# EFFECTIVE COMMUNICATION BETWEEN AMERICAN UNDERGRADUATES AND INTERNATIONAL TEACHING ASSISTANTS

A Dissertation

by

XUEYAN HU

Submitted to the Office of Graduate and Professional Studies of

Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

| | |
|---|---|
| Chair of Committee, | Li-Jen Kuo |
| Committee Members, | Zohreh Eslami |
| | Wen Luo |
| | Monica Neshyba |
| Head of Department, | Michael Anthony de Miranda |

August 2020

Major Subject: Curriculum and Instruction

## ABSTRACT

The trend of non-native English speakers (NNESs) outnumbering native English speakers (NESs) (Crystal, 2012) has challenged the privilege of native varieties of English. With this trend, more and more researchers have realized that successful communication relies on both sides of interlocutors - NNESs as speakers and NESs as listeners. This dissertation investigates effective ways to enhance intelligible communication between NNESs and NESs from the angle of NESs as listeners. It aims to highlight the importance of improving NESs' perception of NNESs' English speech in broader efforts related to diversity, equity, and inclusion.

This dissertation includes three independent but relevant studies regarding the intelligible communication between American undergraduates (UGs) and international teaching assistants (ITAs). The first study is a meta-analytic review. This review used a meta-analytic methods-Robust Variance Estimation to examine the effectiveness of interventions on improving the intelligibility of NNESs' English speech, and to identify factors affecting the effectiveness of these interventions. The second study investigated how different types of intervention activities aiming to ameliorate UGs' biased language ideology could help to improve their perceptions of ITAs' foreign-accented English speech and teaching ability. The third study used the data collected from the pre-intervention survey about UGs' previous experience with ITAs, their openness to foreign-accented English, as well as their perceptions of ITAs' English speech to identify significant factors affecting their perceptions of ITAs' English speech and teaching ability.

Based on the findings of this dissertation, implications about how to improve UGs' attitudes towards ITAs in educational institutions were discussed. Recommendations about how to address methodological issues for future studies were also made at the end of studies.

## ACKNOWLEDGEMENTS

I would like to express my heartiest gratitude to my committee chair, Dr. Li-Jen Kuo, and my academic advisor, Dr. Quentin Dixon, who provided tremendous and invaluable comments and suggestions on the drafts of this dissertation. Without their guidance and persistent encouragement in the past five years, I could not have grown up into an independent researcher to complete this dissertation.

I would also like to thank my committee members, Dr. Zohreh Eslami, Dr. Wen Luo, and Dr. Monica Neshyba, who provided me with helpful insights in the design of the experimental studies and the revision of the drafts of this dissertation.

This dissertation would not have been possible without the financial support of the Marie M. and James H. Galloway Foundation from the Center for Teaching Excellence (CTE) at Texas A&M University. I am especially indebted to the English proficiency consultants from CTE, Mr. Brandon Cooper and Ms. Ginessa Payne, who have been supportive of the data collection for the dissertation in the past two years. My sincere thanks also go to Dr. Ra'sheedah Richardson and Dr. Debra Fowler. Without their generous support and trust for my research, I would not have completed this dissertation and my graduate work.

My special thanks also go to my friends and colleagues. I am grateful for three TAMU undergraduates, Darcy Rydl, Yijie Zhu, and Sung Kim, who joined my Aggie Research Leadership program and helped with my data collection and data entry. Also, I thank my former colleagues at China Three Gorges University, Gaosheng Ni, Min Zhang, Qijia Tan, Wei Sun and Jing He, who offered generous help for my graduate work.

Finally, I wish to express my deep sense of gratitude to my beloved family members: my husband Phil, my sons Ben and Kevin, and my Mom. They all keep me going with their continuous and unparalleled love, help and support. I could not have made it without them.

## CONTRIBUTORS AND FUNDING SOURCES

### Contributors

This work was supervised by a dissertation committee consisting of Dr. Li-Jen Kuo, Dr. Zohreh Eslami, and Dr. Monica Neshyba of the Department of Teaching, Learning, and Culture, and Dr. Wen Luo of the Department of Educational Psychology.

All work conducted for the dissertation was completed by the student independently.

### Funding Sources

Graduate Study was supported by an assistantship from the Marie M. and James H. Galloway Foundation at the Center for Teaching Excellence at Texas A&M University.

## TABLE OF CONTENTS

Page

## LIST OF FIGURES

Page

## LIST OF TABLES

Page

# 1. INTRODUCTION

International graduate students currently account for up to 44% of students in U.S. graduate programs (Okahana & Zhou, 2019). U.S. universities routinely rely on them to serve as instructors in different disciplines. Based on this need, a lot of educational institutions have resorted to improve effective communication between American undergraduates (UGs) and international teaching assistants (ITAs).

In the field of sociolinguistics, American UGs represent native English speakers (NESs) and ITAs represent non-native English speakers (NNESs). The following sections will review previous literature with interventions focusing on either NESs or NNESs to improve successful communication between them.

## Interventions Focusing on Non-native English Speakers

Interventions for non-native English speakers (NNES) have focused on two areas: a) pronunciation and perception of specific sounds**,** and b) communicative strategies. Both types of interventions aim to improve NNES's oral proficiency and enhance the intelligibility of their speech by NESs.

### Instruction on Pronunciation and Perception

Since pronunciation is viewed as "possibly the greatest single barrier to successful communication" by NNESs (Jenkins, 2002, p. 83), many English language teachers have resorted to helping NNESs to improve their pronunciations by instructing about segmentals (e.g., Huthaily, 2008; Liu & Fu, 2011). For example, in Saito's study (Saito, 2019), seventeen Japanese English speakers received four hours of instruction with the specific segmental feature / ɹ/. Compared with students who did not receive any instruction about the features of / ɹ/, students receiving explicit instruction demonstrated improved pronunciation of this sound in both free and controlled speaking activities.

In addition to formative instruction about segmental and suprasegmental features, other studies focus on NNESs' recognition of English speech sounds so as to improve their production performance. These studies have observed participants' successful perception of the target sounds, but participants' production improvements have not been reliably identified (Sakai & Moorman, 2018). For example, in Iverson, Pinet, & Evans' study (2012), twenty adult French NNESs were provided auditory training for English vowels in the form of minimal pairs. These participants demonstrated significant improvement in vowel identification and discrimination tasks, as well as word-reading tasks. In Aliaga-Garcia & Mora's study (2009), 18 adult Spanish NNESs were provided six training sessions about perception and production tasks related to English oral stop sounds /p t k b d g/. However, these participants did not show any improvements in their production of these sounds in mini-dialogue tasks. The conflicting results of the perception interventions on NNESs' speech production were reported to be caused by factors such as length of training, type of training task, and dimensions of the targeted sounds (Aliaga-Garcia & Mora, 2009; Sakai & Moorman, 2018). For example, Iverson, et al. (2012)'s intervention spanned across eight sessions whereas Aliaga-Garcia et al (2009)'s study only included 12-hour training in total.

**Instruction on Communicative Strategies**

Communicative strategies can be operationally defined as strategies used by communicators to efficiently convey their intended meaning clearly. These strategies include paraphrasing, substitution, and asking for clarification (Kaur, 2010, Matsumoto, 2011 & Seidlhofer, 2009). While the importance of communicative strategies for ITAs has been well recognized by researchers, little intervention research has been conducted. To my knowledge, the only intervention study on communicative strategies for ITAs to date was conducted by Gorsuch (2018), who taught four ITAs strategies of using discourse intonations such as pausing and prosodic system of

English to communicate. Gorsuch's (2018) findings on ITAs' discourse performance showed that ITAs could acquire knowledge about discourse intonation very quickly. However, Gorsuch (2018) further specified that ITAs were not able to apply these strategies in their actual communication, because discourse intonation is a procedural knowledge with layers of development within it. Gorsuch (2018) suggested that ITAs' speech fluency should be developed before they can gain improvements in their use of discourse intonations. Therefore, thought groups-a group of approximately two to five words that form a unit of Meaning (Gorsuch, Meyers, Pickering & Griffee, 2011), which helps to increase ITAs' speech fluency, should be taught before prominence and tone choices. The best way to teach thought groups is to develop ITAs' awareness of common signals within English such as pausing before and after a phrase and emphasizing focused words within the phrase.

Gorsuch's (2018) findings revealed that it took time for ITAs to develop communicative skills. This nature of communicative competence development makes it difficult to capture effectiveness of training in most of ITA training programs, because these programs only offer one-semester long courses. However, more studies similar to the one in Gorsuch's article (2018) are needed to examine how ITAs develop their communicative skills as well as the challenges faced by them during the developmental process so as to give more recommendations for instructional practice. For example, future research might explore how the speakers' perceptions of communicative challenges influence these developmental processes and whether speakers' views are reflective of actual problems experienced in communication or rather contribute to the problems (Kennedy & Trofimovich, 2008).

## Interventions Focusing on Native English Speakers

### Improving Listening Ability

Having realized that effective communication relies on both listeners and speakers, many researchers have applied interventions to improve NESs' listening ability. These interventions

can be categorized into three types: explicit instruction, implicit training, and a combination of the two. In Derwing, Rossiter, and Munro's experiment (2002), first-year NESs received explicit instruction about the features of the speech. This study reported a significant improvement in NESs' ability to understand foreign accents, but researchers did not observe any significant improvements in NESs' comprehension or sentence transcription. Similarly, in Villarreal's (2013) study, undergraduates (UGs) in a southeast college were taught the distinctive phonetic features in their international professor's Indian-accented English. This study revealed that while students' attitudes towards Indian-accented English had been positively changed, their comprehension of this foreign-accented English had not improved in comparison to the control group. The researcher further proposed a UG course to develop their listening ability as well as their critical perspective towards language variations.

The typical way for NESs to implicitly learn foreign-accented English is to be exposed to foreign-accented English either from one language background (Bradlow and Bent, 2008; Clarke and Garrett, 2004; Reinisch and Weber, 2012; Sidaras, Alexander & Nygaard, 2009; Sumner, 2011) or from multiple language backgrounds (Baese-Berk, Bradlow & Wright, 2013). A common finding from these studies is that simply interacting with a speaker with an unfamiliar accent is a promising way to help NESs to accommodate these variations to improve their ability of word recognition. More impressively, using a multi-layered research design, Bradlow and Bent (2008) even noticed the generalizability of the improved speech comprehensibility from the speakers they were exposed to, to the speakers they had not been exposed to.

Motivated by the mixed results of explicit and implicit training for improving NESs' listening ability, a recent study conducted by Lindemann, Campbell, Litzenberg and Subtirelu (2016) developed an intervention to examine the combined effects of explicit and implicit training of foreign-accented English for NESs. With exposure to Korean-accented English through sentence transcription (implicit) and learning about linguistic differences between

Korean-accented English and American English (explicit), UGs in a southern university in the U.S. were found to show improvement in their word recognition ability in listening to brief sentences or phrases spoken in Korean accented-English. However, the improvement in participants' comprehension of a longer passage was not observed in this study.

**Improving Attitudes Towards Foreign-accented English**

Research has revealed two effective approaches to improving NESs' attitudes towards foreign-accented English. The first one is intergroup contact where NESs connect with NNESs through unstructured or structured activities. Guided by Contact Theory (Allport, 1954), mentioned previously, many studies designed different structured activities meeting the optimal conditions for intergroup contact (Kang, Rubin, Lindemann, 2015; Staples, Kang, & Wittner, 2014). For example, in Kang, Rubin and Lindemann's study (2015), they designed a problem-solving game for both international teaching assistants (ITAs) and UGs to play together. This study revealed greater improvement for UGs in their evaluations of ITAs' instructional competence ratings and comprehensibility ratings. In this study, they further observed the generalizability of this improvement for ITAs that UGs have never contacted.

The second effective approach to reducing prejudice towards foreign-accented English is to invite individuals to engage in perspective-taking activities with outgroup members. The perspective-taking techniques include watching a video about outgroup members (Dovidio et al., 2004; Weyant, 2007; Manohar & Appiah, 2016), role playing about a situation that NNESs might encounter due to their limited English language proficiency (Hansen, Rakic & Steffens, 2014; Madera, Neal & Dawson, 2011), and structured discussions about NNES' difficulties in the process of learning English (Derwing, Rossiter & Munro, 2002; Staples, Kang & Wittner 2014; Villarreal, 2013). In these studies, participants' attitude change was identified only from the results in self-reported sources such as surveys or interviews (e.g., Dovidio et al, 2004; Hansen, Rakic & Steffens, 2014; Madera, Neal & Dawson, 2011; Manohar & Appiah, 2016).

Measurements about participants' comprehension of the content of the speech samples were not employed in these studies.

## Purpose of the Dissertation

Nowadays, more and more scholars have realized that the improvement of communicative competence has often been narrowly tied to the NNESs in the multilingual context. They argued that the development of skills and competence, as well as the change of attitudes for NESs who engage in communication with NNESs, are of equal importance (Subtirelu & Lindemann, 2016).

Rubin (1992) proposed that the negative attitudes towards ITAs create a mental barrier that inhibits students' ability to understand their ITAs. This deteriorated understanding naturally impairs their academic performance. For example, in Villarreal's study (2013), several participants reported dropping at least one class with an ITA, often claiming that ITAs' accents had played a role in these decisions. This study also found that accents were implicated in altering students' academic plans on a larger scale.

In view of the importance of the role of UGs' attitudes towards ITAs, this dissertation investigated effective ways to enhance intelligible communication between NNESs and NESs from the angle of NESs as listeners. It aims to highlight the importance of improving NESs' perception of NNESs' English speech in broader efforts related to diversity, equity, and inclusion.

## 2. EFFECTIVENESS OF INTERVENTIONS ON IMPROVING NON-NATIVE ENGLISH SPEAKERS' INTELLIGIBILITY: A META-ANALYTIC STUDY

### Introduction

The trend of non-native English speakers (NNESs) outnumbering native English speakers (NESs) (Crystal, 2012) has challenged the privilege of native varieties of English. It has redirected the research of Second Language Acquisition theoretically and pedagogically. Some scholars have proposed a concept of "English as a Lingua Franca" (Firth, 1996; Jenkins, 2002) and argued that the priority of teaching English should be to maximize the mutual intelligibility between NNESs and NESs (Jenkins, 2002). Instead of conforming to the "nativeness principle", which holds that native-like pronunciation and use of English is the ultimate learning goal for English language learners (ELLs) (Levis, 2005), more and more researchers have set the "intelligibility principle" as an important criterion to evaluate NNESs' speech.

By definition, intelligibility refers to "intelligible production and felicitous interpretation of English" (Nelson, 1995, p. 274). This definition actually entails three sub-constructs: intelligibility, comprehensibility and interpretability. In this article, I used the term "overall intelligibility" to cover the three sub-constructs and used "intelligibility" to indicate intelligibility as one of the sub-constructs. Intelligibility is defined as "the extent to which a speaker's message is actually understood by a listener" (Munro & Derwing, 1995a, p. 76). It is usually measured by asking listeners either to transcribe what they hear to check the accuracy of their transcription (Munro & Derwing, 1995b) or to answer some comprehension questions in the form of multiple choice or cloze test. Comprehensibility refers to the listener's ability to understand the meaning of the word or utterance in its given context. It is usually measured by a five-or seven-point Likert scale to check how easy or difficult the speech samples are to understand (e.g., Trofimovich, Lightbown, Halter & Song, 2009; Galante & Thomson, 2017). Interpretability refers to the ability of the listener to understand the speaker's intentions behind the word or

utterance. Interpretability is related to pragmatics and is the most difficult level of speech understanding for listeners (Smith & Nelson, 2019). Its implicit nature makes it difficult to measure (Levis, 2006; Pickering, 2006).

Informed by the concept of intelligibility explained above, more and more researchers have realized that successful communication relies on both interlocutors - NNESs and NESs. On the one hand, NNESs need to fulfill the minimal phonological requirements to be intelligible for effective communication, and many studies have extensively explored methods to improve NNEs' pronunciation (e.g., Kang, Rubin, & Lindemann, 2015; Staples, Kang, & Wittner, 2014). On the other hand, NESs' perceptions of NNESs' foreign-accented English play an important role in communication with NNESs. A growing number of researchers have attempted to either improve NESs' knowledge and familiarity about different varieties of English or improve their attitudes towards NNESs' foreign-accented English (e.g., Galante & Thomson, 2017; Tanner & Landon, 2009).

The current review takes the initiative to meta-analytically synthesize experimental studies which aimed to facilitate the intelligibility of NNESs' foreign-accented English from the angle of either speakers or listeners. This meta-analysis will first report the overall effectiveness of these experimental studies. It will then examine how key factors, such as intervention features and assessment methods, affect the effectiveness of these interventions.

## Literature Review

### Previous Meta-Analytic Reviews

Regarding meta-analytic reviews of interventions focusing on NESs' speaking ability, all of them investigate second language pronunciation instruction (PI). Lee, Jang, & Plonsky (2015) identified in their meta-analysis large effects of PI for both within-group (d=0.89) and between-group (d=0.80) comparisons. This study also indicated that PI with longer interventions, with treatments providing feedback, and with more controlled outcome measures (i.e., requiring a

fixed response from all participants) often produced larger effect sizes. However, the effects of PI were not affected by treatment providers (teacher, researcher, teacher-researcher, or computer), outcome measures (vowels, consonants, stress, intonation, segmentals, suprasegmentals, or rhythm), and use of technology. In another meta-analysis conducted by Mahdi and Khateeb (2019), computer-assisted PI produced medium effect sizes (d=0.68) compared with traditional methods of pronunciation teaching. Mahdi and Khateeb (2019) also revealed that computer-assisted PI is more beneficial for beginner and intermediate learners than advanced learners, and for university students than K-12 school students.

While both of the meta-analyses indicated medium-to-large overall effectiveness for PI, the results of the evaluation are based on the specific constructs of pronunciation, such as accurate pronunciation of specific vowels or consonants. These evaluation criteria demonstrated an implicit focus on the mastery of nativelike pronunciation. This is not appropriate, as the primary purpose of PI should be to help learners become more understandable (Thomson & Derwing, 2015).

Plonsky and colleagues' recent meta-analysis extended their previous one by adopting a new framework for conceptualizing measures of second language PI (Saito & Plonsky, 2019). In addition to specific constructs of pronunciation, four global measures - comprehensibility, intelligibility, perceived fluency, and accentedness - were added to evaluate the effectiveness of second language PI instruction. Similar to the findings in their previous meta-analysis (Lee, Jang, & Plonsky, 2015), when measures about the specific constructs of pronunciation were used, medium effect sizes were identified for both between-group (d=0.68) and within-group comparisons (d=0.73). However, this meta-analysis did not find any significant effects from the PI instruction for between-group comparisons when the global measures were used. This disparity might be attributed to the lack of correlations among all four global measures. While significant correlations among some of these measures (e.g. among comprehensibility,

accentedness, and intelligibility (Munro & Dewing, 1995a), and among comprehensibility and fluency (Derwing, Rossiter, Munro, & Thomson, 2004), have been reported, significant correlations among all the four measures have not been identified in one single study. Therefore, it may not be valid to cluster these four measures as a single construct to evaluate the effectiveness of second language PI.

Unfortunately, none of the existing meta-analyses investigated the overall effectiveness of interventions on improving NESs' perception of NNESs' speech. The existing meta-analyses are confined to examining the extent to which interventions such as contact activities can help to reduce NESs' prejudice against NNESs from the socio-psychological perspectives (e.g., Lemmer & Wagner, 2015; Miles & Crisp, 2014; Pettigrew & Tropp, 2006 & 2008). The growing number of interventions focusing on NESs' perceptions of NNESs' English speech made it a strong case for a meta-analytic review to examine their overall effectiveness and factors affecting the effectiveness.

**Current Meta-Analysis**

The current meta-analysis extends the previous meta-analytic reviews in two aspects. First, it includes studies focusing on either speakers (i.e., NNESs) or listeners (i.e, NESs) to evaluate the effectiveness of interventions on improving the intelligibility of NNESs' English speaking. Second, a more valid global measure - overall intelligibility - is used for the evaluation. That means the effectiveness of interventions will be examined based on how much they can improve the ability of the listeners to recognize individual words or utterances (i.e., intelligibility) or to understand the meaning of the word or utterance in its given context (i.e., comprehensibility).

Meanwhile, this meta-analysis will examine how the effect of interventions varies across the sub-constructs of intelligibility (i.e., intelligibility and comprehensibility), speech task type (i.e., controlled vs. spontaneous), and length of rating scales (short vs. long). Since overall

intelligibility consists of two aspects of verbal communication - how intelligible and how easily a person's speech can be understood, it is essential to examine how each sub-construct impacts the effectiveness of the intervention. Also, previous research has shown that when measuring the overall intelligibility, the length of Likert scales also impacts the result of the overall intelligibility. For example, Flege and colleagues found that finer scales were more useful than restrictive scales for listeners to capture distinctions in the speech (Flege & Fletch, 1992). A 9- or 11-point interval scales are necessary to prevent ceiling effect because a shorter scale (e.g., 5-point or 7-point) was not able to reflect the magnitude of differences in pronunciation especially between L2 learners and NEs perceived by some sensitive raters (Isaacs & Thomson, 2013; Southwood & Flege, 1999). While there are growing numbers of research using scales to measure intelligibility, there is a dearth of research clearly operationalizing constructs in rating scales (Isaacs & Thomson, 2013). It is of great importance for the present meta-analysis to examine how different length of scales would impact the effectiveness of the interventions in terms of NNESs' intelligibility.

The research questions are as follows:

1. What are the effect sizes of interventions on improving the intelligibility of NNESs' speech including the overall effect size of both types of interventions and the effect sizes of each type respectively?
2. How does the effect of interventions vary across the sub-constructs of intelligibility (i.e., intelligibility and comprehensibility), task type (i.e., controlled vs. spontaneous), and rating scales (short vs. long)?

## Methods

### Study Retrieval

Electronic searches were conducted through the database such as ERIC, PsycINFO, ProQuest Dissertations & Theses Global, Linguistics and Language Behavior Abstracts,

Education Source. Based on the existing literature, key items used in the search included: comprehensibility, intelligibility, language proficiency, native English speaker, non-native English speaker. The flow of the retrieval process is summarized in Figure 1. Finally, 23 articles were identified and included in this meta-analysis.

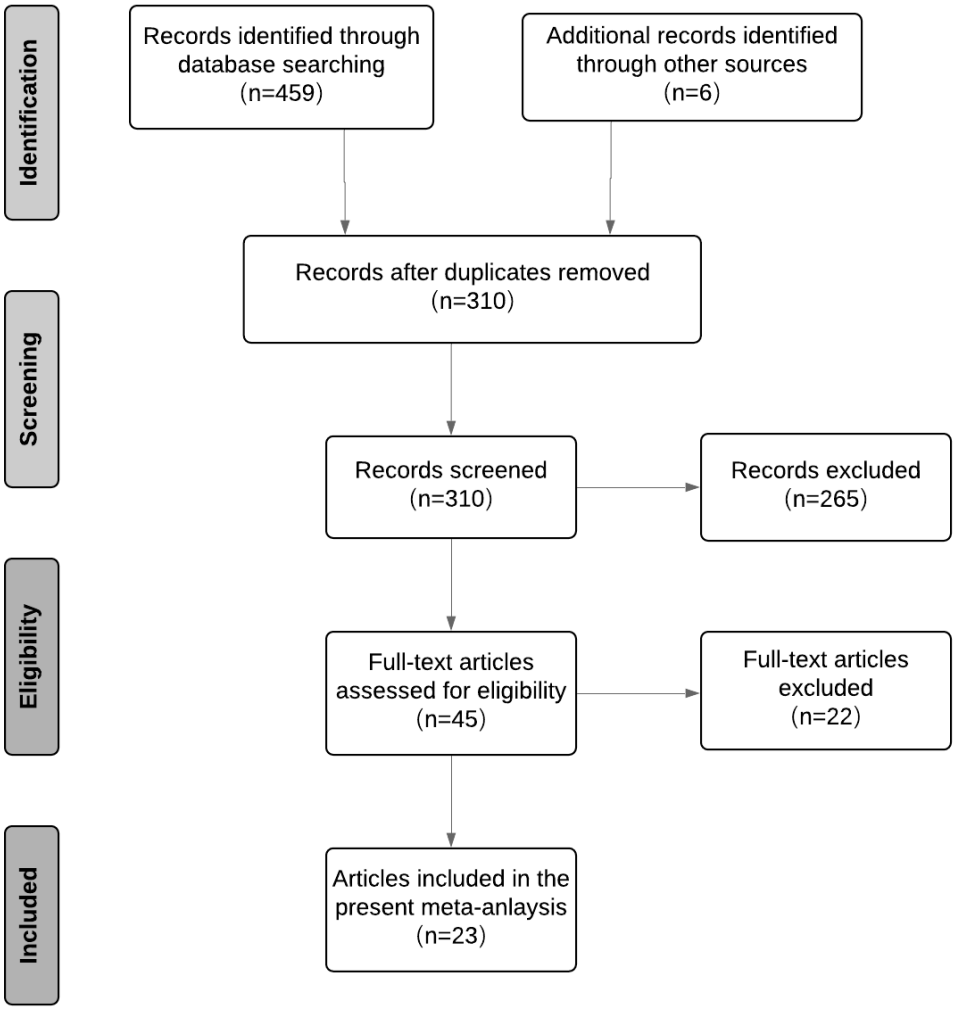


**Figure 1.** PRISMA chart for the study retrieval process

## Inclusion and Exclusion Criteria

Studies had to meet the following criteria to be included in this review: (a) had control-experiment research design; (b) had a comparison group in which participants did not receive any type of intervention or were only instructed in a business-as-usual way; (c) included at least

one measure assessing the intelligibility of non-native English learners' speech that evaluated the impact of the intervention; (d) the target language was English; (e) written in English; and (f) contained the data and statistics necessary to compute a weighted effect size.

Studies were excluded for the following two reasons: (a) exclusively involved participants/speech raters with hearing or speech deficits; (b) had no treatments for speakers/listeners aiming to improve the intelligibility of the speech.

**Coding Procedures**

The researchers developed the coding scheme for the included studies through an iterative process. A coding sheet was created according to the research questions. This sheet included information about publication type (published and unpublished), intervention type (interventions focusing on NNESs/NESs), outcome measures (intelligibility and comprehensibility), speech task type (controlled and spontaneous) and length of the scales (short or long).

There were 16 out of 24 studies using Likert scales. Studies with scales in the length of 7 or below were coded as short and studies with scales in the length of 9 were coded as long. Also, some studies used scales of 1-1000 on a sliding bar for participants to move to indicate their perceptions of the intelligibility or comprehensibility of the speech samples (e.g., Reid, 2019; Saito & Akiyama, 2017). I coded these studies as long because this type of measurement tool offered finer scales than 5- or 7-point scales.

Several iterations of trial coding and revisions were conducted to achieve agreement in coding. If an inconsistency occurred, the coders resolved the inconsistency through discussion to clarify the initial coding criteria and arrive at a consensus.

**Effect Size Calculation**

I used procedures described by Lipsey and Wilson (2001) to calculate an unbiased effect size (Hedge's *g*) for intelligibility and comprehensibility based on reported means and

standard deviations. In the absence of reported means or standard deviations, I employed other procedures prescribed by Lipsey and Wilson (2001) using *p*-values and test statistics ( *t* or *F*) and/or confidence intervals to estimate an effect size.

**Data Analysis**

I analyzed the data with the following three steps using Stata 16 software. First, I conducted an overall effect size calculation. In some of the included studies, outcome measures about both intelligibility and comprehensibility were reported. Since one assumption underlying meta-analyses is that the effect sizes being integrated are statistically independent (Lipsey & Wilson, 2001), I used robust variance estimation (RVE) to integrate statistically dependent effect sizes in meta-analysis (Hedges, Tipton, & Johnson, 2010). At this step, I also tested for differences in overall effect sizes by running sensitivity analyses using the traditional overall effect size calculation. I calculated the effect size by averaging the scores of multiple measures when the holistic score was not available in the study.

Next, I conducted moderator analyses to determine which moderators help explain heterogeneity. Two criteria must be met: a) moderator variable was statistically significant ($p<.05$); and b) the addition of the moderator variable substantially reduced tau-square from the regression model without any moderator variables.

Third, for moderator variables that were statistically significant and substantively reduce tau-square, I conducted stratified RVE analyses to estimate subgroup effect sizes (with confidence intervals) and subgroup heterogeneity. Finally, I used the funnel plot asymmetry assessment (Egger, Smith, Schneider, & Minder, 1997) to check the potential publication bias.

## Results

Thirty-seven effect size estimates were extracted from 24 independent studies in 23 publications. Of these 24 studies, 16 examined interventions on NNESs' English pronunciation and 8 interventions on NESs' perceptions of NNESs' English speech. Some studies measured

both intelligibility and comprehensibility (e.g., Parlak, 2010). Some studies used multiple scoring methods, such as passage comprehension and transcription, to measure intelligibility (e.g., Derwing, Rossiter, & Munro, 2002); some used multiple speaking task types, such as picture narration, video narration, role-play and monologue, to measure the outcome (e.g., Galante & Thomson, 2017). Detailed information about the features of the included studies is summarized in Table 1.

**Table 1.** Summary of included studies

| **Publication** | **Publication Type** | **Intervention Type** | **Outcome Measures** | **Speech Task Type** | **Length of Scale** |
|---|---|---|---|---|---|
| Cooper, et al., 2020 | Pub | NESs | C | Uncontrolled | Short |
| Derwing, Rossiter, & Munro, 2002 | Pub | NESs | I | Controlled | N.A. |
| Galante & Thomson, 2017 | Pub | NNESs | C | Uncontrolled | Long |
| Graff, 2006 | Unpub | NNESs | I | Controlled | Short |
| Jiang, 2018 | Unpub | NNESs | C | Uncontrolled | Short |
| Kang, Rubin, & Lindemann, 2015 | Pub | NESs | C | Uncontrolled | Short |
| Levis & Levis, 2018 | Pub | NNESs | C | Uncontrolled | Long |

**Table 1** Continued

| **Publication** | **Publication Type** | **Intervention Type** | **Outcome Measures** | **Speech Task Type** | **Length of Scale** |
|---|---|---|---|---|---|
| Lindemann, Campbell, Litzenberg, & Subtirelu, 2016 | Pub | NESs | I | Controlled | N.A. |
| Liu, 2008 | Unpub | NNESs | C | Uncontrolled | Long |
| Mahdi, 2019 | Pub | NNESs | C | Controlled | Short |
| Parlak, 2010 | Pub | NNESs | I | Uncontrolled | Short |
| Reid, Trofimovich, & O'Brien, 2019 | Pub | NESs | C | Uncontrolled | N.A. |
| Rubin, 1992 | Pub | NESs | I | Controlled | N.A. |
| Saito, 2011 | Pub | NNESs | C | Uncontrolled &controlled | Long |
| Saito & Akiyama, 2017 | Pub | NNESs | C | Controlled | N.A. |
| Saito & Saito, 2017 | Pub | NNESs | C | Controlled | N.A. |
| Seferoğlu, 2005 | Pub | NNESs | C | Uncontrolled | Short |
| Sidaras, Alexander, & Nygaard, 2009 | Pub | NESs | I | Controlled | N.A. |

**Table 1** Continued

| Publication | Publication Type | Intervention Type | Outcome Measures | Speech Task Type | Length of Scale |
|---|---|---|---|---|---|
| Staples, Kang, & Wittner, 2014 | Pub | NESs | C | Uncontrolled | Short |
| Tanner & Landon, 2009 | Pub | NNESs | C | Uncontrolled | Short |
| Trofimovich, Lightbown, Halter, & Song, 2009 | Pub | NNESs | C | Uncontrolled | Long |
| Yenkimaleki & van Heuven, 2019 | Pub | NNESs | C | Uncontrolled | Short |
| Yeşilçınar, 2019 | Pub | NNESs | C | Uncontrolled | Short |

"N.A." indicates that Likert scales were not used in these studies. "Pub" indicates published. "Unpub" indicates unpublished. "C" indicated comprehensibility. "I" indicated intelligibility.

**Overall Effect Sizes**

Using Stata 16 robumeta command for robust variance estimation (RVE) meta-analysis, I found that all 37 effect size estimates reflected a statistically significant moderate effect size (ES=0.40, $p<0.05$, 95% CI[0.25, 0.54]) overall for improving intelligibility of NNESs' English speech. As for interventions focusing on training NNESs, the 15 studies yielded a slightly larger effect size (ES=0.47, $P<0.05$, 95% CI[0.26, 0.69]). Similarly, a statistically significant effect size (ES=0.28, $P<0.05$, 95% CI[0.05, 0.51]) was also observed in the ten intervention studies focusing on NESs' perceptions and attitudes, although the effect size was smaller than the overall effect size and that for NNES interventions.

At the same time, a certain amount of variability remained that was statistically significant (Q=55.88, p<0.05, I square=33.78%, tau square=0.054). This suggests that variance among the studies could be attributable to factors other than random errors.

**Moderator Analysis**

The moderator analyses (see Table 2) indicated that the measures of intelligibility and comprehensibility, and the speech task type, were significant factors explaining the effect size variations between the included studies. However, the length of scales did not significantly differentiate the effectiveness of interventions.

**Table 2.** Moderator analyses results

| **Moderator Models** | **Moderator Variable Statistics** | | | | **Model Statistics** | |
|---|---|---|---|---|---|---|
| **Baseline & Variables** | Beta | SE | df | 95% CI | $\tau^2$ | $\Delta\tau^2$ |
| **Baseline** | / | / | / | / | 0.047 | n/a |
| **Constructs of overall intelligibility** | 0.40* | 0.11 | 7.89 | 0.14, 0.65 | 0.01 | -0.037 |
| **Speech task type** | 0.35* | 0.12 | 11.51 | 0.087, 0.61 | 0.016 | -0.031 |
| **Length of scales** | -0.044 | 0.093 | 4.72 | -0.29, 0.20 | 0.00 | -0.047 |

*indicates p<0.05

Further analysis of the effect sizes of the sub-groups of each significant moderator showed that the interventions yielded statistically significant effect sizes for comprehensibility (N=17, ES=0.48, P<0.05, 95% CI[0.33, 0.62]), but when measured by intelligibility, the interventions did not have any statistically significant effects (N=6, ES=0.09, P>0.05, 95%CI[-0.15, 0.33]). I also found that spontaneous speech sample activities yielded statistically significant effect sizes (N=17, ES=0.48, P<0.05, 95% CI[0.32, 0.64]), but interventions with controlled speech sample activities did not yield statistically significant effect sizes (N=8, ES=0.16, P>0.05, 95% CI[-0.08, 0.40])

**Publication Bias**

From the funnel plot of these 37 studies, I found that these studies basically scattered symmetrically around the summary effect. There is no obvious void space in the plot. This may indicate that publication bias is not likely for the included studies in the current meta-analysis.

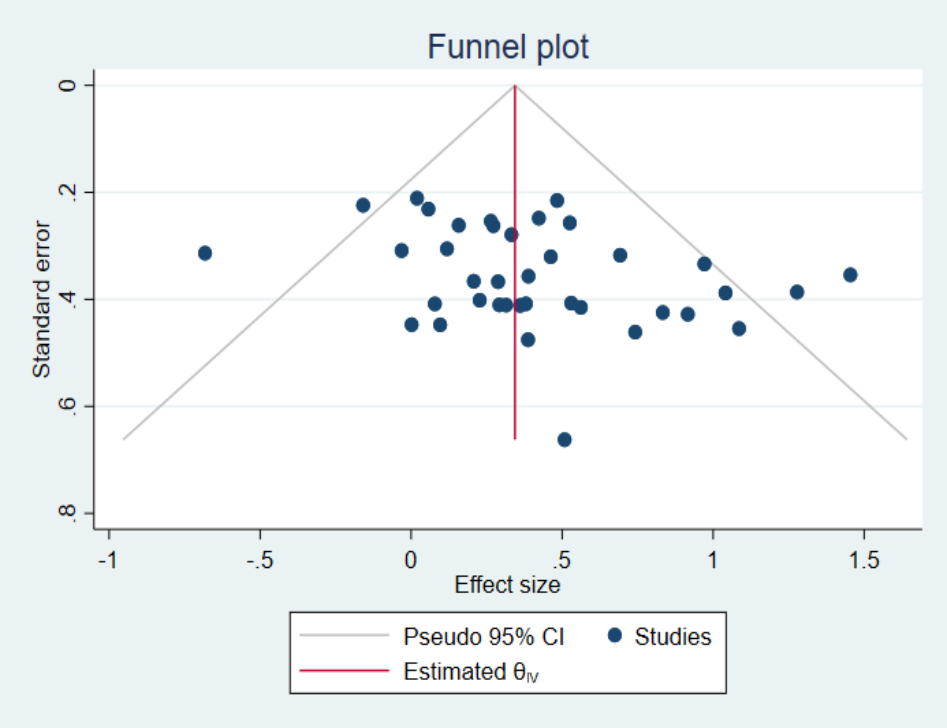


**Figure 2.** Funnel plot

## Discussion

### Overall Effectiveness of Interventions

This study identified 24 studies from 1992 to 2020 using criteria of overall intelligibility to measure the effectiveness of interventions for either NNESs as speakers or NESs as listeners to improve the overall intelligibility of NNESs' English speech. From these 24 studies, I identified 37 effect sizes related to the outcome measures of either intelligibility or comprehensibility. Overall, a statistically significant positive effect was found in these included studies. A meta-analysis run by the RVE method revealed a small to medium overall effect size (ES=0.40) for all the intervention studies included in this meta-analytic review. These significant results provided empirical support for the instructional practice of improving the overall intelligibility of NNESs' English speech. Also, similar results were found in separate meta-analyses for each type of intervention, for either NNESs or NESs. Fifteen interventions for NNESs yielded a larger effect size of 0.47, compared to the effect size of 0.28 from ten interventions for NESs. The significant effect of the intervention on NNESs is in line with the previous meta-analyses which found a medium effect for pronunciation instruction (e.g., Lee, Jang, & Plonsky, 2015; Saito & Plonsky, 2019). More notably, the significant effect of the interventions about NESs as listeners indicated that it was also useful to design interventions focusing on improving NESs' familiarity with NNESs' foreign-accented English or attitudes towards NNESs, in order to improve NESs' perceptions of NNESs' English.

Meanwhile, differences exist in the effect sizes for interventions on NNESs between the current and previous reviews, which together present a more comprehensive understanding of the impact of pronunciation instruction on NNESs' English speech. First, the effect size in this study is much smaller than those identified previously. This discrepancy may be attributed to differences in the measures used in the current and previous reviews. The current review evaluated the effect of pronunciation instruction on intelligibility and comprehensibility, while

previous reviews evaluated the effect on specific proficiency measures, such as segmental and suprasegmental features (e.g., Lee, Jang, & Plonsky, 2015; Mahdi & Khateeb, 2019; Saito & Plonsky, 2019). This effect size discrepancy resulting from the measurement differences suggests that existing pronunciation teaching practices have a greater effect on specific aspects of speech than on overall intelligibility and comprehensibility. Second, as opposed to the result in the current review, a non-significant result about the effectiveness of pronunciation instruction based on global measures such as intelligibility, comprehensibility, accentedness and fluency was found in Saito and Plonsky's study (2019). This difference was likely caused by the different subconstructs of outcome measures used for the effectiveness evaluation. Currently, there is limited evidence for significant correlations among all four measures used in Saito and Plonsky's study (2019). The effects tested by these four unrelated measures may have compromised the impact of pronunciation instruction. It may be more meaningful to evaluate the effectiveness of pronunciation instruction based on the overall intelligibility principle with intelligibility and comprehensibility measures which have proven to be complementary measures for speech understanding (Hustad, 2006).

The significant effect of interventions on NESs found in the current review is consistent with findings from previous research that reported significant relationships between NESs' perceptions of NNESs' English speech, and their attitudinal or emotional reactions (Munro & Derwing, 1995; Dragojevic & Giles, 2016). Munro and Derwing (1995b) proposed two types of comprehensibility: a subjective one and an objective one. The former refers to listeners' perceptions of their ability to understand speakers and the latter refers to their actual ability to understand speakers. NESs' prejudice towards NNESs' English speech often influences their beliefs about their own ability to understand the speech (Derwing & Munro, 1997; Lindemann, 2002; Munro & Derwing, 1995a; Rubin, 1992). If NESs assume that NNESs' foreign-accented English will interfere with their understanding, NESs may readily come to believe that they

cannot understand NESs. This negative association often results in NESs' reluctance to make efforts to understand NESs' speech ( Munro & Derwing, 1995a). These interventions focusing on increasing NESs' familiarity with NNESs' foreign-accented English and on improving NESs' attitudes towards NESs mitigated this negative association, which makes NESs either subjectively feel less difficulty in understanding NESs or more willingness to make efforts to understand NESs.

**Intelligibility and Comprehensibility**

The interventions on comprehensibility yielded statistically significant effect sizes (ES=0.48), but the seven interventions on intelligibility did not show any significant effects (ES=0.09). This result indicates that interventions, whether targeting NNESs or NESs, were effective in making NESs feel that understanding NNESs' speech was less effortful (i.e., more comprehensible), but they did not improve how much NESs can actually understand the speech (i.e., intelligibility). This mixed finding is not surprising because there has not been any evidence showing direct equivalence between the measures of intelligibility and comprehensibility (Kennedy & Trofimovich, 2019). In fact, each measure has its own features. When measuring comprehensibility, Likert scales or 1-1000 sliding bars were used for NESs to self-report how much effort they are taking to understand NNESs' speech. This scoring method shows subjectively how easily NESs feel they can understand NNESs. In contrast, scoring methods for intelligibility are more objective. Among the 11 effect sizes measured by intelligibility, six used word-to-word transcription, three used passage comprehension with multiple questions, and one used cloze tests. These scoring methods were direct and objective ones that tested what the listeners actually understood (Munro & Derwing, 1995).

The non-significant effects of intelligibility measured by objective scoring methods indicates that the current interventions are not sufficient to improve NESs' actual understanding of NNESs. This finding somewhat echoed what was found in Saito and Plonsky's meta-analysis

(2019). In their review, pronunciation instruction was found to facilitate specific aspects of NNESs' speaking such as segmental and suprasegmental accuracy, but its impact on global English pronunciation remained unclear. At the same time, non-significant effects of intelligibility were also found in the interventions focusing on NESs. For example, Lindemann and colleagues (2016) used explicit and implicit training methods for improving NESs' comprehension of NNESs speech. Compared to participants in the control group, the participants in neither of the experimental groups received higher scores in the comprehension of a longer passage. The researchers attributed this failure to the limited time of training and argued that improvement in NESs' actual understanding of NNESs' accented-English might occur from a longer training.

**Rating Scale Length**

This is the first analytical review to consider the rating scale length when examining the effectiveness of interventions on NNESs and NESs. The results showed that the length of scales did not moderate the effectiveness of interventions. This finding supported the observation reported in a correlational study about how rating scale length was related to the judgments of L2 pronunciation (Isaacs & Thomson, 2013). In this study, 40 NESs were randomly assigned to 5-point or 9-point rating scale conditions to evaluate the comprehensibility, accentedness, and fluency of 38 NNESs in Canada. The results showed that there were no differences in mean scores obtained by 5-point and 9-point scales.

I noticed some patterns of scale length choice after a close examination of the distribution of the scale length among different publications (see Table 3). First, more and more recent studies used finer scales, such as 9-point scale and 1-1,000 sliders (e.g., Galante & Thomson, 2017; Levis & Levis, 2018; Reid, Trofimovich, & O'Brien, 2019). This may be due to an assumption that finer scales may help researchers to detect change in listeners' perceptions of the speech samples (Bachman, 1990; Isaacs & Thomson, 2013). Second, studies with relevant

research questions chose the same scale length. For example, a 7-point scale was used in studies conducted by Cooper, et al.(2020), Kang (2008), Kang, Rubin, and Lindemann (2015), and Staples, Kang and Wittner (2014). All of these studies examined how interventions focusing on NESs could improve their perceptions of NNESs' speech. This consistent use of a measurement tool made results comparable across studies (Isaacs & Thomson, 2013).

**Table 3.** Scale length in included studies

| | **<=5-point scale** | **7-point scale** | **9-point scale** | **1-1000 slider** |
|---|---|---|---|---|
| Cooper, et al., 2020 | | x | | |
| Galante & Thomson, 2017 | | | x | |
| Jiang, 2018 | | x | | |
| Kang, Rubin, & Lindemann, 2015 | | x | | |
| Levis & Levis, 2018 | | | x | |
| Liu, 2008 | | | x | |
| Mahdi, 2019 | x* | | | |
| Parlak, 2010 | | | x | |
| Reid, Trofimovich, & O'Brien, 2019 | | | | X |

**Table 3** Continued

| | **<=5-point scale** | **7-point scale** | **9-point scale** | **1-1000 slider** |
|---|---|---|---|---|
| Saito, 2011 | | | x | |
| Saito & Akiyama, 2017 | | | | X |
| Saito & Saito, 2017 | | | | X |
| Seferoğlu, 2005 | x* | | | |
| Staples, Kang, & Wittner, 2014 | | x | | |
| Tanner & Landon, 2009 | x** | | | |
| Trofimovich, Lightbown, Halter, & Song, 2009 | | | x | |
| Yenkimaleki & van Heuven, 2019 | x* | | | |
| Yeşilçınar, 2019 | x* | | | |

* 5-point scale used; **4-point scale used

**Speech Task Type**

The moderating analysis of speech task type (controlled vs. spontaneous) indicated that different types of speech tasks used in the assessment was an important factor explaining the

variations in study results. For this moderator, spontaneous speech activities yielded statistically significant effect sizes (ES=0.48), while the effect size of controlled speech sample activities was smaller (ES=0.16) and non-significant ($p<0.05$). This result is in line with Saito and Plonsky's (2019) findings in their meta-analysis which showed that for the global outcome measure, spontaneous speech tasks yielded smaller effect sizes than controlled speech tasks; in addition, none of these effects sizes were reliable, as 0 was included in the confidence intervals of these two effect sizes.

Spontaneous speech tasks were found to yield statistically significant and larger effect sizes than controlled speech, which contradicts the relevant findings in previous meta-analyses of L2 learning. For example, in Lee, Jang, and Plonsky's meta-analysis about pronunciation instruction (2015), more controlled speech tasks had larger effects than spontaneous speech tasks when measured by specific segmental and suprasegmental measures. Similar findings were reported in a meta-analysis about the impact of computer-assisted pronunciation training (Mahdi & AlKhateeb, 2019).

The mixed findings of the effects of controlled and spontaneous speech tasks used in the assessment can be explained by the nature of the outcome measures examined in these meta-analyses. When speech tasks were rated by global measures, such as intelligibility and comprehensibility, raters focused more on meaning-related aspects of the speech tasks, which helps more in eliciting their perceptions about how much they can understand the speech (i.e., intelligibility) and the effort they needed to understand the speech (i.e., comprehensibility). Therefore, compared to controlled tasks, such as sentence reading (e.g., Sidaras, Alexander, & Nygaard, 2009), spontaneous tasks, such as oral presentation (e.g., Cooper, et al., 2020; Kang, Rubin, & Lindemann, 2015) and picture description (e.g., Levis & Levis, 2018; Liu, 2008), provided more contextual basis to allow listeners (i.e., raters) to understand the meaning of the speech samples. In contrast, for the specific measures, such as the rating of the accurate

application of specific sound knowledge, participants will definitely perform better in the application of knowledge of these specific sounds in the form-focused controlled activities, such as word reading or sentence reading, than in spontaneous activities, such as picture description, in which participants need to focus on both content and form.

## Implications and Recommendations

Successful communication is jointly constructed by both speakers and listeners (e.g., Lindemann & Subtirelu, 2013; Lindsey, King, Hebl, & Levine, 2015; Rubin, 1992). On the one hand, some segmental and suprasegmental elements of pronunciation which were found to be essential to the intelligibility and comprehensibility of the communication should be included in the pronunciation instruction for NNESs. These elements include segmental, prosodic, temporal, lexical, and grammatical features of English in communication (Isaacs & Trofimovich, 2012; Saito, Trofimovich, & Isaacs, 2017). On the other hand, interventions about improving NESs' perceptions of foreign-accented English should also be included. These interventions include those aiming to train NESs to be more familiar with foreign-accented English, and those aiming to improve NESs' attitudes towards NNESs and NNESs' foreign-accented English.

Regarding the selection of the scale length for speech sample rating in future studies, the following recommendations are proposed. First, researchers should consider the scale conventions used in the field to make the results comparable to previous studies. Second, researchers should consider raters' prior rating experience and provide necessary training for raters to understand what each scale point means (Isaacs & Thomson, 2013; Li, Taguchi, & Xiao, 2019).

Several methodological issues about intelligibility measures need to be further examined. First, when scoring intelligibility, some raters were asked to answer some comprehension questions in the form of multiple-choice questions or cloze tests, instead of transcribing the sentences or passages from the speech samples. Since the current review had very strict inclusion

criteria related to the global measure of intelligibility, I had a relatively limited number of included studies that used comprehension measures to score intelligibility (N=3). This made it impossible for us to run a moderator analysis with this factor because the result would be unreliable when the number of studies is less than four (Hedges, Tipton, & Johnson, 2010). Additional intervention studies may use different types of questions related to passage comprehension to further examine the impact of the intervention.

Second, in the current meta-analysis, larger effects were identified in spontaneous speech tasks than controlled tasks, which suggests that for meaning-based ratings of intelligibility and comprehensibility, spontaneous tasks instead of controlled tasks are preferred. Although many researchers have recommended that both types of speech tasks should be adequately used in the assessment of L2 research to provide a more comprehensive picture about the effectiveness of the intervention (Lee, Jang, & Plonsky, 2015; Mahdi & AlKhateeb, 2019) when the intelligibility principle is involved in the assessment, the authentic communication environment provided by spontaneous speech tasks should serve as an optimal instrument to test the impact of an intervention.

# 3. EFFECTS OF INTERVENTION ACTIVITIES ON IMPROVING UNDERGRADUATES' ATTITUDES TOWARDS ITAS AND PERCEPTIONS OF INTERNATIONAL TEACHING ASSISTANTS' SPEECH *[1]

## Introduction

International graduate students currently account for up to 44% of students in U.S. graduate programs (Okahana & Zhou, 2019), and U.S. universities routinely rely on them to serve as instructors in different disciplines. In fact, most universities require international teaching assistants (ITAs) to demonstrate a sufficient level of English proficiency before they are assigned to teaching appointments. In order to fulfill this requirement, many ITA training programs have created training courses or workshops to improve ITAs' pronunciation, which was viewed as the biggest barrier for them to communicate effectively with UGs (Jenkins, 2002, p. 83). However, many undergraduates (UGs) perceive ITAs as unqualified to teach, in part due to their perceptions of ITAs' non-mainstream English and the comprehensibility of ITAs' English. (See "Language Ideology" section for more discussion of this issue.)

Rubin (1992) proposed that UGs' negative attitudes towards ITAs create a mental barrier that inhibits their ability to understand ITAs. This deteriorated understanding naturally impairs their academic performance. For example, in Villarreal's study (2014), several participants reported dropping at least one class with an ITA, often claiming that ITAs' accents had played a role in these decisions. This study also found that ITAs' accents were implicated in altering students' academic plans on a larger scale.

[1] Part of the data reported in this section is reprinted with permission from "The impact of linguistic diversity education on L1 English speakers' ideologies, attitudes, and perceptions of international teaching assistants" by Cooper, B., Payne, G., Hu, X., Dixon, Q., & Kuo, L., 2020, *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference*, Copyright [2020] by *the* 11th Pronunciation in Second Language Learning and Teaching Conference.

Previous research has shown that racial factors rather than linguistic factors often determine the judgements of people's speech. For example, Kang and Rubin (2009) observed the negative effect of perceived ethnicity on listeners' comprehension of speech. In their study, participants who saw a picture of an Asian person while they were listening to an English speech sample scored lower on a test of listening comprehension than those who saw a picture of a Euro-American speaker while listening to the same speech sample. Similarly, in a case study with white and non-white Canadian English speakers, Faez (2012) revealed that white English speakers were more comfortable than non-white English speakers to claim themselves as native-English speakers. This association between racial factors and judgements of people's English language proficiency was also confirmed by Fraser and Kelly's study (2012) where they found small but significant correlations between attitudes toward non-white ethnicities and both intelligibility and comprehensibility of spoken English.

With the growing evidence of native English speakers' (NESs) responsibility for the understanding of non-Native English speakers' (NNESs) speech, more and more researchers have argued that successful communication is jointly constructed by both speaker and listener (e.g., Lindemann & Subtirelu, 2013; Lindsey, King, Hebl, & Levine, 2015; Rubin, 1992). Disregarding the role of the NESs as listeners and placing the whole burden of communication on NNESs could also be considered unethical (Derwing, Fraser, Kang, & Thompson, 2014). Therefore, it is of great importance for us to shift our research focus from speakers to listeners to experiment with effective ways to aid intelligible communication.

**The Role of UGs in ITA-UG Communication**

While there were only a handful of studies examining ITA and UG speech（Looney，2015), there are even fewer studies focusing on how UGs acted in ITA-UG communication. Most of these studies were conducted by Chiang and his colleague (Chiang, 2009; Chiang & Mi, 2008). In their analysis of dyadic interactions between Chinese ITAs and UGs in natural office

hour settings, they observed the use of correction, prevention and reformulation by UGs as strategies to overcome linguistic and cultural differences in communication. Chiang (2009) also revealed that it was UGs' positive attitudes towards ITAs that motivated UGs to take initiative to use these collaborative strategies to achieve mutual understanding. UGs' attitudes towards ITAs play a principal role in achieving successful communication across linguistic and cultural boundaries.

In a similar vein, efforts have been made to examine the relationship between UGs' attitudes toward NNESs and the way they communicate with NNESs. For example, in Linderman's analysis of conversations between UGs and Korean English speakers when they were paired to complete an interactional map task, two non-collaborative communicative strategies used by UGs with negative attitudes towards Koreans were identified: problematizing and avoidance (2002). In the communication, UGs either "consistently frame their NNES partners' contributions as problematic" (p. 431), or "took very little responsibility for the success of the interaction" and "failed to provide crucial feedback to the NNESs" (p. 426). Both of the non-collaborative strategies led to the breakdown of the communication and even the unsuccessful completion of the task.

Subtirelu (2017) examined UGs' discourse in a focus group interview about their international instructors in a university in the southern U.S. His findings revealed that many UGs avoided communicating with ITAs in the classroom rather than attempting to repair the misunderstanding, partly due to their prejudice about ITAs' English communicative ability. For many UGs, their past miscommunication with other ITAs led to their negative assumptions about ITAs' communicative competence, resulting in their avoidance of any communication with different ITAs in other courses. Meanwhile, Subtirelu (2014) identified the collaboration orientation to communication with ITAs from other UGs in his research. These UGs asserted the normality and ubiquity of diversity, linguistic or otherwise, and believed that contact with ITAs

granted them the opportunity to become accustomed to linguistic diversity. Subtirelu (2014) provided evidence to show that UGs' attitudes towards linguistic diversity and their acceptance of foreign-accented English from speakers with different cultural backgrounds may affect their communicative behaviors with ITAs.

The findings of the above studies indicated that problems of miscommunication between UGs and NNESs, to some extent, were caused by UGs' negative attitudes towards the NNESs. These negative attitudes may be rooted in the ideology that English belongs to native speakers, or, as discussed above, previous negative experiences with NNESs being generalized to all NNESs. The purpose of this dissertation study aims to examine the effect of changing UGs' language ideology on enhancing intelligible communication between UGs and ITAs.

## Literature Review

### Improving UGs' Attitudes Towards ITAs with Structured Contact Group Activities

Having recognized the contribution of NESs as listeners to the effective communication with NNESs, many researchers made efforts to improve American UGs' attitudes towards ITAs' foreign-accented English so as to improve their understanding of ITAs. For example, promoting intergroup contact has been identified as one of the effective structured activities to mitigate UGs' negative attitudes towards UGs (Kang, Rubin, Lindemann, 2015; Staples, Kang, & Wittner, 2014). In Staples et al.'s study (2014), through an 8-week contact activity (conversation partners between UGs and ITAs), UGs' perceptions of comprehensibility and accentedness of ITAs' speech as well as their evaluation of ITAs' teaching ability all improved significantly. However, the quality of conversation partner activities varies depending on many factors such as the conversation topic and the personalities of partners. In order to maximize the potential of conversation partnership activities, Kang et al. (2015) designed a one-hour problem-solving game for interactions between UGs and ITAs who had never met in person before the game, utilizing the principles of contact theory (discussed below) to maximize the game's

potential effect. In the game, the participants were of equal status because they were not in a teacher-student hierarchy relationship. They shared common goals because they worked together to solve the puzzles in the game. In order to achieve this goal, all group members cooperated with each other with the support of researchers who helped to establish a relaxing and interactive environment. After this contact activity, greater improvement for UGs in their evaluations of ITAs' instructional competence ratings and comprehensibility ratings was observed, but there was no significant improvement in UGs' attitudes towards ITAs' accents.

These two studies provided evidence for the positive impact of contact activities on enhancing UGs' understanding of ITAs' speech so as to improve their evaluation of ITAs' teaching. However, the inconsistent findings of UGs' perceptions of ITAs' accents still need to be further explored. One possible explanation is the duration of the contact activity. In Staple et al.'s study (2014), UGs were more familiar with ITAs' speech by conversing with ITAs for 50 minutes once a week for 8 weeks. The total contact time between UGs and ITAs in this program is about 7 hours. In contrast, in Kang et al.'s study (2015), UGs only interacted with ITAs for one hour. The different familiarity levels of UGs with ITAs' accents may have led to different results of UGs' scorings of ITAs' speech accentedness in these two programs.

Many UG training programs aiming to improve UGs' attitudes towards ITAs' teaching ability and their speech cannot afford activities lasting more than one or two hours. Kang et al.'s activity (2015) is more practical and easier to replicate. The finding that the contact intervention did not exert any impact on accent standardness ratings revealed the limitations of contact activities in improving UGs' sensitivity to, and tolerance for, ITAs' foreign-accented speech. Contact activities have not explicitly addressed the root assumptions of language ideology such as the idea that L2 English is a 'deficient' form of L1 English (Subtirelu & Lindemann, 2016). It is imperative to learn whether UGs actually regard ITAs' accents negatively in the sense of regarding it as a problem, or if they were simply holding a negative attitude towards non-native

accents without such judgement (Kang et al., 2015). Therefore, explicitly challenging L1 English speakers' ideologies about L2 English would be of great value to further examine this issue.

**Educational Practice of Challenging Undergraduates' Language Ideology**

The fact that there are far more NNESs than NESs (Crystal, 2003) has changed the settings of with whom NESs will speak English. Under this global context of English use in the world, NESs face challenges about their language ideology. Their sense of superiority about their own English accent over other English varieties might be the root of their negative attitudes towards foreign-accented English. This awareness inspired some attempts to include multilingual and multicultural education in UGs' courses (e.g., Eslami, Cassell, & Ates, 2015; Eslami, Moody, & Pashmforoosh, 2019; Villarreal, 2013). Eslami and her colleagues conducted a series of studies about how the education of World Englishes could change UGs' attitudes towards linguistic diversity (Eslami, Cassell, & Ates, 2015; Eslami, Moody & Pashmforoosh, 2019). They embedded into U.S. pre-service teachers' UG courses multiple in-class and online educational activities such as listening to presentations by David Crystal about different varieties of World Englishes and the biggest challenges for English language teachers, watching videos about different American accents and writing reflections on these English varieties, and online discussions. The findings showed positive results in promoting pre-service teachers' awareness of language diversity. These pre-service teachers showed appreciation of the language diversity English language learners bring into the classroom. Their survey responses indicated that they would make their future ELL students feel welcome and not feel discrimination because of their language.

However, their studies did not measure pre-service teachers' attitude change before and after the instruction to provide more solid evidence of attitude change. Also, their studies did not examine whether this self-reported attitude change about their language ideology actually changed their behavior towards the speakers of English varieties. Further studies are needed to

examine the effect of enhanced understanding of language diversity on UGs' communication with NNESs in terms of the comprehensibility of ITAs' language.

In terms of the format, most of the educational programs about linguistic diversity reported in previous studies either included multiple activities embedded in the course curriculum, which lasted 10 weeks (e.g., Eslami et al., 2015, 2019) or were delivered in the format of video or audio (Eslami et al., 2019; Villarreal, 2013). This time-demanding educational practice made it challenging to be replicated in other educational programs where time is in short supply. It is also possible that the impact of the educational content on promoting the awareness of language diversity will be compromised by the lack of interaction in audio listening or video watching especially for a program relatively limited in time. Therefore, since long interventions may not always be practical, it is necessary to create a compact workshop aiming to enhance understanding of linguistic diversity, which is easily tailored to any educational program, such as new UGs' orientation or short-term pre-service and in-service teacher education. As Eslami et al. (2015) advocated, it is of great importance to examine different instructional activities and program models in different contexts to promote the awareness and acceptance of language varieties.

## Theoretical Framework

### Language Ideology

Language ideology refers broadly to the ways in which people and groups conceptualize language. It is a network of beliefs that language users hold, either tacitly or overtly, about language and its assumed relation to other aspects in their environments, especially other individuals and social groups, which stem either from explicit teaching or implicit socialization (Subtirelu & Lindemann, 2016).

In the current study, standard language ideology is defined as the belief that there is a single, "correct" form of English spoken by educated individuals (Lippi-Green, 1997; Milroy,

2001). Thus, the term *standard language ideology* is interchangeable with *standard English language ideology* in the current study. Standard or mainstream English, which is spoken by primarily white, upper-middle class Americans, is often claimed to be "unaccented" (Kubota & Lin, 2006), though of course it is accented like all other English varieties. People with this standard English language ideology would hold a "standardness" principle when judging a person's speech. However, in the context where English is used as a lingua franca, the intelligibility principle, which eschews a deficit-oriented basis, rather than the standardness principle is more appropriate and practical to apply for enhancement of effective communication (Jenkins, 2006; Seidlhofer, Breiteneder, & Pitzl, 2006) .

NNESs usually suffer both prejudice and negative stereotypes affected by the standard language ideology conceived by NESs. For example, NESs of a "standard" variety may actually judge NNESs less intelligent (Lindemann, 2003), and less competent (Boyd, 2003). This social norm which associates non-mainstream accents in English with different races and ethnicities may cause unequal status between NESs and NNESs. NESs may take it for granted that they may release their portion of the burden of comprehension in communication or even refuse to communicate when confronting speakers with English they consider accented. This may result in issues such as not obtaining information, eliciting negative responses from others, and exclusion in communication (Gluszek & Dovidio, 2010)

**Contact Theory**

Contact theory was developed by Gordon W. Allport (1954), who stated that, under appropriate conditions, interpersonal contact was one of the most effective ways to reduce prejudice between majority and minority group members. Allport (1954) claimed that when people have the opportunity to communicate with others, they would be more able to understand and appreciate different points of views involving their way of life. This new appreciation and understanding could help to diminish the issues of stereotyping, prejudice, and discrimination.

Allport (1954) also pointed out that properly managed contact between the groups should be guaranteed so as to reduce these problems and lead to better interactions. Therefore, optimal contact activities should meet the following criteria:

- Equal status. Both groups must engage equally in the relationship. Differences in academic backgrounds, wealth, skill, or experiences should be minimized if these qualities will influence perceptions of prestige and rank in the group;
- Common goals. Both groups must work on a problem/task by pooling their efforts and resources and share this as a common goal;
- Intergroup cooperation. No competition exists between group members;
- Support of authorities. Both groups must acknowledge some authority that supports the contact and interactions between the groups and encourage friendly, supportive, egalitarian attitudes;
- Personal interaction. The contact situation needs to involve informal, personal interaction with outgroup members for them to mingle with one another.

Guided by the above criteria, a number of studies have designed contact activities aiming to improve NESs' attitudes towards NNESs and their foreign-accented English ( e.g., Kang, Rubin, Lindemann, 2015; Staples, Kang, & Wittner, 2014). The contact hypothesis has received extensive empirical validation in a positive way. A meta-analytic review of the effects of intergroups contact on intergroup attitudes has identified significantly improved attitudes based on 713 independent samples from 515 studies (Pettigrew & Tropp, 2006).

**Intelligibility**

In this dissertation, intelligibility is used in its broadest sense, which means "intelligible production and felicitous interpretation of English" (Nelson, 1995, p. 274). This definition actually entails three sub-constructs of intelligibility: intelligibility, comprehensibility and interpretability.

Intelligibility, in a narrow sense, refers to the ability of the listener to recognize individual words or utterances. It is usually measured by asking listeners to transcribe what they hear to check the accuracy of their transcription (Munro & Derwing, 1995).

Comprehensibility refers to the listener's efforts to understand the meaning of the word or utterance in its given context. It is usually measured by a five or seven-point Likert scale to check how easy or difficult the speech samples are to understand (e.g., Trofimovich, Lightbown, Halter & Song, 2009; Galante & Thomson, 2017).

Interpretability refers to the ability of the listener to understand the speaker's intentions behind the word or utterance. This level of speech understanding is the most difficult one for listeners because listeners may be able to recognize the words in the speech and understand its meaning, but have difficulty understanding the speaker's intentions behind it (Gallego, 1990). That is to say, interpretability is more related to the ability of understanding meaning in context. The implicit nature of this construct makes it naturally difficult to measure (Pickering, 2006).

**Current Study**

The current study compares how UGs' language ideology and attitudes towards ITAs will be affected by both types of interventions: contact experience with ITAs and language ideology education. At the same time, this study extends previous studies on language ideology education by quantitatively measuring the change of UGs' perceptions about ITAs' speech and teaching before and after the intervention. In order to achieve these objectives, three types of intervention activities were developed: contact activity, language ideology workshop activity, and contact activity plus language ideology workshop activity.

The following two research questions will be addressed in the study:

1. How will UGs' linguistic ideology and attitudes towards ITAs be affected by interventions targeting language ideology education, contact with ITAs, or both, respectively?

2. What are the respective impacts of three types of intervention on UGs' perceptions of ITAs' comprehensibility, accentedness, teaching ability, and English language proficiency?

## Methods

### Recruiting Process

Participants were recruited through the internal email systems among UGs from a university in the southern part of the U.S. In the recruiting email, UGs were told that they would participate in a couple of activities held by the researchers' department for a research study and they would be compensated for their time invested in the study. In this email, a link for a screening survey was attached. If UGs were interested in the study, they could fill out the survey, which was used to check their eligibility for the study. In order to be eligible to participate in this study, UGs were required to meet at least four requirements: a) they were UGs whose first language is American English; b) they were at least 18 years old on the day of the first activity for the study; c) they did not have any identified hearing problems; d) they could foresee that they would be available for the activities of the study.

Three hundred and eighty-seven students filled out the screening survey and showed their interest in this study. At the end of the screening process, two hundred and forty-eight qualified UGs were contacted and visited the language lab on campus to participate in the first study activity. However, due to unexpected conflicts of UGs' academic schedules with the study activities, 126 UGs failed to make all the required appointments for the study, so only 122 UGs completed the entire study.

### Participants

These 122 participants were from about 40 different academic departments across the university campus. Among these participants, 81 were female and 41 were male. There were 56

freshmen, 22 sophomores, 26 juniors and 36 seniors. As for their language backgrounds, 100 participants were native monolingual English speakers and 22 students were multilingual English speakers. These students were randomly assigned to four different groups: contact only group, workshop only group, combined group (contact activity plus workshop), and control group. Please see Table 4 for the demographic information across the four groups.

At the same time, twenty ITAs were recruited to take part in the contact activity with UGs. These ITAs and UGs had never met before the contact activity.

**Table 4.** Participants' demographic information

| **Group** | | **Contact only group** | **Workshop only group** | **Combined group** | **Control group** |
|---|---|---|---|---|---|
| **Gender** | Male | 6 | 10 | 13 | 11 |
| | Female | 20 | 20 | 16 | 26 |
| **Academic Year** | First-year | 4 | 12 | 10 | 8 |
| | Sophomore | 9 | 4 | 9 | 8 |
| | Junior | 8 | 7 | 4 | 7 |
| | Senior | 5 | 7 | 6 | 14 |
| **Language background** | Monolingual English speaker | 22 | 28 | 23 | 27 |
| | Multilingual speaker | 4 | 2 | 6 | 10 |
| **Total** | / | 26 | 30 | 29 | 37 |

## Materials

*Speech Samples*

There were 8 speech samples used in both the pre- and post- tests. Each speech sample was a 2-minute long excerpt from an 8-minute long videotaped micro-teaching lesson about a specific topic such as sociology, political science, or nutrition. To minimize the ceiling effect, I had undergraduate raters listen to a speech sample from a NES who spoke standard American English about how to develop the leadership. I placed this sample as the first one before all the other samples. If the raters were first scrutinizing this sample, the other samples would be scored lower. The rest of the seven micro-teaching videos were presented by ITAs who agreed to allow the researcher to use their videos in the research. These excerpts were all converted into audio recordings. These 2-minute excerpts were carefully chosen from the original oral presentations by considering the following factors: a) the content was neither too abstruse nor including vocabulary words beyond UGs' high school knowledge level about the specific subject; b) selected excerpts were continuous streams of speech; c) there was no interruption from the audience's voice or background noise. A pilot study had been conducted with two UGs outside of the study. They listened to these speech samples and had confirmed that these samples were understandable in content and of high sound quality.

At the same time, these seven samples represented four typical foreign-accented English spoken by the three largest population of ITAs in the southern areas of the U.S.: Chinese-accented English, Indian-accented English, Farsi-accented English and Spanish-accented English. The English language proficiency, as measured by their speaking scores on the TOEFL, varied among speakers (see Table 5).

**Table 5.** Speech samples

| Sample No. | Accent | Proficiency (TOEFL Speaking Score) | Topic |
|---|---|---|---|
| 0 | American English | / | Leadership skills |
| 1 | Spanish | 20 | Concrete |
| 2 | Spanish | 23 | Sociology |
| 3 | Indian | 23 | Monty Hall problem |
| 4 | Persian | 20 | Engineering |
| 5 | Chinese | 18 | Cognitive psychology |
| 6 | Chinese | 19 | Chemistry |
| 7 | Chinese | 18 | Nutrition |

*Background Survey*

The background survey consists of three sections in addition to questions about UGs' gender, major and academic year. The first section is about UGs' multicultural and multilingual experience (MCMLE). In this part, UGs were asked four major questions with a five-point rating scale for each question to indicate their previous multicultural and multilingual experience. For example, UGs were asked "How would you rate your proficiency in a language other than English?" with five options from not proficient in any foreign languages to proficient in one/two/three/more than three foreign languages. If they were proficient in at least one foreign language, they were asked to list these languages. This answer was used to further analyze if

their learning experience of a specific language affected their attitudes towards native speakers of this language. The second and third sections are questions about UGs' attitudes towards ITAs and about UGs' language ideology.

**Research Design**

The UG participants were informed about the purposes and activities of the study, signed consent forms, completed surveys and rated speech samples in the first and last activities for the study. They were randomly assigned to one of the four groups: control group (CG), contact only group (CO), workshop only group (WO) and the combined group (C&W). The contact only group took part in a 1-hour contact activity. The workshop only group participated in a 1-hour workshop. Workshop and contact activity combined group took part in the contact activity as well as the workshop, which lasted about two hours in total. The control group rated speech samples at the same time as the other three groups without taking part in any of the intervention activities.

**Intervention**

*Contact Activity*

This study used the same contact activity-mystery puzzle conducted in Kang et al.'s study (2015). In this activity, five to six UGs played a mystery puzzle game with two ITAs who had never met these UGs. The entire activity took about 1 hour. In this mystery puzzle activity, twenty clues were evenly and randomly distributed among members of groups. Then UGs and ITAs orally shared their clues with each other and figured out the solution to the puzzle by synthesizing these clues collaboratively without showing the clues.

The design of this activity aimed to create a cooperative communication atmosphere where UGs and ITAs shared equal status and worked together towards a common goal. Prior to this activity, they did not know each other. They were not instructors and students but collaborators. Since each clue was important to the solution of the puzzle, each participant

contributed to the solution of the puzzle by describing their clue. This type of interaction facilitated personal interaction among group members. At the same time, this was an activity supported by the research program to ensure that this event could proceed smoothly. The whole process was under researchers' guidance and monitoring. Therefore, this was an activity consistent with the core principles of contact theory.

*Workshop on Language Ideology*

In this study, an interactive face-to-face workshop about linguistic diversity was developed for UGs. This workshop lasted about 1 hour and was delivered by two consultants from the English Language Proficiency program (ELP) at the Center for Teaching Excellence (CTE) at Texas A&M University. This workshop created interactive activities for participants to: a) identify the source of judgements about different varieties of English, b) discuss the merits and dangers of language standardization, c) interrogate the linguistic merits of "standard" or mainstream English, d) trace the rise of English as a lingua franca, e) weigh the feasibility and desirability of accent modification, and f) identify the factors that affect listening comprehension. Please see Appendix B for a list of activities in the workshop.

**Measurement**

*Comprehensibility*

Comprehensibility ratings were collected while UGs listened to the speech samples. All groups listened to and rated all 7 speech samples in a random order with the samples from native English speakers always placed as the first one.

Following the conventions in the previous studies (e.g., Kang et al., 2010; Kang, Rubin, & Lindemann, 2015), five semantic differential measures items were used to test comprehensibility. These items included statements such as this speaker (1) was easy to understand, (2) was comprehensible, (3) was clear, (4) required little effort to understand, and (5) made it simple to grasp the meaning. A 7-point Likert scale, as in Kang et al. (2015), was used

to measure each sub-measure of the comprehensibility. Using seven points instead of the traditional five points allowed for finer discrimination of differences in the rated items. These five sub-measures have been tested by Cronbach's Alpha and reached a strong reliability of 0.90.

*English Language Proficiency*

UGs' perceptions of the ITAs' English proficiency was measured by a 7-point Likert scale with 1 representing low proficiency, 4 representing moderate proficiency and 7 representing high proficiency (Kang, et al., 2015).

*Accentedness*

Accentedness was defined as how much a speaker's accent differs from the subjective norm of listeners (Derwing & Munro, 2005). Following the convention in previous studies (e.g., Kang et al., 2010; Kang, Rubin, & Lindemann, 2015), four semantic differential measures items were used to test accentedness. These items included statements such as, the speaker (1) speaks with a foreign accent; (2) is a non-native speaker of English; (3) has a strong accent; (4) speaks like a non-native speaker of English. A 7-point Likert scale was used to measure each sub-measure of comprehensibility. These sub-measures reached a reliability level of 0.83 (Cronbach's alpha).

*Evaluation of ITAs' Teaching ability*

ITAs' teaching ability was measured by two semantic differential scales. These two sub-measures were adapted from an earlier scale of ITAs' instructional quality (Kang et al., 2015 & Staple et al., 2014). The earlier version of this measure included nine sub-items such as teaching effectiveness, teaching qualifications, classroom management, availability, and knowledge materials. Because the speech samples in this study were only 2-minute excerpts selected from simulation teaching presentations rather than authentic classroom lectures, it is not valid for UGs to rate all of the nine sub-items. I therefore used 7-point Likert scales to indicate how much they

agreed or disagreed to the following two statements: a) this speaker is an effective teacher; b) this speaker is a qualified teacher (1=strongly agree; 7=strongly disagree).

*Attitudes Towards ITAs*

Two sub-statements with a 7-point Likert scale were designed to measure UGs' attitudes towards ITAs ( 1=strongly agree; 7=strongly disagree). UGs needed to choose from the 7-point Likert scale ratings to indicate how much they agreed or disagreed with the statements: (a) If there are multiple sections of a course, I will take the section taught by a native English-speaking teaching assistant instead of a non-native English-speaking teaching assistant; (b) Native English-speaking teaching assistants are usually more effective in teaching than non-English speaking teaching assistants due to the latter's foreign-accented English. These two sub-measures have the marginally acceptable reliability of 0.70.

*Language Ideology*

Two statements were designed to measure UGs' language ideology: a) I speak English with no accent because I am a native English speaker; b) I am a native English speaker, so I do not need to modify my speech into standard American English. These two statements were designed to capture how much the UGs believed themselves to be native speakers of "standard" American English, and thus had no responsibility to modify their own speech to be more comprehensible. UGs needed to choose from the 7-point Likert scale ratings (1=strongly agree; 7=strongly disagree) to indicate how much they agreed or disagreed with the statements. These two sub-measures have the reliability of 0.76 (Cronbach's alpha).

**Data Collection**

Data were collected during the first and last visits of UGs in each group. During the first visit, all four groups of UGs completed an online background survey with the self-evaluation of their attitudes towards ITAs and their linguistic ideology. Then each of them listened to the speech samples. While they were listening to the speech samples, UGs were asked to answer

questions about how they evaluated speakers' comprehensibility, accentedness, English language proficiency, and teaching ability. These questions were printed out and UGs provided their written responses.

After the first visit, all three experimental groups went through different activities according to the research design in the following four weeks. In one to three days after the activities, UGs in all four groups made their last visit to complete the same online survey questions about their attitudes towards ITAs and their language ideology. They rated the same speech samples again and answered the same questions about speakers' comprehensibility, accentedness, English language proficiency, and teaching ability. The speech samples were randomly sequenced in pre- and post-ratings for a counterbalanced design. Please see Table 6 for the data collection process.

The whole data collection process was conducted in a language lab where UGs used a desktop computer to listen to the samples with a headphone. Before rating the samples, the survey was completed online through Qualtrics when UGs were physically in the lab.

**Table 6.** Data collection procedure

| **Time** | **Activity** | **Workshop Only** | **Contact Only** | **Workshop & Contact** | **Control Group** |
|---|---|---|---|---|---|
| Week 1 | Survey+rating | √ | √ | √ | √ |
| Week 5-a | Contact activity | | √ | √ | |
| Week 5-b | Workshop | √ | | √ | |
| Week 6 | Survey+rating | √ | √ | √ | √ |

## Results

To answer the research questions, the data were analyzed using a 4 (Groups) × 2 (pre and post) Repeated Measures Factorial ANOVA. A separate ANOVA was run for each of the six dependent variables. Post hoc tests were conducted for each of the groups and times.

Table 7 shows cell means for the six dependent variables, broken down by group (control, contact only, workshop only, and combined) and by time (pretest and posttest). The speech ratings were all averaged across the seven ITA speech samples for each rater.

**Table 7.** Descriptive statistics of dependent variables

| | | **Language Ideology( Mean, SD)** | **Attitude towards ITA (Mean, SD)** | **Comprehensibility (Mean, SD)** | **Accentedness (Mean, SD)** | **Language proficiency (Mean, SD)** | **Teaching ability (Mean, SD)** |
|---|---|---|---|---|---|---|---|
| **Control (N=37)** | Pre | 2.08 (0.92) | 3.89 (1.43) | 3.68 (0.88) | 1.89 (0.57) | 4.46 (1.17) | 2.81 (0.97) |
| | Post | 2.78 (1.72) | 3.78 (1.40) | 3.38 (0.89) | 2 (0.62) | 4.76 (1.12) | 2.84 (1.07) |
| **Contact only (N=30)** | Pre | 2.4 (1.07) | 3.63 (1.27) | 3.8 (0.85) | 1.9 (0.55) | 3.97 (1.07) | 3.17 (0.83) |
| | Post | 4.1 (1.56) | 3.7 (1.21) | 3.3 (0.65) | 2.17 (0.8) | 4.53 (0.90) | 2.83 (0.83) |

**Table 7** Continued

| | | **Language Ideology( Mean, SD)** | **Attitude towards ITA (Mean, SD)** | **Comprehensibility (Mean, SD)** | **Accentedness (Mean, SD)** | **Language proficiency (Mean, SD)** | **Teaching ability (Mean, SD)** |
|---|---|---|---|---|---|---|---|
| **Workshop only （N=29）** | Pre | 2.45 （1.15） | 3.41 （1.18） | 3.69 （0.81） | 1.72 (0.59） | 4.45 （0.91） | 3.34 （0.9） |
| | Post | 3.34 （1.76） | 3.79 （1.45） | 3.48 （0.83） | 2 (0.53） | 4.76 （1.02 （ | 3.07 （1.00） |
| **Combined （N=26）** | Pre | 2.65 (1.67) | 3.96 (1.25) | 3.46 (0.76) | 2.27 (0.67) | 4.38 (1.02) | 2.92 (0.74) |
| | Post | 2.62 (1.42) | 3.46 (1.30) | 3 (0.75) | 1.92 (0.80) | 4.81 (0.94) | 2.73 (0.67) |

**Effects of Interventions on UGs' Language Ideology and Attitudes Towards ITAs**

The first question is to examine the effects of interventions on UGs' language ideology and their attitudes towards ITAs. No statistically significant main or interaction effects were identified in ANOVAs of UGs' attitudes towards ITAs.

In terms of UGs' language ideology, results from a 4 (Groups) × 2 (pre and post) Repeated Measures Factorial ANOVA analysis show a significant main effect of time (F [1, 118]=34.64, P=0.00, partial eta squared=0.227). Participants' language ideology score after the intervention (Mpost=3.21, SD=0.15) was higher than that before the intervention (Mpre=2.40, SD=0.11). The time×group interaction effect was significant (F [3, 118]=6.23, P=0.00, partial eta

squared=0.14) (Figure 3). The language ideology scores for the three intervention groups were higher than that for the control group in the pretest. In the posttest, the language ideology score for the control, contact only group, and workshop only group was higher than that before the intervention, but the score for the combined group was lower. Follow-up analysis was conducted with the Tukey's HSD pairwise procedure and revealed that there is no statistically significant difference in the language ideology ratings between the four groups( F [3, 118]=1.56, p=0.21) in the pre-intervention ratings. The post-intervention ratings were significantly different between the control group and the workshop-only group (F(1, 118)=11.64; p=0.00) but not contact group and combined group.

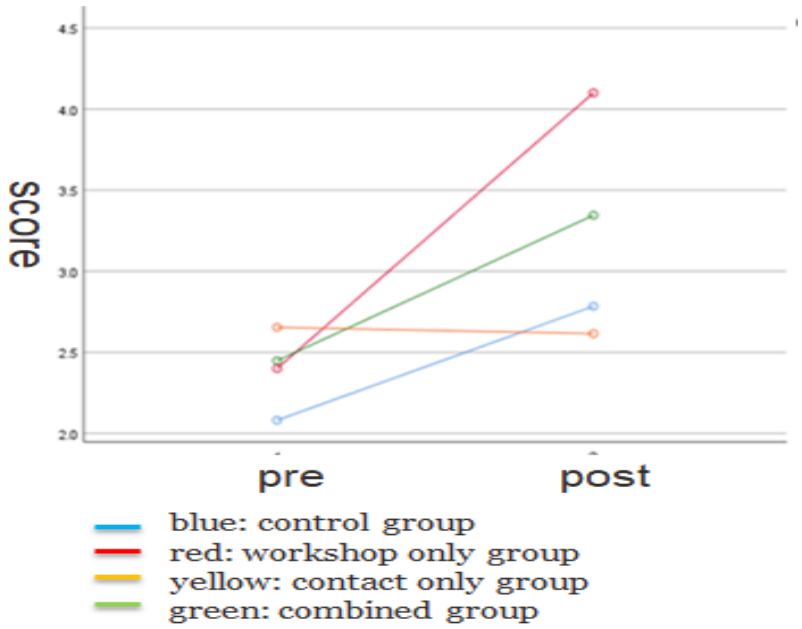


**Figure 3.** Interaction effect between time and group for language ideology

**Effects of Interventions on UGs' Perceptions of ITAs' Comprehensibility, Accentedness, Teaching ability and English Language Proficiency**

The 4 (Groups) × 2 (pre and post) Repeated Measures Factorial ANOVA for the ratings of ITAs' comprehensibility showed a statistically significant main effect for time (F [1,

118]=37.70, P=0.00, partial eta squared=0.24), indicating that UGs rated ITAs as more comprehensible on the posttest (Mpost=3.29, SD=0.072) than they did on the pretest (Mpre=3.66, SD=0.076). However, no significant interaction effect was found between time and treatment groups; thus, this score change cannot be attributed to the intervention.

For UGs' language proficiency ratings, the 4 (Groups) × 2 (pre and post) Repeated Measures Factorial ANOVA shows a statistically significant main effect of time (F [1, 118]=36.64, P=0.00, partial eta squared=0.24), indicating that UGs rated ITAs more proficient in English on the posttest (Mpost=4.71, SD=0.092) than they did on the pretest (Mpre=4.32, SD=0.096). However, no significant interaction effect was found between time and treatment groups, thus this score change cannot be attributed to the intervention.

For UGs' accentedness ratings, the 4 Group (Control, Contact-only, Workshop-only, and Combined) × 2 Time (Pre and Post) Repeated Measures Factorial ANOVA shows a statistically significant main effect of Time (F [1, 118]=7.88, P=0.01, partial eta squared=0.063), indicating that UGs rated ITAs less accented on the posttest (Mpost=2.02, SD=0.064) than they did on the pretest (Mpre=1.83, SD=0.053). However, the interaction between Time and Group was not significant, which suggests that the improvement over time cannot be attributed to the intervention.

For UGs' teaching ability ratings, the 4 (Groups) × 2 (Pre and Post) Repeated Measures Factorial ANOVA shows a statistically significant main effect of Time (F [1, 118]=8.93, P=0.00, partial eta squared=0.070); UGs evaluated ITAs' teaching ability higher on the posttest (Mpost=2.87, SD=0.084) than they did on the pretest (Mpre=3.061, SD=0.080). However, no significant interaction effect was found between time and treatment groups. Thus, this improvement cannot be attributed to the intervention.

## Discussion

### Effects of Interventions on UGs' Language Ideology and Attitudes Towards ITAs

This study revealed significant intervention effects of the workshop on UGs' self-evaluation of their language ideology. This result indicates that participation in a one-hour seminar on language ideology taught UGs to accept different varieties of English. It added to the current literature that language ideology educational practice in educational institutions is effective for improving UGs' openness towards foreign-accented English. Compared with previous educative activities about UGs' openness towards foreign-accented English which employed multiple sessions of activities in the classroom(e.g.,Eslami, Cassell, & Ates, 2015; Eslami, Moody, & Pashmforoosh, 2019), the workshop in this study was a compact and interactive tool that can be delivered independently for any educational institution.

However, I did not observe improved ratings of UGs' attitudes towards ITAs after any of the interventions. This study measured attitudes towards ITAs by examining UGs' willingness to take ITAs' class and if they believed non-native English speaking ITAs were less effective than native-English speaking TAs due to ITAs' foreign-accented English. The absence of intervention effect on UGs' attitudes towards ITAs indicated that none of these interventions helped to change their willingness to take ITAs' course and their evaluation of ITAs' teaching ability compared to native English-speaking TAs.

### Effects of Interventions on UGs' Perceptions of ITAs' Comprehensibility, Accentedness, Teaching Ability and English Language Proficiency

In this study, I observed the main effects of time but no interaction effects between time and group for these four dependent variables in all groups after the interventions. The main effect of time is likely to be caused by UGs' familiarity with these speech samples, because I used the same speech samples in pre- and post-intervention tests. In the post-intervention test, UGs were

at least more familiar with the content of the speech samples, which likely improved their perceptions of these four outcome measures.

The current study did not show significant rating changes in UGs' perceptions of ITAs' accentedness for UGs in any groups. This finding is consistent with the findings in Kang et al.'s study (2016), which suggests that perception of accentedness may be more involved with speech samples' linguistic features rather than listeners' perceptions.

In this study, contact activity did not lead to improvement of UGs' perceptions of ITAs' comprehensibility, teaching ability and English language proficiency. This result is different from previous studies where contact activities (i.e., 1-hour contact activity in Kang et al.'s study [2015] or conversation partnership activity in Staple et al's study [2014]) helped to improve UGs' perceptions of these outcome measures. This inconsistency may be explained by the research design of the intervention. In Kang et al.'s study (2015), ITAs participating in the contact activity were those whose speech samples were rated by UGs. The same type of speech samples was also used in Staples et al.'s study (2014). In contrast, in the current study, the ITAs who participated in the contact activity were different from those whose speech samples were rated by UGs. This may also explain the absence of intervention effect on the UGs' self-evaluation of their attitudes towards ITAs as measured in this study. This discrepancy between the current study and the previous studies tends to strengthen Staple et al's finding that exposure to and familiarity with specific ITAs could be helpful for making ITAs' speech easier for UGs to understand and for improving UGs' evaluation of ITAs' teaching ability and language proficiency.

In the current study, the absence of the intervention effect for UGs in the workshop group is unexpected. Nonetheless, it suggests the importance of interaction and familiarity with ITAs for improving UGs' perceptions of ITAs' speech. Conceptual education about language ideology

does not seem to be sufficient in changing UGs' perception of ITAs' speech and their teaching ability.

The non-significant changes for UGs' perceptions of ITAs' speech comprehensibility, English language proficiency or teaching ability in the combined group maybe caused by the fatigue of UGs. UGs in the combined group provided their post-intervention ratings at the end of almost two hours of intervention activities after a whole day of classes and work. This may cause fatigue for them. The self-reported nature of the measures of language ideology and attitudes makes the scores sensitive to UGs' personal conditions like fatigue.

## Limitations and Implications

This study aimed to study the effect of three types of interventions on UGs' attitudes towards foreign-accented English (i.e., language ideology) and ITAs, especially their perceptions of ITAs' comprehensibility, accentedness, English language proficiency, and teaching ability. Unfortunately, I did not observe as much improvement for these outcome measures as I expected. This may be caused by some measurement issues. In this study, I used 7-point Likert scales to measure UGs' scoring changes, but these seven points might not be sufficient to catch these changes. Recently, studies used 1-1000 sliding bars to measure comprehensibility and accentedness (e.g., Galante & Thomson, 2017; Levis & Levis, 2018), because it allows more room for listeners to locate their ratings of speech samples, which helps to catch their perception change more accurately.

The second measurement issue is concerned with statements used for measuring UGs' language ideology and their attitudes towards ITAs. The alpha value for the consistency of the statements for UGs' language ideology, and attitudes towards ITAs are 0.69, and 0.72 , respectively, which is marginally reliable. Currently, little existing research provides guidelines for measuring these two outcome variables. More research might be needed to further identify relatable elements which help to measure UGs' ideology and attitudes towards ITAs to help to

make intervention decisions (i.e., who will be the target participants). Particularly, the statements used in this study to measure language ideology may reflect UGs' attitudes toward their own language variety rather than their openness toward understanding and even accepting non-mainstream varieties of English.

Another limitation is the lack of qualitative data. Focus group interviews or surveys could have been used to further explore how UGs' language ideology, and attitudes towards ITAs might be changed by these interventions. Information from interviews and surveys may help to identify changes in UGs' language ideology and attitudes towards ITAs, which might not be detected by quantitative data due to the measurement issues mentioned above. Questions about how they feel about these interventions in interviews and surveys may also help to improve the design of these intervention activities.

In addition to adding qualitative data, an advanced data-analysis method may be used for analyzing UGs ratings of speech samples. Instead of aggregating ratings of seven speech samples, future studies might use multilevel models to model scores for each speech sample. This would help to capture the variation of UGs' scores across the speech samples, and may shed more light on the characteristics of the speech samples that may affect change or lack of change of UGs' perceptions of their comprehensibility.

In spite of the limitations above, this study contributes to the body of research by examining the impact of contact activities on UGs' perceptions of ITAs teaching speech. The findings of this study imply that structured contact activities conducted between UGs and ITAs could contribute to more effective communication between ITAs and UGs. For example, departments may hold a "Meet Your Teachers" event and invite both UGs and ITAs to attend. In this event, they can play the game used in the contact activity in this study. This event is supposed to be scheduled before UGs register for the courses in the incoming semester to avoid some issues caused by extensive contact between professors and UGs beyond academic affairs

during the semester. At the same time, the findings of this study imply that promoting educational practice such as a workshop about language ideology in educational institutes could help to improve UGs' acceptance of foreign-accented English.

# 4. LINGUISTIC AND PSYCHOSOCIAL FACTORS AFFECTING AMERICAN UNDERGRADUATES' EVALUATION OF INTERNATIONAL TEACHING ASSISTANTS

## Introduction

The number of international teaching assistants in the U.S. has increased dramatically in recent decades. Many undergraduates (UGs) perceive ITAs as unqualified to teach, in part due to their negative perceptions of ITAs' non-mainstream English. This negative attitude towards ITAs creates a mental barrier that inhibits students' ability to understand their ITAs (Rubin, 1992). Therefore, it is important for us to shift our research attention from ITAs to undergraduates to facilitate effective communication between them.

### Linguistic and Nonlinguistic Factors Predicting NNESs' Speech Comprehensibility

Previous studies have investigated the linguistic influences on the comprehensibility of L2 English based on listeners' judgements. For example, in Saito and colleagues' study (2017), twenty native English speakers rated forty French English learners' picture narratives. Their rating results indicated that vowel/consonant errors, word stress, fluency, lexis and grammar were associated with comprehensibility. Kang, Rubin and Pickering (2010) examined how some suprasegmental measures predicted NNESs' comprehensibility. In their study, 188 monolingual U.S. undergraduates rated NNESs' oral summarization of a passage. The rating results indicated that suprasegmental fluency, rising tones, and boundary markers were statistically significant predictors for NNESs' speech comprehensibility.

NESs may fail to understand NNESs, simply because they readily assumed that NNESs' foreign-accented English was difficult to understand even though ITAs' English was understandable (Derwing & Munro, 1997, 2009; Munro & Derwing, 1995a). Evidence of this negative association has been found in previous studies. For example, in Lindemann's study

(2002), NESs with either positive or negative attitudes towards Koreans were asked to complete an interactive map task with Korean English speakers. The analyses about NESs' communication strategies used in the tasks suggested that NESs with positive attitudes towards Koreans were more active in the interaction than NESs with relatively negative attitudes.

### The Current Study

While a growing number of studies have examined linguistic dimensions of comprehensibility evaluated by inexperienced and experienced raters, few studies have used speech samples from ITAs' teaching performance to make their results more relatable to undergraduates' judgements about ITAs' speech comprehensibility and teaching ability. While the issues of UGs' attitudes towards ITAs have been widely discussed in many studies (e.g., Rubin, 2001), few empirical studies have investigated the predictive factors of UGs' attitudes towards ITAs' affecting UGs' perceptions of ITAs' speech comprehensibility and teaching ability.

This study aims to fill this gap in literature by examining how language and non-language factors affecting UGs' judgements of ITAs' speech comprehensibility and teaching ability. This study aims to answer the following questions:

1. How are ITAs' segmental and suprasegmental errors related to UGs' perceptions of ITAs' speech comprehensibility and teaching ability?
2. How are UGs' perceptions of ITAs' speech comprehensibility and teaching ability related to UGs' attitudes towards ITAs?

## Methods

### Participants

*Undergraduate raters*

One hundred and ninety-eight undergraduates were recruited from about 40 different academic departments across TAMU campus. They were recruited through the university

internal email systems (detailed recruitment process was explained in the previous study; see p.47). See Table 8 for the demographic information of these undergraduate raters.

**Table 8.** Undergraduates' demographic information

| | | **Number** | **Percentage** |
|---|---|---|---|
| **Gender** | Male | 65 | 32.83% |
| | Female | 131 | 66.16% |
| | Undisclosed | 2 | 1% |
| **Academic Year** | First-year | 68 | 34.34% |
| | Sophomore | 32 | 16.16% |
| | Junior | 45 | 22.73% |
| | Senior | 53 | 26.77% |
| **Language background** | Monolingual English speaker | 163 | 82.32% |
| | Multilingual speaker | 35 | 17.68% |
| **Contact experience with ITAs** | Never taught by ITSs | 63 | 31.82% |
| | Taught by ITAs whom UGs did not have any issues with | 132 | 66.67% |
| | Taught by ITAs whom UGs had any issues with | 3 | 1.5% |
| **Total** | / | 26 | 100% |

*Speakers*

The seven speakers were recruited from an international teaching assistant training program from the Center of Teaching Excellence (CTE) at TAMU. These participants received training from CTE and gave micro-teaching presentations at the end of the training. They all signed consent forms to allow the researcher to use their recorded micro-teaching presentations in the research. See Table 5 for the demographic information about these speakers.

**Measurement**

*Comprehensibility*

Comprehensibility ratings were collected while UGs listened to the speech samples. All participants listened to and rated a sample from a native English speakers, followed by all seven speech samples in a random order.

Following the conventions in the previous studies (e.g., Kang, Rubin, & Lindemann, 2015), five semantic differential measures items were used to test comprehensibility. These items included statements such as: this speaker (1) was easy to understand; (2) was comprehensible; (3) was clear; (4) required little effort to understand; (5) made it simple to grasp the meaning. A 7-point Likert scale was used to measure each sub-measure of the comprehensibility with 1 representing high comprehensibility and 7 representing low comprehensibility. These five sub-measures have a strong reliability of 0.91.

*Evaluation of ITAs' teaching* ability

ITAs' teaching ability was measured by two semantic differential scales. These two sub-measures were adapted from an earlier scale of ITAs' instructional quality (Kang et al., 2015 & Staple et al., 2014). The earlier version of this measure included nine sub-items such as teaching effectiveness, teaching qualification, classroom management, availability and knowledge materials. Because the speech samples in this study were only 2-minute excerpts selected from simulated teaching presentations rather than authentic classroom lectures, it is not valid for UGs

to rate all of the nine sub-items. I therefore used 7-point Likert scales to indicate how much they agreed or disagreed to the following two statements: a) this speaker is an effective teacher; b) this speaker is a qualified teacher. I used 1 to represent high teaching ability and 7 to represent low teaching ability. These two sub-measures have a strong reliability of 0.84.

*Attitudes towards ITAs*

Two sub-statements with a 7-point Likert scale were designed to measure UGs' attitudes towards ITAs with 1 representing negative attitudes and 7 representing positive attitudes. UGs needed to choose from the 7-point Likert scale ratings to indicate how much they agreed or disagreed with the statements: (a) If there are multiple sections of a course, I will take the section taught by a native English-speaking teaching assistant instead of a non-native English-speaking teaching assistant; (b) Native English-speaking teaching assistants are usually more effective in teaching than non-English speaking teaching assistants due to the latter's foreign-accented English. These two sub-measures have the marginally acceptable reliability of 0.72.

*Experience with ITAs*

Information about UGs' experience with ITAs was obtained by having them specify how many courses taught by international instructors/international teaching assistants they had taken before this semester. They may enter 0, if they have not taken any course taught by international instructors/international teaching assistants.

*ITAs' segmental and suprasegmental errors*

ITAs' segmental and suprasegmental errors were evaluated with five questions of how they rated the speakers' segmental errors, word stress errors, intonation errors, speech rate, and inappropriate pauses (Crowther, Trofimovich, Saito, & Isaacs, 2015). Segmental errors were defined as the incorrect pronunciation of a sound in a word. Word stress errors were defined as placing the stress on the wrong syllable within a word. Intonation errors were defined as using the wrong pitch for a sentence. Inappropriate pauses were defined as placing pauses in a wrong

place of a sentence or no pause in a sentence. UGs were provided a 1-10 scales to indicate how frequently the ITA made segmental, word stress, intonation errors and inappropriate pauses (1 indicated frequent and 10 indicated infrequent), and to indicate how fast the ITA spoke (1 indicated too slow and 10 indicated too fast).

**Materials**

*Speech samples*

There were 8 speech samples used in the study. Each speech sample was a 2-minute long excerpt from an 8-minute long videotaped micro-teaching lesson about a specific topic such as sociology, political science, or nutrition. To minimize the ceiling effect, one of the speech samples was from an NES about how to develop leadership. I placed this sample at the beginning before all the other samples. The rest of the seven micro-teaching videos were presented by the seven ITAs. These excerpts were all converted into audio recordings. These 2-minute excerpts were carefully chosen from the original oral presentations by considering the following factors: a) the content was neither too abstruse nor including vocabulary words beyond UGs' high school knowledge level about the specific subject; b) selected excerpts were continuous streams of speech; c) there was no interruption from the audience or background noise. A pilot study had been conducted with two UGs outside of the study. They listened to these speech samples and had confirmed that these samples were understandable in content and of high quality in sound effect.

*Rating sheets*

On a question sheet, I printed questions regarding speakers' comprehensibility, teaching ability, and segmental and suprasegmental errors. On another sheet, I printed eight tables for UGs to write their answers to all the questions for each speech sample (see Appendix A).

*Survey*

The background survey consisted of demographic questions about UGs' gender, major and academic year. After these questions, UGs were asked to indicate their experience with ITAs and their attitudes towards ITAs. This survey was designed in a digital form through Qualtrics.

**Data Collection**

The whole data collection process was conducted in a language lab at CTE within one hour. In the first 10 minutes, one of the researchers gave UGs specific instructions about the basic procedures and explained each question. When explaining items about segmental and suprasegmental errors, the researcher gave definitions and one or two examples to illustrate what exactly was meant by segmental errors, word stress errors, intonation error, and inappropriate pauses. In case that UGs may forget the meaning of these items, a brief note about the definition and examples of each item was provided below the relevant questions.

After the directions, UGs started completing the Qualtrics survey first. Then UGs used a desktop to listen to the samples with a headphone. While they were listening to the speech samples for the first time, UGs were asked to answer questions about how they evaluated speakers' comprehensibility and teaching ability. Then UGs listened to each sample again and evaluated ITAs' segmental and suprasegmental errors. These evaluation questions were printed out and UGs provided their written questions.

There were six desk computers in the language lab, so there were at most six UGs participating in the study at one time. In order to avoid time conflict with the normal business of the language lab at CTE, the data collection was scheduled from 8:00 to 9:00 in the morning before the normal business hours. The whole study lasted about one year, including the recurring recruitment process and the data collection process.

## Results

### Data Analyses

The speech ratings were all averaged across the 7 ITA speech samples for each rater. To answer the research question, I performed two groups of hierarchical multiple regression analyses with Stata 16 software. In each group of regression models, ITAs' speech comprehensibility and teaching ability served as dependent variables, respectively. Each group of models contained five hierarchical multiple regression models regarding five items related to ITAs' segmental and suprasegmental errors. For each group of models, I first entered the variable about UGs' experience with ITAs because previous research had provided abundant evidence showed that raters' previous experience with speakers significantly correlated to their ratings of speakers' speech (Huang, 2013; Isaacs, Thomson, 2013). In the second step, I included one of the five variables about UGs' ratings of ITAs' segmental and suprasegmental errors. In the last step, I included the variable about UGs' attitudes towards ITAs. Table 9 shows the descriptive data and intercorrelations of all these variables. Tables 10 & 11 show the results of multiple regression analyses for each dependent variable.

**Table 9.** Descriptive data and intercorrelations of variables (N=198)

| | **1.** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
|---|---|---|---|---|---|---|---|---|---|
| **1.Comprehensibility** | / | | | | | | | | |
| **2.Teaching ability** | 0.61* | / | | | | | | | |
| **3. Previous experience with ITAs** | -0.14 | -0.14* | / | | | | | | |

**Table 9** Continued

| | **1.** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** |
|---|---|---|---|---|---|---|---|---|---|
| **4.Segmental errors** | -0.55* | -0.46* | 0.09 | / | | | | | |
| **5.Word stress** | -0.58* | -0.50* | 0.14 | 054.* | / | | | | |
| **6.Intonation errors** | -0.49* | -0.49* | 0.12 | 0.34* | 0.54* | / | | | |
| **7.Speech rate** | -0.11 | -0.06 | -0.06 | 0.09 | 0.15* | 0.06 | / | | |
| **8.Inappropriate pauses** | 0.01 | -0.01 | -0.22* | 0.08 | 0.04 | 0.13 | 0.17* | / | |
| **9. Attitudes towards ITAs** | -0.31* | -0.35* | 0.14 | 0.26* | 0.19* | 0.24* | 0.06 | 0.03 | / |
| **Mean** | 3.66 | 3.12 | 2.33 | 6.19 | 5.89 | 6.68 | 5.22 | 2.53 | 3.66 |
| **SD** | 0.87 | 0.89 | 2.87 | 0.87 | 1.23 | 1.10 | 1.04 | 2.58 | 1.39 |

**Predicting Effect of ITAs' segmental and suprasegmental Errors on UGs' Perceptions**

The first research question concerned the relationship between ITAs' segmental and suprasegmental errors and UGs' perceptions of ITAs' speech comprehensibility and teaching ability. First, as shown in Tables 10 and 11, I found that UGs' previous experience with ITAs was a statistically significant factor affecting their perceptions of ITAs' teaching ability ($p <$

0.5), but not of ITAs' speech comprehensibility. The more courses UGs have taken with ITAs, the more positive UGs will perceive for ITAs' teaching ability.

Second, in terms of the five sub-measures of ITAs' segmental and suprasegmental errors, I found that ITAs' segmental errors, word stress errors, and intonation errors statistically significantly predicted two dependent variables when their previous experience with ITAs was controlled. When more segmental, word stress and intonation errors were perceived, the ITAs were predicted to have lower comprehensibility and teaching ability (See Table 10). However, ITAs' speech rate and their pauses did not show any statistically significant relationship with the three dependent variables.

**Predicting Effect of UGs' Attitudes Towards ITAs on UGs' Perceptions**

The second research question concerned the relationship between UGs' attitudes towards ITAs and UGs' perceptions of ITAs' speech comprehensibility and teaching ability. Tables 10 &11 show that UGs' perceptions of ITAs' speech comprehensibility and teaching ability were significantly affected by their attitudes towards ITAs when their previous experience with ITAs and ITAs' speech errors were controlled. A more positive attitudes from UGs towards ITAs predicted UGs' perceptions of higher comprehensibility and teaching ability of ITAs,

**Table 10.** Results of multiple regression models predicting ITAs' speech comprehensibility

| **Models** | **Step 1: UGs' previous experience with ITAs** | | | **Step 2: segmental and suprasegmental errors** | | | **Step 3: UGs' attitudes towards ITAs** | | |
|---|---|---|---|---|---|---|---|---|---|
| | Beta | SE | $R^2$ | Beta | SE | $\triangle R^2$ | Beta | SE | $\triangle R^2$ |
| **Previous experience** | -0.04 | 0.02 | 0.02 | -0.01 | 0.02 | 0.437** | -0.00 | 0.02 | 0.015** |

**Table 10** Continued

| **Models** | **Step 1: UGs' previous experience with ITAs** | | | **Step 2: segmental and suprasegmental errors** | | | **Step 3: UGs' attitudes towards ITAs** | | |
|---|---|---|---|---|---|---|---|---|---|
| **a.segmental errors** | / | / | / | -0.32** | 0.06 | | -0.29** | 0.06 | |
| **b.word stress** | | | | -0.19** | 0.05 | | -0.20** | 0.05 | |
| **c.intonation errors** | | | | -0.19** | 0.05 | | -0.17** | 0.05 | |
| **d. speech rate** | | | | -0.03 | 0.05 | | -0.02 | 0.05 | |
| **e. inappropriate pauses** | | | | -0.02 | 0.02 | | -0.03 | 0.02 | |
| **Attitudes** | **/** | **/** | **/** | / | / | / | -0.08* | 0.03 | |

*p<0.05; **p<0.01

**Table 11.** Results of multiple regression models predicting ITAs' teaching ability

| **Models** | **Step 1: UGs' previous experience with ITAs** | | | **Step2: segmental and suprasegmental errors** | | | **Step 3: UGs' attitudes towards ITAs** | | |
|---|---|---|---|---|---|---|---|---|---|
| | Beta | SE | $R^2$ | Beta | SE | $\triangle R^2$ | Beta | SE | $\triangle R^2$ |
| **Previous experience with ITAs** | -0.04* | 0.02 | 0.02 | -0.01 | 0.02 | 0.345** | -0.01 | 0.02 | 0.03** |
| **a.segmental errors** | | | | -0.25** | 0.07 | | -0.21** | 0.07 | |
| **b.word stress** | | | | -0.15** | 0.06 | | -0.15** | 0.05 | |
| **c.intonation errors** | | | | -0.24** | 0.06 | | -0.21** | 0.06 | |
| **d. speech rate** | | | | 0.00 | 0.05 | | 0.01 | 0.05 | |
| **e. inappropriate pauses** | | | | 0.02 | 0.02 | | 0.02 | 0.02 | |
| **Attitudes** | / | | / | / | / | / | -0.13 | 0.04 | |

*p<0.05; **p<0.01

## Discussion

This study found that UGs' previous experience with ITAs was a statistically significant factor affecting their perceptions of ITAs' teaching ability. This means when UGs took more courses with ITAs, they would perceive higher ability for ITAs' teaching. This conclusion makes more sense when I further investigate UGs' contact experience with ITAs. In this study, 68.17%

(N=135) of UGs had taken courses with ITAs and only three of them reported that they had issues with ITAs and chose to drop the course taught by ITAs (see Table 8). This means that most of UGs had non-negative experience with ITAs, which could help to explain that UGs' previous experience with ITAs contributed to their evaluation of ITAs' teaching ability. However, in this study UGs' previous experience with ITAs was not a statistically significant predictor for UGs' perception of ITAs' speech comprehensibility. This finding was consistent with that in previous studies which supported that NESs' experience with NNESs did not affect NESs' perception of NNESs' comprehensibility (Munro, Derwing, & Morton, 2006; Kennedy & Trofimovich, 2008).

In this study, I observed that comprehensibility was associated with ITAs' segmental errors, word stress and intonation but not with their speech rates and inappropriate pauses. This finding was not consistent with that of Saito, Trofimovich and Isaacs' research (2017). In their study, Saito et al. (2017) found that comprehensibility was associated with various sources of linguistic information including speech rates, pauses as well as vowel/consonant errors, word stress and intonations. The absence of the role of fluency and pauses in predicting the speech comprehensibility in this study was likely attributed to the measurement issues for "speech rate" and "pauses". In the current study, ITAs' speech rate was measured by UGs' subjective judgements of how fast or slow ITAs' speech was. Similarly, for the construct of pauses which was defined as placing pauses in a wrong place of a sentence or no pause in a sentence, the judgement for appropriate pauses varied among UGs. The inconsistent criteria of fluency and pauses for each UGs confounded the results of their judgement, which messed their relationships with ITAs' comprehensibility. In contrast, vowel/consonant errors, word stress, and intonations were more consistently defined for UGs in the current study. This made the results associated with these three variables more reliable and thus consistent with previous studies. This

explanation can also be applied to the mixed predictive effect of different linguistic factors for ITAs' teaching ability.

This study also found that UGs' perceptions of ITAs' comprehensibility and teaching ability were predicted by their attitudes towards ITAs. This finding confirmed previous studies' findings that NESs' attitudes towards speakers affected their perceptions of NNESs' speech and characteristics such as intelligence and ability (Lindemann & Subtirelu, 2013). This finding also added evidence to demonstrate that the blame of ITAs' lower comprehensibility and poorer teaching ability may be placed on UGs' biased attitudes towards ITAs to some extent. In this study, I did find that ITAs' segmental and suprasegmental errors predicted UGs' perceptions of ITAs' speech comprehensibility and teaching ability. However, when these linguistic factors were controlled for, UGs' attitudes still played a role in their perceptions of both outcome variables. This indicated that UGs' evaluation of ITAs' speech comprehensibility and their teaching ability were jointly affected by ITAs' speaking ability and UGs' attitudes towards ITAs.

## Implications

The findings of this study reveal the effects of UGs' attitudes towards ITAs on UGs' perceptions of ITAs' speech comprehensibility and teaching ability. This means that improving UGs' attitudes towards ITAs is of great importance to enhance effective communication between UGs and ITAs, and to establish UGs' trust in ITAs' teaching ability. First of all, UGs' academic consultants need to encourage UGs to take courses taught by ITAs and serve as a mediator between UGs and ITAs. The present study showed that when UGs took more courses with ITAs, they gave higher evaluations of ITAs' teaching ability. The exposure with ITAs provided UGs opportunities to understand ITAs, which helps to reduce UGs' stereotypes of ITAs. At the same time, academic consultants need to serve as an intermediary between UGs and ITAs to help to address any issues between them so as to facilitate a positive relationship with them (Plakans, 1997). As indicated by the findings of this study, when UGs had more positive contact

experience with ITAs, they felt it was easier to understand ITAs' speech. This will definitely benefit UGs' academic learning in courses taught by ITAs. Second, in addition to taking courses with ITAs, UGs should be encouraged to participate in programs creating opportunities for UGs to interact with ITAs. These programs should be mutually beneficial and convenient for both UGs and ITAs in order to ensure positive experiences between them. Conversation partnership between UGs and ITAs which enhance multicultural awareness of UGs and improved ITAs' oral communication ability, and structured 1-hour contact activity which promote a cooperative relationship between UGs and ITAs within short time period would be good options (Kang, Rubin & Lindemann, 2016; Staples, Kang, & Wittner, 2014)

What is equally important for successful communication between UGs and ITAs is ITAs' role in the cross-cultural communication in their classrooms. Researchers from critical applied linguistics have pointed out that language use is also associated with social, cultural and political dynamics (Kubota, 2012). In terms of conversation partnership activities, the host departments should consider the formation of cultural practice as well as the construction of cultural knowledge. This means in addition to including useful topics about cultural differences (e.g. educational practice) between U.S. and other countries where ITAs came from to help them to gain cultural knowledge, the conversation partnership activity should include activities to teach ITAs how to adopt the target cultural practices, such as how to show your care or respect to American UGs in an appropriate way.

The methodological issues about the measurement of speech rate and pauses indicate that specific guidelines and explicit and detailed training for speech raters, especially inexperienced raters, is important for the reliability of the speech rating (Isaacs & Thomson, 2013). Inexperienced raters benefit from more guidance about the nature of the construct being measured for reliability reasons. Researchers may use multiple examples to explain ratings of

different score points, and provide handy notes for raters to refer to during the rating process to help to improve the reliability of the rating.

# 5. CONCLUSIONS

In this dissertation, these three independent but relevant studies worked together to highlight an important fact that successful communication is jointly constructed by both speaker and listener (Lindemann & Subtirelu, 2013; Lindsey, King, Hebl, & Levine, 2015; Rubin, 1992). In addition to taking efforts to improve ITAs' speaking ability, we should create effective programs to help UGs improve their attitudes towards ITAs and their perceptions of ITAs' speech according to the intelligibility principle.

This dissertation started with a meta-analytic review in which I evaluated the effectiveness of interventions focusing on either NNESs or NESs to improve NNESs' overall intelligibility (i.e. intelligibility or comprehensibility). In this review, the robust variation estimations of 24 independent studies identified significant effects for all interventions (g=0.40), and interventions for NNESs (g=0.47) and NESs (g=0.28), respectively. This review further revealed that interventions helped to understand NNESs more easily but not better, and that spontaneous rather than controlled speech activities yielded significant effects for interventions. This review also found that rating scale length used in the comprehensibility measurement did not affect the effect of interventions.

The second study of the dissertation examined the impact of two intervention activities: contact activity and language diversity workshop, on improving American UGs' attitudes towards ITAs and their perceptions of ITAs' speech. One hundred and twenty-two TAMU undergraduates from various academic departments were randomly assigned to control group and three intervention groups where they participated in different activities: contact activity only, workshop only, contact activity and workshop combined. This study found that language ideology workshop was helpful for improving UGs' openness towards ITAs' foreign-accented English. Surprisingly, in this study, none of the interventions had statistically significant impact on improving UGs' attitudes towards ITAs and their perceptions of ITAs' comprehensibility,

accentedness, English language proficiency, and teaching ability. These unexpected results revealed the importance of UGs' exposure to and familiarity with ITAs and ITAs' accented speech in helping UGs understand ITAs more easily and mitigating UGs and ITAs' relationship.

The last study of the dissertation investigated linguistic and nonlinguistic factors predicting American UGs' evaluation of ITAs' comprehensibility and teaching ability. I collected data from 198 TAMU UGs regarding their previous experience with ITAs, their ratings of seven ITAs' speech samples in terms of ITAs' segmental and suprasegmental errors, and their attitudes towards ITAs. Hierarchical multiple regression analyses indicated that when UGs took more courses with ITAs, they would perceive higher ability for ITAs' teaching, and that comprehensibility was associated with ITAs' segmental errors, word stress and intonation errors. This study also found that UGs' perceptions of ITAs' comprehensibility and teaching ability were uniquely predicted by their attitudes towards ITAs after UGs' previous experience with ITAs and ITAs' segmental and suprasegmental errors were statistically controlled. These results stressed the importance of improving UGs' attitudes towards ITAs.

Taken together, the results from these three studies had several implications for future educational practices aiming to enhance effective communication between American UGs and ITAs. First, interactive and compact workshops about language ideology should be advocated in educational institutions. This type of workshop can be embedded as one of lectures in freshmen orientation programs. An early intervention will be more helpful for UGs to adapt themselves to the increasing diversity on campus. This will also help them to develop positive expectations towards ITAs and ITAs' courses.

Second, structured contact activities between UGs and ITAs should also be promoted by UGs academic departments and ITAs training programs. For example, before UGs start registering courses for the upcoming semester, UGs' academic department may hold an event for UGs and ITAs to meet together and play the game used in the second study. Also, a long-term

conversation partnership could be established between UGs and ITAs. For UGs, it is a great opportunity for them to become familiar with foreign-accented English and to gain a deeper understanding about cultures other than their own.

Third, ITA training programs should include sessions focusing on improving ITAs' classroom communicative strategies. In this dissertation, UGs' perceptions of ITAs' speech were only based on ITAs' audio materials. There is no doubt that ITAs' comprehensibility and intelligibility would be greatly improved if their actual presentations were shown to UGs. In the actual presentation, ITAs may use their gestures and other visual materials to present their lectures especially when their language ability is relatively limited. Unfortunately, due to the space and time limitation in this dissertation, I was not able to examine this subject. Future study should investigate effective methods or techniques which could help ITAs improve their classroom communicative competence.

Last but not the least, it is of equal importance to provide NESs with instruction on communicative strategies. Previous studies have recognized several effective strategies used by NESs, including (1) using intonation to indicate the problems in the communication (Mauranen, 2006; Watterson, 2008); (2) asking specific questions or requesting for clarification (Kaur, 2010; Kirkpatrick, 2011; Mauranen, 2006; Watterson, 2008); (3) repairing communication difficulties by intentionally connecting to the topic and context, modifying the segmental phonemes of the utterances (O'Neal, 2015 & 2016), clearly correcting the misused words and rephrasing the sentences (Watterson, 2008; Pickering, 2009; Kaur, 2010; Kirkpatrick, 2011; Matsumoto, 2011); and (4) preventing difficulties in communication by checking understanding of the listeners (Kaur, 2010) and utilizing accommodations such as adjusting speed of speech and avoiding the use of idioms (Seidlhofer, 2009). Further studies are needed to explore effective techniques to teach NESs how to use these strategies in their communication with NNESs.

## REFERENCES

*indicates articles included in the meta-analyses

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In Watkins, M. A., Rauber, A. S., & Baptista, B. O. (Eds.) *Recent research in second language phonetics/phonology: Perception and production* (pp. 231-248). Cambridge Scholars Publishing, Newcastle.

Allport, G. W. (1954). *The nature of prejudice*. Reading, Addison Wesley, Boston.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, 133(3), EL174-EL180.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford university press, Oxford.

Boyd, S. (2003). Foreign-born teachers in the multilingual classroom in Sweden: The role of attitudes to foreign accent. *International journal of bilingual education and bilingualism*, 6(3-4), 283-295.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707-729.

Chiang, S. Y. (2009). Dealing with communication problems in the instructional interactions between international teaching assistants and American college students. *Language and Education*, 23(5), 461-478.

Chiang, S. Y., & Mi, H. F. (2008). Reformulation as a strategy for managing 'understanding uncertainty' in office hour interactions between international teaching assistants and American college students. *Intercultural education*, 19(3), 269-281.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, 116(6), 3647-3658.

Cooper, B., Payne, G., Hu, X., Dixon, Q., & Kuo, L. (forthcoming). The impact of linguistic diversity education on L1 English speakers' ideologies, attitudes, and perceptions of international teaching assistants. In O. Kang, S. Staples, K. Hirschi, & K. Yaw (eds.), *Proceedings of the 11th Pronunciation in Second Language Learning and Teaching Conference*. Ames: Iowa State University.

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility?. *The Modern Language Journal*, 99(1), 80-95.

Crystal, D. (2012). *English as a global language*. Cambridge: Cambridge university press.

Crystal, D. (2003). *English as a Global Language* (Second edition). Cambridge University Press, Cambridge。

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in second language acquisition*, 19(1), 1-16.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL quarterly*, 39(3), 379-397.

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

Derwing, T. M., Fraser, H., Kang, O., & Thomson, R. I. (2014). L2 accent and ethics: Issues that merit attention. In A. Mahboob & L. Barratt (Eds.), *Englishes in multilingual contexts* (pp. 63–80). Springer, Dordrecht.

*Derwing, T. M., Rossiter, M. J., & Munro, M. J. (2002). Teaching native speakers to listen to foreign-accented speech. *Journal of Multilingual and Multicultural Development*, 23(4), 245-259.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language learning*, 54(4), 655-679.

Dovidio, J. F., Ten Vergert, M., Stewart, T. L., Gaertner, S. L., Johnson, J. D., Esses, V. M., & Pearson, A. R. (2004). Perspective and prejudice: Antecedents and mediating mechanisms. *Personality and Social Psychology Bulletin*, 30(12), 1537-1549.

Dragojevic, M., & Giles, H. (2016). I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research*, 42(3), 396–420.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, 315(7109), 629-634.

Eslami, Z. R., Cassell, E., & Ates, B. (2015). Changing Preservice Teachers' Attitudes Toward Linguistic Diversity by Introducing a World Englishes Perspective. In *Teaching at Work* (pp. 81-103). Sense Publishers, Rotterdam.

Eslami, Z., Moody, S., & Pashmforoosh, R. (2019). Educating Pre-Service Teachers about World Englishes: Instructional Activities and Teachers' Perceptions. *TESL-EJ*, 22(4), n4.

Faez, F. (2012). Linguistic identities and experiences of generation 1.5 teacher candidates: Race matters. *TESL Canada Journal*, 29 (6), 124-124.

Firth, A. (1996). The discursive accomplishment of normality: On 'lingua franca' English and conversation analysis. *Journal of pragmatics*, 26(2), 237-259.

Flege, J. E. and Fletcher, K. L. 1992. Talker and listener effects on degree of perceived foreign accent. *Journal of the Acoustical Society of America*, 91(1): 370–389.

Fraser, C., & Kelly, B. F. (2012). Listening between the lines: Social assumptions around foreign accents. *Australian Review of Applied Linguistics*, 35(1), 74-93.

*Galante, A., & Thomson, R. I. (2017). The effectiveness of drama as an instructional approach for the development of second language oral fluency, comprehensibility, and accentedness. *TESOL Quarterly*, 51(1), 115-142.

Gallego, J. C. (1990). The intelligibility of three nonnative English-speaking teaching assistants: An analysis of student-reported communication breakdowns. *Issues in Applied Linguistics*, 1(2), 219-237.

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and social psychology review*, 14(2), 214-237.

Gorsuch, G. (2018). A Teaching Practicum Course and its Effects on International Teaching Assistants' Discourse Intonation. *TESL-EJ*, 22(2), 1-36.

Gorsuch, G., Meyers, C. M., Pickering, L., & Griffee, D. T. (2011). English communication for international teaching assistants. *The Electronic Journal for English as a Second Language*, 14(4), 1-3.

*Graff, M. M. (2006). *A study of Rosetta Stone's effectiveness on improving English pronunciation.* California State University, Dominguez Hills.

Hansen, K., Rakić, T., & Steffens, M. C. (2014). When actions speak louder than words: Preventing discrimination of nonstandard speakers. *Journal of Language and Social Psychology*, 33(1), 68-77.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1 (1), 39-65.

Huang, B. H. (2013). The effects of accent familiarity and language teaching experience on raters' judgments of non-native speech. *System*, 41(3), 770-785.

Hustad, K. C. (2006). A closer look at transcription intelligibility for speakers with dysarthria: Evaluation of scoring paradigms and linguistic errors made by listeners. *American Journal of Speech-Language Pathology*, 15(3), 268-277.

Huthaily, K. Y. (2008). Second language instruction with phonological knowledge: Teaching Arabic to Speakers of English (Unpublished doctoral dissertation). The University of Montana: Missoula.

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, 10(2), 135-159.

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

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(1), 145-160.

Jenkins, J. (2002). A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied linguistics*, 23(1), 83-103.

Jenkins, J. (2006). Global lntelligibility and local diversity: Possibility or Poroolox. In Rubdy, R., & Saraceni, M. (Eds.). *English in the World: Global Rules, Global Roles* (pp. 32-68). London: A&C Black.

*Jiang, Y. (2018). *Development and effectiveness of online intonation training modules to improve Chinese speakers' English speech* (Unpublished doctoral dissertation), University of California: Santa Barbara.

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.

Kang, O., Rubin, D. O. N., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94(4), 554-566.

*Kang, O., Rubin, D., & Lindemann, S. (2015). Mitigating US undergraduates' attitudes toward international teaching assistants. *TESOL Quarterly*, 49(4), 681-706.

Kaur, J. (2010). Achieving mutual understanding in world Englishes. *World Englishes*, 29(2), 192-208.

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.

Kennedy, S., & Trofimovich, P. (2019). Comprehensibility: A Useful Tool to Explore Listener Understanding. *Canadian Modern Language Review*, 75(4), 275-284.

Kirkpatrick, A. (2011). English as an Asian lingua franca and the multilingual model of ELT. *Language Teaching*, 44(2), 212-224.

Kubota, R., & Lin, A. (2006). Race and TESOL: Introduction to concepts and theories. *TESOL quarterly*, 40(3), 471-493.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), 345-366.

Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, 45(2), 152-168.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369-377.

Levis, J. M. (2006). Pronunciation and the assessment of spoken language. In Hughes, R (Ed). *Spoken English, TESOL and Applied Linguistics: Challenges for Theory and Practice* (pp. 245-270). Springer: London.

*Levis, J. M., & Levis, G. M. (2018). Teaching High-Value Pronunciation Features: Contrastive Stress for Intermediate Learners. *CATESOL Journal*, 30(1), 139-160.

Li, S., Taguchi, N., & Xiao, F. (2019). Variations in Rating Scale Functioning in Assessing Speech Act Production in L2 Chinese. *Language Assessment Quarterly*, 16(3), 271-293.

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.

Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of sociolinguistics*, 7(3), 348-364.

Lindemann, S., & Subtirelu, N. (2013). Reliably biased: The role of listener expectation in the perception of second language speech. *Language Learning*, 63(3), 567-594.

*Lindemann, S., Campbell, M. A., Litzenberg, J., & Subtirelu, N. C. (2016). Explicit and implicit training methods for improving native English speakers' comprehension of nonnative speech. *Journal of Second Language Pronunciation*, 2(1), 93-108.

Lindsey, A., King, E., Hebl, M., & Levine, N. (2015). The impact of method, motivation, and empathy on diversity training effectiveness. *Journal of Business and Psychology*, 30(3), 605-617.

Lippi-Green, R. (1997). Teaching children how to discriminate: What we learn from the Big Bad Wolf. In Lippi, R., Donati, S., Lippi-Green, R., & Donati, R (Eds). *English with an Accent: Language, Ideology, and Discrimination in the United States* (pp.79-103). Psychology Press, New York.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-analysis*. SAGE publications, Inc.

Liu, Q., & Fu, Z. (2011). The Combined Effect of Instruction and Monitor in Improving Pronunciation of Potential English Teachers. *English Language Teaching*, 4(3), 164-170.

*Liu, Y. (2008). *The effectiveness of integrating commercial pronunciation software into an ESL pronunciation class* (Unpublished master thesis). Iowa State University, Ames.

Looney, S. D. (2015) Interaction and Discourse Markers in the ITA-led Physics Laboratory. In G. Gorsuch (Ed.), *Talking Matters: Research on Talk and Communication of International Teaching Assistants* (pp. 75–108). New Forums Press, Stillwater.

Madera, J. M., Neal, J. A., & Dawson, M. (2011). A strategy for diversity training: Focusing on empathy in the workplace. *Journal of Hospitality & Tourism Research*, 35(4), 469-487.

*Mahdi, H. S., & Al Khateeb, A. A. (2019). The effectiveness of computer-assisted pronunciation training: A meta-analysis. *Review of Education*, 7(3), 733-753.

Manohar, U., & Appiah, O. (2016). Perspective taking to improve attitudes towards international teaching assistants: The role of national identification and prior attitudes. *Communication Education*, 65(2), 149-163.

Matsumoto, Y. (2011). Successful ELF communications and implications for ELT: Sequential analysis of ELF pronunciation negotiation strategies. *The Modern Language Journal*, 95(1), 97-114.

Mauranen, A. (2006). Signaling and preventing misunderstanding in English as lingua franca communication. *International Journal of the Sociology of Language*, 177 (1): 123-150

Miles, E., & Crisp, R. J. (2014). A meta-analytic test of the imagined contact hypothesis. *Group Processes & Intergroup Relations*, 17(1), 3-26.

Milroy, J. (2001). Language ideologies and the consequences of standardization. *Journal of sociolinguistics*, 5(4), 530-555.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning*, 45(1), 73-97.

Munro, M.J., & Derwing, T.M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and speech,* 38(3), 289-306.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in second language acquisition*, 28(1), 111-131.

Nelson, C. L. (1995). Intelligibility and world Englishes in the classroom. *World Englishes*, 14(2), 273-279.

Okahana, H. & Zhou, E. (2019). Graudate Enrollment and Degrees: 2008-2018. retrieve online: https://cgsnet.org/ckfinder/userfiles/files/CGS_GED18_Report_web.pdf

O'Neal, G. (2015). Consonant clusters and intelligibility in English as a Lingua Franca in Japan: Phonological modifications to restore intelligibility in ELF. *Pragmatics and Society*, *6*(4), 615-636.

O'Neal, G. (2016). Intelligibility and Segmental Phoneme Repair Strategies in English-as-a-Lingua-Franca Interactions among Chinese and Japanese Speakers of English. *Chinese Journal of Applied Linguistics*, *39*(4), 379-400.

*Parlak, Ö. (2010). Does pronunciation instruction promote intelligibility and comprehensibility? *SPLIS News*, 7 (2), 1-5.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of personality and social psychology*, 90(5), 751-783.

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European journal of social psychology*, 38(6), 922-934.

Pickering, L. (2006). Current research on intelligibility in English as a lingua franca. *Annual Review of Applied Linguistics*, 26 (1), 219-233.

Pickering, L. (2009). Intonation as a pragmatic resource in ELF interaction. *Intercultural Pragmatics*, 6(2), 235-255.

Plakans, B. S. (1997). Undergraduates' experiences with and attitudes toward international teaching assistants. *TESOL quarterly*, 31(1), 95-119.

*Reid, K., Trofimovich, P., & O'Brien, M. (2019). Social attitudes and speech ratings: effects of positive and negative bias on multiage listeners' judgments of second language speech. *Studies in Second Language Acquisition*, 41(2), 419-442.

Reinisch, E., & Weber, A. (2012). Adapting to suprasegmental lexical stress errors in foreign-accented speech. *The Journal of the Acoustical Society of America*, 132(2), 1165-1176.

*Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher education*, 33(4), 511-531.

Rubin, D. L. (2001). Help! My professor (or doctor or boss) doesn't talk English. In J. N. Martin, T. K. Nakayama, & L. A. Flores (Eds.) *Readings in cultural contexts* (pp. 127- 140). Mountain View, Mayfield.

*Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language awareness*, 20(1), 45-59.

Saito, K. (2019). The role of aptitude in second language segmental learning: The case of Japanese learners' English/ɹ/pronunciation attainment in classroom settings. *Applied Psycholinguistics*, 40(1), 183-204.

Saito, K., & Akiyama, Y. (2017). Video-based interaction, negotiation for comprehensibility, and second language speech learning: A longitudinal study. *Language learning*, 67(1), 43-74.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, 69(3), 652-708.

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, 38(4), 439-462.

*Saito, Y., & Saito, K. (2017). Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners. *Language Teaching Research*, 21(5), 589-608.

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1), 187-224.

*Seferoğlu, G. (2005). Improving students' pronunciation through accent reduction software. *British Journal of Educational Technology*, 36(2), 303-316.

Seidlhofer, B. (2009). Common ground and different realities: World Englishes and English as a lingua franca. *World Englishes*, 28(2), 236-245.

Seidlhofer, B., Breiteneder, A., & Pitzl, M. L. (2006). English as a lingua franca in Europe: Challenges for applied linguistics. *Annual review of applied linguistics*, 26 (1), 3-34.

*Sidaras, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125(5), 3306-3316.

Smith, Larry, and Nelson, Cecil (2006). World Englishes and issues of intelligibility. In Braj B. Kachru, Yamuna Kachru, and Cecil L. Nelson (eds.), *The Handbook of World Englishes* (pp. 428–445). Blackwell, Malden.

Southwood, M. H. and Flege, J. E. 1999. Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics*, 13 (5), 335–349.

*Staples, S., Kang, O., & Wittner, E. (2014). Considering interlocutors in university discourse communities: Impacting US undergraduates' perceptions of ITAs through a structured contact program. *English for Specific Purposes,* 35 (1), 54-65.

Subtirelu, N. C. (2017). Students' orientations to communication across linguistic difference with international teaching assistants at an internationalizing university in the United States. *Multilingua,* 36(3), 247-280.

Subtirelu, N.(2014). A language ideological perspective on willingness to communicate. *System* 42 (1), 120–132.

Subtirelu, N. C., & Lindemann, S. (2016). Teaching first language speakers to communicate across linguistic difference: Addressing attitudes, comprehension, and strategies. *Applied Linguistics*, 37(6), 765-783.

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition,* 119(1), 131-136.

*Tanner, M., & Landon, M. (2009). The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning & Technology*, 13(3), 51-65.

Thomson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36(3), 326-344.

*Trofimovich, P., Lightbown, P. M., Halter, R. H., & Song, H. (2009). Comprehension-based practice: The development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition*, 31(4), 609-639.

Watterson, M. (2008). Repair of non-understanding in English in international communication. *World Englishes*, 27(3/4), 378-406.

Villarreal, D. (2013). Closing the Communication Gap between Undergraduates and International Faculty. *CATESOL Journal*, 24(1), 8-28.

Weyant, J. M. (2007). Perspective Taking as a Means of Reducing Negative Stereotyping of Individuals Who Speak English as a Second Language. *Journal of Applied Social Psychology*, 37(4), 703-716.

*Yenkimaleki, M., & van Heuven, V. J. (2019). The relative contribution of computer assisted prosody training vs. instructor-based prosody teaching in developing speaking skills by interpreter trainees: An experimental study. *Speech Communication*, 107(1), 48-57.

*Yeşilçınar, S. (2019). Using the Flipped Classroom to Enhance Adult EFL Learners' Speaking Skills. PASAA: Journal of Language Teaching and Learning in Thailand, 58 (2), 206-234.

## APPENDIX A

UNDERGRADUATES' SPEECH RATING SHEET AND ANSWER SHEET

Speech Rating Sheet

**Please listen to the speech samples once and indicate how much you agree or disagree with the following statements in Section 1, 2, 3 and 4.**

**Section 1:**

**a. This speaker's speech was easy to understand.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**b. This speaker's speech was comprehensible.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**c. This speaker's speech was clear.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**d. This speakers' speech requires little effort to understand.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**e. It was simple to grasp the meaning of this speaker's speech.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**Section 2:**

**The speaker's English proficiency is:**

1) very low

2) low

3) somewhat low

4) moderate

5) somewhat high

6) high

7) very high

**Section 3:**

**a. This speaker speaks with a foreign accent.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**b. This speaker is a non-native speaker of English.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**c. This speaker has a strong accent.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**d. This speaker has an unfamiliar English accent.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**e. This speaker speaks like a non-native speaker of English.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**Section 4**

**a. This speaker is an effective teacher.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**b. This speaker is a qualified teacher.**

1) strongly agree

2) agree

3) somewhat agree

4) neutral

5) somewhat disagree

6) disagree

7) strongly disagree

**Section 5.**

**Please Listen to this speech again and choose the corresponding number to answer the following questions in Section 5.**

**a. How do you rate the speakers' segmental errors?**

1 2 3 4 5 6 7 8 9 10

frequent infrequent

**b. How do you rate the speakers' word stress errors?**

1 2 3 4 5 6 7 8 9 10

frequent infrequent

**c. How do you rate the speakers' intonation errors?**

1 2 3 4 5 6 7 8 9 10

frequent infrequent

**d. How do you rate the speaker's speech rate?**

1 2 3 4 5 6 7 8 9 10

too slow too fast

**e. How do you rate the speakers' inappropriate pauses?**

1 2 3 4 5 6 7 8 9 10

frequent infrequent

**Notes for Section 5:**

- **Segmental error: the incorrect pronunciation of a sound in a word:**

  E.g., error, invitation

- **Word stress error: placing the stress on the wrong syllable within a word**

  E.g., uni-ver-sity vs. uni-ver-sity; ho-tél vs. *hó-tel

- **Intonation error: using the wrong pitch for a sentence.**

  E.g., An expression like "It is sunny outside, isn't it?" Or a single word like "really" can express completely different feelings with rising tone or falling tones.

- **Inappropriate pauses: placing pauses in a wrong place of a sentence or no pause in a sentence.** e.g., Native speakers / of American English / usually speak / like this.

Speech Sample Rating

Answer Sheet

Name:

Date:

Form type:

**Speech Sample 0**

| **Section 1** | **a.** | **b.** | **c.** | **d.** | **e.** |
|---|---|---|---|---|---|
| **Section 2** | | / | / | / | / |
| **Section 3** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 4** | **a.** | **b.** | / | / | / |
| **Section 5** | **a.** | **b.** | **c.** | **d.** | **e.** |

**Speech Sample 1**

| **Section 1** | **a.** | **b.** | **c.** | **d.** | **e.** |
|---|---|---|---|---|---|
| **Section 2** | | / | / | / | / |
| **Section 3** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 4** | **a.** | **b.** | / | / | / |
| **Section 5** | **a.** | **b.** | **c.** | **d.** | **e.** |

**Speech Sample 2**

| **Section 1** | **a.** | **b.** | **c.** | **d.** | **e.** |
|---|---|---|---|---|---|
| **Section 2** | | / | / | / | / |
| **Section 3** | **a.** | **b.** | **c.** | **d.** | **e.** |

| | | | | | |
|---|---|---|---|---|---|
| **Section 4** | **a.** | **b.** | **/** | **/** | **/** |
| **Section 5** | **a.** | **b.** | **c.** | **d.** | **e.** |

**Speech Sample 3**

| | | | | | |
|---|---|---|---|---|---|
| **Section 1** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 2** | | **/** | **/** | **/** | **/** |
| **Section 3** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 4** | **a.** | **b.** | **/** | **/** | **/** |
| **Section 5** | **a.** | **b.** | **c.** | **d.** | **e.** |

**Speech Sample 4**

| | | | | | |
|---|---|---|---|---|---|
| **Section 1** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 2** | | **/** | **/** | **/** | **/** |
| **Section 3** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 4** | **a.** | **b.** | **/** | **/** | **/** |
| **Section 5** | **a.** | **b.** | **c.** | **d.** | **e.** |

**Speech Sample 5**

| | | | | | |
|---|---|---|---|---|---|
| **Section 1** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 2** | | **/** | **/** | **/** | **/** |
| **Section 3** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 4** | **a.** | **b.** | **/** | **/** | **/** |
| **Section 5** | **a.** | **b.** | **c.** | **d.** | **e.** |

**Speech Sample 6**

| | | | | | |
|---|---|---|---|---|---|
| **Section 1** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 2** | | / | / | / | / |
| **Section 3** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 4** | **a.** | **b.** | / | / | / |
| **Section 5** | **a.** | **b.** | **c.** | **d.** | **e.** |

**Speech Sample 7**

| | | | | | |
|---|---|---|---|---|---|
| **Section 1** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 2** | | / | / | / | / |
| **Section 3** | **a.** | **b.** | **c.** | **d.** | **e.** |
| **Section 4** | **a.** | **b.** | / | / | / |
| **Section 5** | **a.** | **b.** | **c.** | **d.** | **e.** |

## APPENDIX B

LINGUISTIC DIVERSITY WORKSHOP

List of activities in the workshop**:**

- Activity 1-Voices and associations

  You'll hear several different people reading aloud the same text. Based on your immediate gut reaction, fill out a survey.

- Activity 2-Movie director challenge

  Imagine that you are directing an animated film with the following characters:

  - a boss/supervisor
  - a villain
  - a hero
  - a dumb person

  Which voice would you cast for each part? Why? (Answer silently, not aloud)

- Activity 3-Record your own voice

  Please read the following sentences loudly and make a recording of your voice while you are reading. Then listen to your own voice.

  *Don fell into the pool at dawn. Ben found a rope in a bin and used it to pull Don out.*

- Activity 4-What's your position

  To what extent do you agree or disagree to the following statement?

  *Language standardization is a good thing.*