	21
	High	6.26	0.75	5
	Total	5.80	0.86	33
Caucasian	Low	6.23	0.43	8
	Medium	5.65	0.78	21
	High	5.65	0.95	11
	Total	5.76	0.80	40
Total	Low	6.21	0.62	15
	Medium	5.60	0.81	42
	High	5.84	0.91	16
	Total	5.78	0.82	73

Table U4*ANOVA Summary Table for Superiority*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	5.51a	5	1.10	1.71	0.144	0.11
Intercept	1929.84	1	1929.84	2997.18	0	0.98
Condition	0.34	1	0.34	0.53	0.471	0.01
Exposure to NESs	4.45	2	2.23	3.46	0.037	0.09
Condition * Exposure to NESs	1.31	2	0.65	1.01	0.368	0.03
Error	43.14	67	0.64			
Total	2486.51	73				
Corrected Total	48.66	72				

a R Squared = .113 (Adjusted R Squared = .047)

Table U5*Descriptive Statistics for Intelligibility*

Condition	Exposure to NESs	Mean	Std. Deviation	N
Asian	Low	3.14	1.77	7
	Medium	4.10	0.97	20
	High	3.80	1.10	5
	Total	3.84	1.22	32
Caucasian	Low	4.13	0.99	8
	Medium	4.50	0.76	20
	High	4.70	0.48	10
	Total	4.47	0.76	38
Total	Low	3.67	1.45	15
	Medium	4.30	0.88	40
	High	4.40	0.83	15
	Total	4.19	1.04	70

Table U6*ANOVA Summary Table for Intelligibility*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	13.15a	5	2.63	2.74	0.026	0.18
Intercept	889.10	1	889.10	926.27	0	0.94
Condition	7.80	1	7.80	8.12	0.006	0.11
Exposure to NESs	5.01	2	2.50	2.61	0.081	0.08
Condition * Exposure to NESs	1.24	2	0.62	0.64	0.529	0.02
Error	61.43	64	0.96			
Total	1301.00	70				
Corrected Total	74.59	69				

a R Squared = .176 (Adjusted R Squared = .112)

Table U7*Descriptive Statistics for Comprehensibility*

Condition	Exposure to NESs	Mean	Std. Deviation	N
Asian	Low	5.00	1.83	7
	Medium	5.32	1.45	21
	High	6.40	0.68	5
	Total	5.42	1.48	33
Caucasian	Low	5.84	1.79	8
	Medium	4.92	1.87	21
	High	6.36	1.31	11
	Total	5.50	1.79	40
Total	Low	5.45	1.80	15
	Medium	5.12	1.66	42
	High	6.38	1.13	16
	Total	5.46	1.65	73

Table U8*ANOVA Summary Table for Comprehensibility*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	22.66a	5	4.53	1.75	0.135	0.12
Intercept	1751.54	1	1751.54	676.20	0	0.91
Condition	0.25	1	0.25	0.10	0.758	0.00
Exposure to NESs	16.52	2	8.26	3.19	0.048	0.09
Condition * Exposure to NESs	4.30	2	2.15	0.83	0.441	0.02
Error	173.55	67	2.59			
Total	2374.31	73				
Corrected Total	196.21	72				

a R Squared = .115 (Adjusted R Squared = .049)

Table U9*Descriptive Statistics for Accentedness*

Condition	Exposure to NESs	Mean	Std. Deviation	N
Asian	Low	6.34	0.86	7
	Medium	5.98	0.93	21
	High	6.56	0.98	5
	Total	6.15	0.93	33
Caucasian	Low	6.83	0.36	8
	Medium	6.52	0.71	21
	High	6.78	0.42	11
	Total	6.66	0.59	40
Total	Low	6.60	0.67	15
	Medium	6.25	0.86	42
	High	6.71	0.62	16
	Total	6.43	0.80	73

Table U10*ANOVA Summary Table for Accentedness*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	7.17a	5	1.43	2.50	0.039	0.16
Intercept	2327.40	1	2327.40	4066.01	0	0.98
Condition	2.38	1	2.38	4.15	0.046	0.06
Exposure to NESs	2.43	2	1.21	2.12	0.128	0.06
Condition * Exposure to NESs	0.27	2	0.13	0.23	0.792	0.01
Error	38.35	67	0.57			
Total	3058.68	73				
Corrected Total	45.52	72				

a R Squared = .157 (Adjusted R Squared = .095)

Table U11*Descriptive Statistics for Teaching Quality*

Condition	Exposure to NESs	Mean	Std. Deviation	N
Asian	Low	5.68	0.87	7
	Medium	5.50	0.96	21
	High	6.24	0.44	5
	Total	5.65	0.90	33
Caucasian	Low	5.92	0.84	8
	Medium	5.39	1.10	21
	High	5.90	1.24	11
	Total	5.63	1.10	40
Total	Low	5.81	0.83	15
	Medium	5.44	1.02	42
	High	6.01	1.05	16
	Total	5.64	1.01	73

Table U12*ANOVA Summary Table for Teaching Quality*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	4.97a	5	0.99	0.97	0.441	0.07
Intercept	1833.29	1	1833.29	1793.77	0	0.96
Condition	0.08	1	0.08	0.07	0.786	0.00
Exposure to NESs	4.59	2	2.30	2.25	0.114	0.06
Condition * Exposure to NESs	0.62	2	0.31	0.30	0.739	0.01
Error	68.48	67	1.02			
Total	2396.20	73				
Corrected Total	73.45	72				

a R Squared = .068 (Adjusted R Squared = -.002)

Appendix V: Interaction Effects of Exposure to NNEs and Condition on Dependent Variables

Table V1

Descriptive Statistics for Social Attractiveness

Condition	Exposure to NNEs	Mean	Std. Deviation	N
Asian	Low	5.51	1.00	9
	Medium	5.31	1.11	20
	High	5.75	1.15	4
	Total	5.42	1.06	33
Caucasian	Low	5.83	1.04	10
	Medium	5.25	1.01	23
	High	5.68	0.78	8
	Total	5.48	0.99	41
Total	Low	5.68	1.01	19
	Medium	5.28	1.04	43
	High	5.70	0.86	12
	Total	5.45	1.01	74

Table V2*ANOVA Summary Table for Social Attractiveness*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	3.49a	5	0.70	0.66	0.654	0.05
Intercept	1635.04	1	1635.04	1551.60	0	0.96
Condition	0.05	1	0.05	0.05	0.823	0.00
Exposure to NNEs	2.82	2	1.41	1.34	0.269	0.04
Condition * Exposure to NNEs	0.51	2	0.26	0.24	0.785	0.01
Error	71.66	68	1.05			
Total	2274.53	74				
Corrected Total	75.14	73				

a R Squared = .046 (Adjusted R Squared = -.024)

Table V3*Descriptive Statistics for Superiority*

Condition	Exposure to NNEs	Mean	Std. Deviation	N
Asian	Low	5.87	0.73	9
	Medium	5.69	0.95	20
	High	6.14	0.76	4
	Total	5.80	0.86	33
Caucasian	Low	5.96	0.71	10
	Medium	5.64	0.83	23
	High	5.82	0.81	8
	Total	5.75	0.79	41
Total	Low	5.92	0.70	19
	Medium	5.66	0.88	43
	High	5.93	0.78	12
	Total	5.77	0.82	74

Table V4*ANOVA Summary Table for Superiority*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1.53a	5	0.31	0.44	0.819	0.03
Intercept	1815.67	1	1815.67	2606.80	0	0.98
Condition	0.12	1	0.12	0.18	0.674	0.00
Exposure to NNEs	1.33	2	0.67	0.96	0.389	0.03
Condition * Exposure to NNEs	0.28	2	0.14	0.20	0.818	0.01
Error	47.36	68	0.70			
Total	2514.45	74				
Corrected Total	48.90	73				

a R Squared = .031 (Adjusted R Squared = -.040)

Table V5*Descriptive Statistics for Intelligibility*

Condition	Exposure to NNEs	Mean	Std. Deviation	N
Asian	Low	3.44	1.51	9
	Medium	4.00	1.05	19
	High	4.00	1.41	4
	Total	3.84	1.22	32
Caucasian	Low	4.50	0.85	10
	Medium	4.33	0.80	21
	High	4.75	0.46	8
	Total	4.46	0.76	39
Total	Low	4.00	1.29	19
	Medium	4.18	0.93	40
	High	4.50	0.91	12
	Total	4.18	1.03	71

Table V6*ANOVA Summary Table for Intelligibility*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	9.73a	5	1.95	1.95	0.098	0.13
Intercept	912.62	1	912.62	914.19	0	0.93
Condition	6.67	1	6.67	6.68	0.012	0.09
Exposure to NNEs	1.15	2	0.57	0.58	0.565	0.02
Condition * Exposure to NNEs	1.75	2	0.88	0.88	0.421	0.03
Error	64.89	65	1.00			
Total	1317.00	71				
Corrected Total	74.62	70				

a R Squared = .130 (Adjusted R Squared = .064)

Table V7*Descriptive Statistics for Comprehensibility*

Condition	Exposure to NNEs	Mean	Std. Deviation	N
Asian	Low	5.11	1.63	9
	Medium	5.40	1.52	20
	High	6.19	0.83	4
	Total	5.42	1.48	33
Caucasian	Low	5.10	1.89	10
	Medium	5.25	1.98	23
	High	6.41	0.77	8
	Total	5.44	1.81	41
Total	Low	5.11	1.72	19
	Medium	5.32	1.76	43
	High	6.33	0.76	12
	Total	5.43	1.66	74

Table V8*ANOVA Summary Table for Comprehensibility*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	12.69a	5	2.54	0.91	0.48	0.06
Intercept	1646.92	1	1646.92	590.97	0	0.90
Condition	0.01	1	0.01	0.00	0.967	0.00
Exposure to NNEs	10.46	2	5.23	1.88	0.161	0.05
Condition * Exposure to NNEs	0.31	2	0.15	0.06	0.947	0.00
Error	189.50	68	2.79			
Total	2383.31	74				
Corrected Total	202.19	73				

a R Squared = .063 (Adjusted R Squared = -.006)

Table V9*Descriptive Statistics for Accentedness*

Condition	Exposure to NNEs	Mean	Std. Deviation	N
Asian	Low	6.18	0.74	9
	Medium	6.09	1.01	20
	High	6.35	1.05	4
	Total	6.15	0.93	33
Caucasian	Low	6.58	0.63	10
	Medium	6.65	0.61	23
	High	6.73	0.49	8
	Total	6.65	0.58	41
Total	Low	6.39	0.69	19
	Medium	6.39	0.86	43
	High	6.60	0.70	12
	Total	6.42	0.79	74

Table V10*ANOVA Summary Table for Accentedness*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	4.96a	5	0.99	1.67	0.155	0.11
Intercept	2189.60	1	2189.60	3671.65	0	0.98
Condition	2.64	1	2.64	4.43	0.039	0.06
Exposure to NNEs	0.25	2	0.12	0.21	0.814	0.01
Condition * Exposure to NNEs	0.13	2	0.06	0.11	0.899	0.00
Error	40.55	68	0.60			
Total	3099.64	74				
Corrected Total	45.52	73				

a R Squared = .109 (Adjusted R Squared = .044)

Table V11*Descriptive Statistics for Teaching Quality*

Condition	Exposure to NNEs	Mean	Std. Deviation	N
Asian	Low	5.75	0.65	9
	Medium	5.50	1.03	20
	High	6.17	0.62	4
	Total	5.65	0.90	33
Caucasian	Low	5.89	0.94	10
	Medium	5.44	1.21	23
	High	5.86	0.88	8
	Total	5.63	1.09	41
Total	Low	5.82	0.79	19
	Medium	5.47	1.11	43
	High	5.96	0.79	12
	Total	5.64	1.00	74

Table V12*ANOVA Summary Table for Teaching Quality*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	3.55a	5	0.71	0.69	0.632	0.05
Intercept	1762.55	1	1762.55	1714.60	0	0.96
Condition	0.08	1	0.08	0.08	0.784	0.00
Exposure to NNEs	3.35	2	1.68	1.63	0.204	0.05
Condition * Exposure to NNEs	0.34	2	0.17	0.17	0.849	0.01
Error	69.90	68	1.03			
Total	2427.06	74				
Corrected Total	73.45	73				

a R Squared = .048 (Adjusted R Squared = -.022)