L2 Speaking at the University:
Naturalistic Changes, Rater Judgments, and Teacher Training

Larissa Buss

A Thesis
In the Department
of
Education

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Education) at
Concordia University
Montreal, Quebec, Canada

June 2018

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Larissa Buss

Entitled: L2 Speaking at the University: Naturalistic Changes, Rater Judgments, and Teacher Training

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Education)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____	Chair
Dr. Diane Pesco	
_____	External Examiner
Dr. Amanda Baker	
_____	External to Program
Dr. Diane Querrien	
_____	Examiner
Dr. Pavel Trofimovich	
_____	Examiner
Dr. Walcir Cardoso	
_____	Thesis Supervisor
Dr. Sara Kennedy	

Approved by _____
Dr. Richard Schmid, Department Chair

August 6, 2018 _____
Dr. André Roy, Dean, Faculty of Arts and Science

ABSTRACT

L2 Speaking at the University: Naturalistic Changes, Rater Judgments, and Teacher Training

Larissa Buss, Ph.D.
Concordia University, 2018

This dissertation explores issues related to the development and evaluation of second language (L2) oral performance and to teacher training in L2 pronunciation pedagogy. Study 1 investigated naturalistic changes in L2 graduate student presentations given at the beginning and end of the students' first two terms of study at an English-medium university. The presentations were evaluated for accentedness, comprehensibility, fluency, topic structure clarity, and overall quality by native English listeners. Links between the listener evaluations and the speakers' use of paratones (measured as pitch increase at topic shifts) were also explored. The participants became significantly less accented and more comprehensible, but no other changes were found, and pitch increase values were not correlated with any of the listener ratings. Listener comments provided more intricate insights into the role of intonation in perceptions of structural clarity and overall quality of L2 presentations.

Study 2 involved participants from the same pool as those in Study 1, but examined longitudinal changes over four time points. Samples taken from L2 graduate student presentations mostly in engineering and computer science were evaluated for accentedness, comprehensibility, fluency, content, organization, and speaking style by two groups of listeners: content specialists and non-specialists. Overall, no significant changes were found, but an analysis of individual performance revealed that some speakers appeared to improve in fluency and speaking style. The main difference between the listener groups was that specialists valued content and organization more than non-specialists, and their comments tended to be more specific and mention a wider range of assessment criteria.

Study 3 explored the effects of training in pronunciation pedagogy on student teacher cognitions. Pre-service teachers from one university attended a course in phonology and pronunciation teaching, while a comparable group from another university received no specific training in pronunciation teaching. Data were collected through questionnaires and interviews. The treatment participants developed more favorable views of explicit pronunciation teaching and became more confident in their ability to teach pronunciation than the comparison group. The interviews revealed other cognition changes and several positive aspects of the course that influenced cognition development, as well as potential areas for improvement.

Acknowledgements

I would like to express my deep gratitude to my amazing supervisor, Dr. Sara Kennedy, for supporting me and helping to push me forward at every stage of my PhD. Thank you for your always timely and to-the-point feedback and for being so understanding when new motherhood slowed me down in my work (twice!).

I would also like to thank my committee members, Dr. Pavel Trofimovich and Dr. Walcir Cardoso, for their contributions to the studies in this dissertation and for being inspiring role models as researchers and teachers.

Thank you to all the professors, staff members, and peers from the Department of Education who were part of my PhD journey or contributed in some way to my research. Specifically, I wish to acknowledge the assistance given by William Zullo, who helped with part of the data collection for my first study, and Ross Sundberg, who kindly verified some of my data coding.

Finally, I am grateful to my family for encouraging me in my endeavours, especially to my beloved husband, Djonatan Buss, who spent many evenings and weekends caring for our two small children while I finalized this manuscript. I am so thankful that God gave me you to share my life with. I dedicate this dissertation to you and our boys.

This research was supported by a doctoral scholarship from CAPES, Brazil [grant n.0369/12-8], and by a SSHRC grant awarded to Sara Kennedy.

CAPES

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

Contribution of Authors

Studies 1 and 2 of this thesis are coauthored with my supervisor, Dr. Sara Kennedy, who collected the speaker data used in both studies, contributed to the conception of Study 2, and assisted in preliminary statistical analyses of part of the data from Study 2. As first author, I collected the listener data for both studies, conducted the data analyses, and wrote the manuscripts with supervision by Dr. Sara Kennedy. Study 3 is single authored.

Table of Contents

List of Figures

List of Tables

Chapter 1: General Introduction

Introduction

Just over thirty years ago, second language (L2) instructors who wished to teach pronunciation had to rely mostly on their intuitions and observations of student behavior, as resources and directions beyond drills and minimal pair practice were limited (Derwing & Munro, 2015). In the past few years, research on L2 pronunciation teaching and learning has accumulated, and a plethora of instructional materials is currently available to second language teachers. Before anything else, however, teachers need to establish pedagogical priorities. Based on the recommendations in Derwing (2008) and Derwing and Munro (2008), this decision should take into consideration at least three main factors: (1) the students' needs, (2) evidence-based information about the relative importance of different pronunciation features to communication, and (3) evidence-based information about the aspects of L2 speech that tend to develop naturally and those that may require intervention.

Consideration of students' needs may involve determining the pronunciation features or individual words that are most challenging for them, the students' objectives for learning the L2, and the specific real-life situations in which they need to be able to communicate effectively. Pronunciation research can help to determine how students' performance in these specific situations is impacted by non-target-like pronunciation, the features of speech that are most valued by the students' interlocutors, and the aspects that are less prone to naturalistic improvement, thus requiring focused instruction. All of this information is valuable for selecting instructional targets so that teachers can make the most of their limited classroom time. Nonetheless, while pronunciation research has made great strides in recent years, there are still many underexplored issues pertaining to the assessment and naturalistic acquisition of L2 pronunciation. Study 1 and Study 2 of this dissertation address these topics in the context of L2 graduate student academic presentations in engineering and computer science.

Study 1 sought to determine whether L2 graduate students who had just arrived in an English-speaking environment would show naturalistic improvements in their pronunciation, oral presentation skills, and use of intonation to signal topic structure six months after starting their programs. It also sought to explore the impact of the speakers' use of intonation for discourse

structuring on listener assessments of their pronunciation and oral presentations. Study 2 also investigated naturalistic changes in L2 oral performance, but included the issue of subject-matter knowledge and how it may affect listener perceptions. A comparison was made between the ratings and written comments provided by subject-matter specialists in engineering and computer science and non-specialists when evaluating L2 academic presentations on speech- and content-related constructs.

Like much of the research in the field of L2 pronunciation, Studies 1 and 2 were conducted in the hopes of ultimately helping English as a Second Language (ESL) instructors to make informed decisions about what to focus on when teaching pronunciation. However, these findings are to no avail if ESL teachers are unwilling to address pronunciation in their classrooms. Previous studies conducted in different countries have concluded that ESL teachers may be reluctant to teach pronunciation because they lack confidence and/or would like more training in the field (Baker, 2011; Buss, 2016; Foote, Holtby, & Derwing, 2011; Macdonald, 2002). In Study 3, I turn from acquisition to teacher training by looking at the effects of training in pronunciation pedagogy on pre-service teacher cognitions (TC). Teaching English as a Second Language (TESL) students from one university attended a course in phonology and pronunciation teaching, while comparable students from another university received no specific training in pronunciation teaching. The objective was to determine whether training would help the treatment students develop cognitions that promote more pronunciation instruction in their present or future ESL teaching.

The general literature review that follows will address important issues in L2 pronunciation which relate to the three studies. It begins by defining the key concepts of accentedness, intelligibility, comprehensibility, and fluency. Except for intelligibility, these dimensions of pronunciation are evaluated in Studies 1 and 2 through listener ratings. Intelligibility is defined due to its importance in the field of L2 pronunciation, and because of the need to distinguish it from the other dimensions. Two constructs used for the evaluation of the L2 oral presentations in Studies 1 and 2 are also explained: speaking style and topic structure clarity. I then discuss evidence related to listener evaluations and naturalistic acquisition of L2 pronunciation and how it can help to inform pronunciation teaching. The chapter concludes with a discussion of evidence-based teacher training and TC related to L2 pronunciation.

General Review of the Literature

Dimensions of L2 Pronunciation and Oral Presentations

Accentedness, intelligibility, comprehensibility, and fluency. According to Derwing and Munro (2015, p. 2), *pronunciation* can be defined as "the ways in which speakers use their articulatory apparatus to create speech". When assessing L2 pronunciation, there are different dimensions that can be focused on. The current literature distinguishes between four basic dimensions: accentedness, intelligibility, comprehensibility, and fluency. *Accentedness* is the "listener's perception of how different the speaker's accent is from that of the L1 [first language] community" (Derwing & Munro, 2005, p. 385). A foreign accent usually results from the interference of some phonological characteristics of the speaker's L1 with their L2, and, while its degree can be influenced by a variety of factors, its strongest predictor is age of L2 learning (Piske, MacKay, & Flege, 2001). L2 speakers who started learning the language as adults almost invariably have some trace of a foreign accent. However, accentedness does not necessarily interfere with the speaker's ability to communicate in the L2. The most important dimensions for communication are instead intelligibility and comprehensibility.

Intelligibility is defined as "the extent to which a listener actually understands an utterance", while *comprehensibility* relates to the "listener's perception of how difficult it is to understand an utterance" (Derwing & Munro, 2005, p. 385). It is important to distinguish between these two dimensions because listeners might rate an utterance as highly comprehensible believing that they understood it correctly when they actually did not (Munro, 2008). More commonly, however, listeners will understand an utterance perfectly but rate it as less comprehensible because understanding it required additional effort or longer processing time. Thus, comprehensibility ratings are usually harsher than intelligibility scores (Derwing & Munro, 1997). Although it may be more objective to evaluate how much comprehension has actually taken place, comprehensibility ratings can give a general idea of a listener's ability to understand what is said. They are a practical and common way of evaluating "broadly-defined intelligibility" (Trofimovich & Isaacs, 2012, p. 906).

It is generally believed that a strong foreign accent automatically reduces intelligibility and comprehensibility, but research suggests that this is not always the case. In fact, these

notions are relatively independent. According to Derwing and Munro (2005), even though listeners tend to consider unintelligible or incomprehensible utterances as heavily accented, the opposite does not necessarily happen. This means that a speech sample may be considered heavily accented but highly comprehensible at the same time. Likewise, accented speech samples may prove entirely intelligible to native listeners (Derwing & Munro, 1997; Munro & Derwing, 1995).

Finally, *fluency* is related to the rate and fluidity of speech. Fluent speech flows easily without many pauses, hesitations, filled and unfilled pauses, self-repetitions, and other dysfluency markers (Derwing & Munro, 2015). Although fluency in an L2 tends to increase with overall language proficiency, even advanced L2 speech and L1 speech can vary in terms of fluency, based on individual speaker traits and on situational factors, such as nervousness, stress, fatigue, etc. According to Lennon (1990), fluency differs from other aspects of oral proficiency because it is "purely a performance phenomenon", and the problem with dysfluent speech is that it shifts the listener's attention from the message to the process of speech planning and production:

> fluency is an impression on the listener's part that the psycholinguistic processes of speech planning and speech production are functioning easily and efficiently. Dysfluency markers, as it were, make the listener aware of the production process under strain. To some extent, then, fluency reflects the speaker's ability to focus the listener's attention on his or her message by presenting a finished product rather than inviting the listener to focus on the working of the production mechanisms (p. 391-392).

In Studies 1 and 2, the pronunciation of L2 graduate students was assessed through listener ratings of accentedness, comprehensibility, and fluency. Having discussed these basic dimensions of speech commonly found in the literature, I will now turn to two constructs devised for evaluating L2 oral presentations in Studies 1 and 2.

Topic structure clarity and speaking style. The *topic structure* of a presentation, lecture, or any other form of monologue is the organization of the discourse into different topics. A discourse topic can be considered "an aggregate of coherently related events, states, and referents" (Chafe, 1994, p. 121). Listeners and readers seem to have an intuitive notion of "what is being talked about" in a given stretch of discourse, and discourse analysts have appealed to

this notion many times in order to segment spoken and written discourse (G. Brown & Yule, 1983, p. 71). However, as mentioned by G. Brown and Yule (1983), the notion of "topic" is difficult to pin down, and sometimes it can be hard to determine discourse topic without access to the speaker's intentions. According to the authors, "it is speakers and writers who have topics, not texts" (p. 68). Therefore, in order to identify the beginnings and endings of large portions of discourse, one can resort to topic boundary markers; that is, elements used by speakers to signal *topic shifts*.

Topic boundary markers include lexical discourse markers such as "To begin with", "The next thing was" (Chaudron & Richards, 1986) and intonation, which G. Brown, Currie, and Kenworthy (1980, p. 26) claim to be "the strongest indicator that the speaker is changing direction within the overall topic area". Topic shifts in L1 English are characterized by a lowering of the pitch level at the end of one topic, often followed by a long pause, and a jump to extra high pitch at the beginning of a new topic (Tench, 1996; S. Thompson, 2003; Yule, 1980). These cues organize spoken speech into what have been called *paratones* or *intonational paragraphs*: the spoken equivalent of paragraphs (G. Brown, 1977). Previous studies suggest that some L2 speakers use fewer discourse structuring devices than L1 speakers in academic presentations and lectures, which may render their speech less effective (Tyler, 1992; Tyler, Jefferies, & Davies, 1988). The construct *topic structure clarity* (also referred to as simply *structure clarity*) was used in Study 1 in order to assess how effectively L2 graduate students giving oral presentation were able to signal topic shifts, thus conveying the macro-organization of their discourse. Topic structure clarity was evaluated by listeners through scalar ratings and defined to them as:

> The degree to which a speaker's presentation has a clear topic structure. If it is unclear or hard to identify when one topic ends and another one starts, then the presentation has low structure clarity (i.e., the structure is very unclear). If it is clear to the listener when one topic ends and another one starts, then it has high structure clarity (i.e., the structure is very clear).

Although no mention of intonation is found in the definition, the *structure clarity* ratings were intended as an impressionistic measure related to paratones, in order to explore whether more native-like use of intonation at topic shifts would increase their salience to the listener. The

second construct devised for this dissertation, used in Study 2, is *speaking style*.[1] This construct was also evaluated through scalar ratings and was explained to the listeners as:

> The degree to which the speaker uses a dynamic speaking voice. If the speaker uses a range of intonation (changes in pitch) which makes his/her speech attractive, then the speech is highly dynamic. If the speaker changes his/her intonation very little, speaking in a monotone, then the speech is monotonous.

Public speaking manuals often refer to the need for speakers to vary their pitch in order to convey vitality and engage listeners (e.g., Atkinson, 2005; Lucas, 2009). Flat, monotone speech is considered boring and may cause the listener's attention to wander away from the speaker's message. A study by Hincks (2005) found evidence of the relationship between pitch variations and listener perceptions of liveliness. She derived a metric called PVQ (pitch variation quotient) from the standard deviation of the mean fundamental frequency (F0) measured in short samples of L2 undergraduate student presentations. A selection of these samples was presented to eight university teachers of English, who rated them using an undivided scale from "lively" to "monotone". As predicted, there was a significant positive correlation between PVQ and liveliness ratings. The construct *speaking style* is related to liveliness and is featured in Study 2 to assess potential changes to this facet of the speakers' use of pitch over time, as well as explore its importance to subject-matter specialists and non-specialists.

What Underlies Evaluations of L2 Speech

The influence of paratones. The dimensions discussed above were measured in Studies 1 and 2 using scalar ratings provided by native English-speaking listeners. Scalar ratings are very common assessment tools in L2 pronunciation research. However, they do not tell us much about how raters interpret the dimensions assessed and about the aspects of speech they focus on when making their judgments (Isaacs & Thomson, 2013). Several studies have addressed this issue by investigating how specific linguistic aspects feed into holistic ratings, including suprasegmental features (e.g., Anderson-Hsieh, Johnson, & Koehler, 1992; Derwing, Rossiter, Munro, &

[1] This construct is not derived from the sociolinguistic construct of *style*.

Thomson, 2004; Hahn, 2004; Munro & Derwing, 1998). However, few of them have directly investigated paratones.

In Trofimovich and Isaacs (2012) and Isaacs and Trofimovich (2012), pitch range was investigated as a measure that was "influenced by Wennerstrom's (2001) notion of paratones" (Isaacs & Trofimovich, 2012, p. 483). The premise was that a narrower pitch range, measured as the difference between the highest and lowest F0, would indicate fewer paratones in the stories (Isaacs & Trofimovich, 2012) and a wider pitch range would describe animated voices and indicate use of changes in pitch to convey meaning and emphasis (Trofimovich & Isaacs, 2012). The speech samples analyzed in the two studies were produced from 40 French speakers of L2 English and had been elicited through a picture-based narrative task. They were analyzed for 19 different pronunciation, lexical, grammatical, and discourse-level variables, and the measures were correlated with scalar judgments of accentedness and comprehensibility provided by 60 native English listeners who were undergraduate students in non-linguistic programs. In order to further explore the aspects underlying accentedness and comprehensibility ratings, these studies included introspective reports in which three experienced ESL teachers discussed the rationale for their evaluations. The findings revealed that pitch range was not a significant aspect for either of the listener ratings.

Kang (2010) also investigated the relative weight of pitch range and other suprasegmental measures for accentedness and comprehensibility ratings, but her speech samples consisted of five-minute lectures given by 11 international teaching assistants (ITAs). The ratings were carried out by 58 undergraduate students who were native speakers of English. According to the results of her regression analysis, pitch range was the best predictor of accentedness ratings, accounting for 24% of the variance, but did not significantly contribute to predicting the variance in comprehensibility ratings. A follow-up study, Kang (2012) found that pitch range was also related to oral proficiency ratings and teaching competence ratings provided by 70 undergraduate students for the same speech samples in Kang (2010).

Turning to studies that examined paratones more directly, Wennerstrom (1998) conducted acoustic analyses of several aspects of discourse intonation in the speech of 18 L2 graduate students who were native speakers of Mandarin. The speakers were recorded giving a 10-12-minute lecture in their fields and their production was evaluated by three raters based

primarily on comprehensibility and presence of "errors". These scores were used in a multiple regression analysis with the intonation variables as predictors. The paratone variable, measured as a percentage of increase in pitch range from the end of one topic to the beginning of the next topic, was the only significant predictor of the production scores.

Kang, Rubin, and Pickering (2010) explored the relative importance of 29 suprasegmental measures to listener ratings of oral proficiency and comprehensibility in 26 speech samples taken from iBT TOEFL examinations. The listeners were 188 native English-speaking undergraduate students. Four of the suprasegmental measures were related to the paratone: (1) number of low termination tones, which consisted of low terminations followed by high-key resets, (2) average height of onset pitch, (3) average height of termination pitch, and (4) average paratone pause length. A hierarchical cluster analysis followed by a multiple regression revealed that the cluster called *boundary marking*, which included number of silent pauses and low termination tones, had positive relations to the comprehensibility and oral proficiency ratings. Conversely, the *pitch height* cluster, a group of pitch variation parameters including the average height of paratone onsets and terminations, was negatively related to the listener ratings.

As can be seen, the findings regarding the relevance of pitch range in general and paratones in particular to listener evaluations of L2 pronunciation are inconsistent, and further evidence is needed before these aspects of speech can be suggested as important pedagogical targets for ESL and English for Academic Purposes (EAP) courses. Another issue that is particularly relevant for EAP instruction and remains underexplored is the role of subject-matter knowledge in listeners' perceptions and evaluations of L2 speech.

The influence of listener-related variables. Previous research has investigated the influence of some listener-related variables on listener perceptions of L2 speech. The most researched one is probably familiarity. Gass and Varonis (1984) investigated the effects of four types of familiarity on listeners' ability to transcribe accented utterances: familiarity with the topic, familiarity with nonnative speech in general, familiarity with a particular nonnative accent, and familiarity with a particular speaker. While all of these factors had a facilitating effect on intelligibility, the most important one was familiarity with topic. Studies on the effects of accent familiarity have yielded conflicting results. Carey, Mannell, and Dunn (2011) found that 99 International English Language Testing System (IELTS) examiners gave higher ratings to L2

pronunciations to which they had prolonged exposure. In addition, a significant proportion of nonnative-speaker (NNS) raters were more favorable toward candidates from their home countries. While further studies have corroborated these findings (e.g., Kraut & Wulff, 2013), others, such as Munro, Derwing, and Morton (2006) have not found strong effects of foreign accent familiarity or have found mixed results (e.g., Kennedy & Trofimovich, 2008). Related to this discussion is the issue of nonnative-speaker listeners and whether/how their perceptions may differ from those of native-speaker listeners. This emerging area of research is outside the scope of this dissertation, and in order to ensure a baseline of listening comprehension, all of the listeners recruited in Studies 1 and 2 were native English speakers.

Another listener-related factor that has yielded varied findings is linguistic training. I. Thompson (1991), for example, submitted samples of Russian-accented English to two groups of listeners: language experts and untrained listeners. The listeners with linguistic training were found to be more lenient and reliable in their ratings of accentedness than were untrained listeners. Similarly, Saito, Trofimovich, and Isaacs (2017), which used data from Trofimovich and Isaacs (2012) and Isaacs and Trofimovich (2012), found that raters with linguistic and pedagogical experience provided more positive judgments of accentedness and comprehensibility than inexperienced raters, and they were also more consistent in evaluating some linguistic variables. Okamura (1995), on the other hand, found that teachers of Japanese as a foreign language were more critical than non-teachers when evaluating Japanese learners' grammar, fluency, appropriateness, vocabulary, comprehensibility, and pronunciation. Finally, Isaacs and Thomson (2013) did not observe significant differences between comprehensibility, accentedness, and fluency ratings given by ESL teachers with postsecondary training and by novice raters, although the former were slightly more consistent and, according to verbal reports and interviews, arrived at their ratings in different ways than the latter.

In the studies above and others in the literature, "expert" or "trained" judges refer to listeners with training in linguistics or language teaching. However, the role of other forms of training on ratings of L2 speech is still unclear. Specifically, it is unknown whether the perceptions of subject-matter specialists differ significantly from those of non-specialists when L2 speech is produced in the context of academic or professional tasks, such as a presentation or lecture given by an international graduate student. Also, little is known about the aspects that

specialists pay most attention to when listening to L2 speech in these contexts. Prior studies suggest that subject-matter specialists are more concerned with the accomplishment of communicative goals than with the quality of language *per se* (Elder, 1993; Purpura, 2016). However, these studies tended to either provide specialists with inflexible quantitative instruments (e.g., Brown, 1995) or to ask them to give general feedback without much guidance (e.g., Jacoby, 1998). No known studies have prompted subject-matter specialists to rate different dimensions of pronunciation (accentedness, comprehensibility, etc.) while also providing comments to explain their evaluations. This information is important because EAP is often taught by teachers who have little or no expertise in their students' fields, and these teachers' perceptions may not coincide with those of a specialized audience. Thus, for teachers to be able to address their L2 graduate students' real needs, they should be informed of potential differences in how subject-matter specialists perceive L2 speech and oral presentations compared to non-specialists and of the aspects that are most valued by the specialists.

In addition to addressing the contribution of paratones and listener subject-knowledge to pronunciation ratings, Studies 1 and 2 also include a time dimension, whereby longitudinal changes are examined in the oral performance of L2 graduate students in their first two terms at an English-medium university. Indeed, knowing the aspects of speech that students most struggle with and those that are relevant to their specific communication needs is still not enough to justify the selection of targets for instruction. It is also important to identify the aspects that are most likely to develop naturalistically and how much development can take place without explicit instruction, so that the teacher can focus on the areas that require intervention. The next section discusses research findings on naturalistic development of L2 speech.

Naturalistic Development of L2 Oral Performance

It is well-documented fact that many ESL speakers continue to have pronunciation problems even after years of residence in an English-speaking country. For these "fossilized" learners, explicit instruction has proven effective in promoting positive changes in pronunciation (Derwing, Munro, Foote, Waugh, & Fleming, 2014; Derwing, Munro, & Wiebe, 1997). However, it has been suggested that pronunciation fossilization only takes place after an initial period of rapid incidental learning, usually during the speakers' first year of immersion (Flege,

1988). This incidental learning, which results from a sudden increase in L2 exposure and practice, has not been fully researched. It is unknown, for example, whether some dimensions of pronunciation (accentedness, comprehensibility, intelligibility, etc.) or some features of speech (segments, stress, intonation, etc.) are more likely to develop naturalistically than others. Such knowledge is relevant because it may allow ESL teachers working with recent immigrants to invest more of their limited class time on aspects that are harder to learn incidentally.

The evidence so far suggests that the oral skills of adult L2 speakers can indeed develop naturalistically within the first year of residence in an L2 environment (Flege, 1988) and even after the first year, although this development may be dependent on a number of variables, such as amount of native-speaker input (Flege & Liu, 2001), age of arrival (W. Baker, 2010; Trofimovich & Baker, 2006), L1, and willingness to communicate (Derwing & Munro, 2013; Derwing, Munro, & Thomson, 2007). The longitudinal studies reported in Derwing, Munro, and Thomson (2007) and Derwing and Munro (2013) investigated two groups of immigrants to Canada: 11 Mandarin speakers and 11 Slavic speakers. These immigrants received two years of basic ESL classes as part of the Language Instruction for Newcomers to Canada (LINC) program and were evaluated at different points in time by untrained listeners. By the end of the second year, the Slavic language speakers were perceived as significantly more fluent, more comprehensible, and less accented, whereas the Mandarin speakers' performance did not change significantly. After five years without further ESL instruction and a total of seven years living and working in Canada, the Mandarin group continued to show little or no improvement. The Slavic group, on the other hand, was once again perceived as more comprehensible and fluent, although their accent scores did not change.

In another study featuring L2 speakers from the same pool of participants used in Derwing et al. (2007) and Derwing and Munro (2013), Munro and Derwing (2008) examined longitudinal changes in the English vowel production of 44 L1 speakers of Mandarin and Slavic languages. Data was collected at two-month intervals over their first year of residence in Canada and was assessed for intelligibility by four phonetically trained native-speaking listeners. Both Mandarin and Slavic language speakers were found to improve in vowel intelligibility, but the largest changes occurred in the first half of the year. A follow-up study by Munro, Derwing, and Saito (2013) included an analysis of accuracy through acoustic measurements in addition to an

intelligibility assessment to explore vowel changes in the speech of 13 Mandarin and 18 Slavic language speakers between Years 1 and 7 of residence. The analyses revealed that some improvement did seem to have occurred after the first year, but it was limited and not statistically significant.

The studies above lend support to the claim that pronunciation will change more quickly within the first year of immersion (Flege, 1988), although this does not preclude slower changes in subsequent years and does not guarantee that all speakers will show improvement, since individual differences may also play a role in the process. It also appears that accentedness may be more resistant to change without instruction than the other dimensions, which may be partly due to that fact that accent has less communicative value (Derwing & Munro, 2013), since a speaker can be perfectly intelligible and comprehensible in spite of a strong accent (Munro & Derwing, 1995). Further evidence is needed in order to confirm these findings. Notably, although the participants in the studies above did not receive explicit pronunciation training, they were enrolled in full-time intensive ESL classes for two years, which may have given them more opportunities for L2 input, output, and noticing (Schmidt, 1990) than what is typically experienced by other recent arrivals, such as international graduate students. Furthermore, L2 graduate students have higher proficiency than the participants in the studies above (who were all beginners), given that they have met the minimum scores in standardized English tests required by their universities. This begs the question of whether these learners can improve their speech naturalistically, and if so, which aspects exhibit the most improvement.

Lastly, although there has been some research on the development of oral academic discourse, most qualitative studies have focused on academic discourse socialization through engagement in oral presentations (e.g., Morita, 2000; Zappa-Hollman, 2007). These studies examine how L2 graduate students are socialized into underlying values of academic discourse as the students prepare and perform oral presentations; the studies also explore the principal difficulties students face in the socialization process. Little is known, however, about specific changes in the nature of L2 graduate student presentations. It is possible that some aspects of L2 graduate students' oral presentations may improve without instruction, simply by their taking part in academic activities in an L2 environment, while other aspects may require direct teaching.

Therefore, comparisons of L2 oral performance over time may help to inform the selection of targets for ESL and EAP programs aimed at L2 graduate students.

Teacher Training in L2 Pronunciation

The sections above discussed issues that are relevant to evidence-based teaching of L2 pronunciation, particularly when it comes to selecting instructional targets for teaching L2 graduate students. Research on these issues contributes with evidence that can be useful not only to ESL instructors, but also to those involved in teacher training. However, in addition to evidence-based information on L2 pronunciation teaching itself, teacher trainers need evidence-based guidelines for training pre-service teachers in pronunciation pedagogy. They must know, for example, the needs of teachers in training and the training methods that are most likely to produce positive outcomes, which include the development of empirically-grounded cognitions related to pronunciation, willingness to teach pronunciation, and effective teaching skills in pronunciation.

Teacher cognition is defined by Borg (2003, p. 81) as the "cognitive dimension of teaching – what teachers know, believe, and think" and is a main influence on teaching practices (e.g., Ghaith & Yaghi, 1997; Smylie, 1988). Very few studies have explored the effect of training on TC and teaching practices related to L2 pronunciation. Baker (2011a, 2014) found preliminary evidence that teacher training in pronunciation pedagogy can have positive effects. In her first study (Baker, 2011a), inteviews and classroom observations indicated that, for three of the five teacher participants, the greatest influence in their teaching of pronunciation had been a graduate-level course in pronunciation pedagogy. The other two teachers had not received the same level of training, and one of them, who had little or no training in pronunciation pedagogy, reported lack of confidence to teach pronunciation. In Baker (2014), the same three teachers who had graduate-level training were found to use a larger range of pronunciation teaching techniques than the other two, although most of them were controlled rather than communicative techniques.

A longitudinal case study reported by Burri (2015a, 2015b) and Burri, Baker, and Chen (2017) explored the cognition development of 15 ESL student teachers during a graduate course in pronunciation pedagogy at an Australian university. Findings from interviews, focus groups,

questionnaires, and classroom observations indicated that the training helped to increase the students' awareness of the importance of suprasegmentals and of nonnative-speakers' ability to teach pronunciation (Burri, 2015b). Nonnative speakers seemed to have benefited from increased awareness of their own speech and from the perception that their pronunciation improved with the course, while native speakers benefited from their interactions with nonnative peers. The participants also came to believe that the goal of pronunciation teaching should not be accent elimination (Burri, 2015a) and became more aware of the benefits of kinesthetic/tactile teaching techniques, although they still appeared to favor controlled activities, which were the kind they had experienced as L2 learners (Burri et al., 2017).

More research is warranted to explore the effects of training in L2 pronunciation pedagogy in other contexts. Several of the participants in Burri's research had years of teaching experience and some had previous training in TESL or related areas. Study 3 sought to investigate whether similar outcomes would be observed for pre-service teachers in undergraduate programs who have little or no teaching experience. It is also sought to determine how training in pronunciation pedagogy uniquely contributed to cognition development in the field by factoring out potential influences from other training in TESL. Evidence from this line of research can help to inform those involved in curriculum and course design in L2 pronunciation pedagogy, who ultimately will be faced with the same challenge referred to at the beginning of this chapter of selecting instructional targets that will be most beneficial to their students (in this case, ESL teachers).

Introduction to Study 1

The first study in this dissertation had two main objectives: (1) to explore naturalistic changes in the pronunciation and oral presentation skills of L2 graduate students giving academic presentations in English, including their use of paratones, and (2) to investigate how changes in pitch to signal topic structure relate to ratings of accentedness, comprehensibility, fluency, topic structure clarity, and overall quality given by native-speaker listeners with a background in TESL or applied linguistics. The speakers were all recent arrivals to a multicultural and effectively bilingual (French-English) city in Canada and were assessed at the beginning and end of their first six months of study in an English-medium university, within the time frame that is hypothesized to be a "window of maximal opportunity" for adults to improve L2 pronunciation without instruction (Derwing & Munro, 2015, p. 43; Flege, 1988). The study aimed to determine whether L2 graduate students who are already proficient in English and may have limited native-speaker input can benefit from this window and which dimensions of speech are more prone to incidental improvement. Finally, relations were explored between the speakers' use of intonation to signal the topic structure of their presentations and listener evaluations of these presentations. The goal here was to confirm or disconfirm previous claims and evidence that paratones may be important to the effectiveness of L2 presentations and lectures (e.g., Pickering, 2004; Tyler et al., 1988; Wennerstrom, 1998).

Chapter 2: Study 1

Naturalistic Changes and Discourse Structuring in L2 Graduate Student Presentations

Abstract

Given the importance of L2 speaking and academic presentations to international graduate students in English-speaking universities (Berman & Cheng, 2010), the present study sought to determine whether 10 L2 graduate students beginning their studies at an English-medium university would show naturalistic improvements in their speech, oral presentation skills, and use of paratones as discourse structuring devices six months after starting their programs. It also sought to explore potential links between the speakers' use of paratones and listener evaluations of their L2 speech and oral presentations. The method involved collecting short presentations given by the students at the beginning and at the end of the time interval and having them evaluated by 11 native speakers for accentedness, fluency, comprehensibility, topic structure clarity (i.e., how easily the listener could identify a change of topic), and overall quality. The presentations were coded for discourse topics, and the intonation cues associated with paratones were acoustically measured and averaged. Time 1 and Time 2 data were statistically compared to determine any changes, and paratone values were correlated with the listener ratings to explore potential relationships. Written comments provided by the listeners were used to corroborate and make sense of the quantitative findings. The findings revealed that the participants became significantly less accented and more comprehensible at Time 2, despite having had no instruction in speaking and possibly limited native-speaker input. However, no significant changes were found in fluency, topic structure clarity, overall quality, and use of paratones, and the latter was not linearly correlated with any of the listener ratings. A combined interpretation of the quantitative and qualitative data and an analysis of individual cases provided more intricate insights into the role of paratones in perceptions of structural clarity and overall quality of L2 oral presentations.

Introduction

Graduate students in different fields are expected to perform a number of oral tasks, which include interacting with professors and peers, taking part in class discussions, giving presentations in class and at conferences, and sometimes lecturing as teaching assistants or instructors. Therefore, for the growing number of international graduate students in English-speaking universities, oral proficiency in the L2 is essential. A survey conducted in Canada by Berman and Cheng (2010) found that nonnative English-speaking (NNS) graduate students considered speaking to be the most difficult skill and carrying out oral presentations to be the most challenging task they were expected to perform at university. Furthermore, the survey found evidence that these language difficulties affected the students' academic achievement compared to their native-speaking (NS) peers, while no such evidence was found for L2 speakers in undergraduate programs.

In fact, previous studies have reported several issues in the academic discourse of L2 graduate students which may lead to lower comprehensibility and negative evaluations from listeners, including problems with fluency and discourse structuring (Pickering, 2001, 2004; Rounds, 1987; Tyler et al., 1988; Wennerstrom, 1998). It is not clear, however, whether and to what extent these students can improve their oral performance without receiving focused instruction during the course of their studies. The present study had two main objectives. The first one was to investigate naturalistic changes in the pronunciation and oral presentation skills of 10 L2 graduate students at an English-medium university located in an effectively bilingual (French-English) Canadian city. The students gave oral presentations at the beginning and at the end of their first 6 months of study and the presentations were evaluated by native-speaker listeners for accentedness, fluency, comprehensibility, topic structure clarity (i.e., how easily the listener could identify a change of topic), and overall quality. The second main objective of this study was to explore potential relationships between listener evaluations and the speakers' use of paratones (i.e., use of pitch to signal new topics). Therefore, the intonation cues associated with paratones were acoustically measured in all of the presentations.

Paratones in L1 and L2 Discourse

One of the reasons why oral presentations can be particularly challenging to L2 graduate students is because, in addition to entailing the normal cognitive demands of speaking in an L2, they require speakers to create a coherent discourse that can be easily processed by the listener in real time, without the possibility of relying on the contextual features, turn-taking, and negotiation of meaning that are present in casual conversation (S. Thompson, 1994). Cohesion in L1 English spoken discourse can be achieved through a number of devices, such as clause relations, lexical discourse markers, patterns of repetition, and prosody. The role of intonation in signaling to the listener the relative importance of discourse segments and the interrelationships between them has been well documented. Pitch height can be used, for example, to distinguish between "new" and "given" information and to convey emphasis or contrast (Chun, 2002). At boundaries, it can indicate a connection between two utterances of the discourse or express finality (Wennerstrom, 2001). Yet another key function of intonation in monologue is to signal the topic structure or macro-organization of the discourse. English speakers use prosodic cues to segment their spoken discourse into units that are analogous to paragraphs in writing, each corresponding to a distinct discourse topic. These units have been called paratones (G. Brown, 1977), phonological paragraphs (Tench, 1996; S. Thompson, 2003), sequence chains (Barr, 1990), and intonational paragraphs (Levis & Pickering, 2004; Pickering, 2004). The phonetic cues associated with a new paratone are a high peak on the first prominent syllable and an expanded pitch range (Pierrehumbert & Hirschberg, 1990), followed by "a series of lesser peaks" (Yule, 1980, p. 36) and a narrowing of the pitch range. At the end of a topic or paratone, there is a lowering of the pitch level and often a slowing down, followed by a lengthy pause (Tench, 1996; Yule, 1980). The last prominent syllable in a paratone is characterized by extra low pitch (S. Thompson, 1994).

Several studies have analyzed the use of paratones to signal topic structure in L1 English, but these studies vary in terms of the frameworks, labels, and measurements used. S. Thompson (2003) analyzed phonological paragraphs in authentic and pedagogically prepared academic lectures in English. These paragraphs were characterized by "low pitch on the final tonic syllable of the paragraph followed by a jump up to high pitch on the first prominent syllable of the new paragraph" (p. 9). Pitch tended to be exceptionally high at the beginning of a new paragraph

compared to other places, and the end of a paragraph was often characterized by extra low pitch, decreased volume or speed, laryngealization (i.e., creaky voice), and/or a long pause.

Grosz and Hirschberg (1992) provide further evidence that these intonational features are in fact associated with discourse structure. In their study, discourse segments were defined as groups of utterances that contribute to the same underlying purpose or intention of the speaker/writer, according to Grosz and Sidner's (1986) theory of discourse structure. A group of subjects were trained to label the discourse features of transcribed news stories based on this theory, either from text alone or from text and speech. There was considerable agreement among labelers regarding segment beginnings and segment endings. Furthermore, these discourse structural elements were found to significantly correlate with intonational features identified by an acoustic-prosodic analysis. Phrases labeled as initiating discourse segments were produced in a larger pitch range and followed by shorter pauses than other phrases, while phrases ending segments were followed by longer pauses.

An early study by Lehiste (1975) sought to determine whether listeners pick up on these intonational cues to identify the discursive context of an utterance. She had a native English speaker record six paragraphs of text that consisted of different orderings of the same three sentences. The sentences were then isolated and randomized, and native listeners were asked to determine whether they had been produced in isolation, at the beginning of a paragraph, in the middle of a paragraph, or at the end of a paragraph. The listeners were able to consistently identify sentences produced with high F_0 peaks as being paragraph initial, suggesting that intonation was in fact used as a cue to context.

While there seems to be consistency in how L1 English speakers produce and interpret the prosodic cues associated with paratones, this is not true for L2 speakers. Wennerstrom (1994) analyzed how 10 native speakers (NSs) and 30 nonnative speakers (NNSs) of English from three language groups used intonation to signal meaning and structure their discourse. The speakers were recorded reading a two-paragraph text with several intonational contrasts. For the analysis of paratones, the researcher compared the pitch ranges of two syntactically and lexically parallel sentences produced in different positions: paragraph medial and paragraph initial. The NSs and the Spanish speakers used a significantly higher pitch range in the paragraph-initial sentence than

in the paragraph-medial one. For the Thai and Japanese speakers, however, no significant differences were found between the two sentences.

Pickering (2004) analyzed intonational paragraphs in the teaching discourse of American and Chinese teaching assistants (TAs) at a North American university. She found that the NSs used pitch consistently to structure their speech into coherent sections. Not only were the boundaries of these sections signalled by prosodic cues, but also by co-occurring lexical and topic-related cues. The analysis of the NNS data revealed a compression of overall pitch range and, as a result, less pronounced pitch peaks than in the NS data. Furthermore, the prosodic cues that the L2 speakers produced often did not coincide with cues at other levels of the discourse and could not be used reliably by listeners as an organization cue.

Wennerstrom (1998) provided preliminary evidence that weaker command of paratones may negatively affect listener ratings of L2 speech. She analyzed several components of discourse intonation in the speech of 18 L1 speakers of Mandarin lecturing in English. All of the participants were attending graduate programs and were enrolled in a 10-week course ESL focused on developing their pronunciation, teaching, and presentation skills, including explicit instruction in intonation. As a final exam for this course, they were required to deliver short lectures that were scored by three raters on four categories: production, organization, classroom management, and a holistic score. The ratings for production, which were primarily based on level of comprehensibility, were used in a multiple regression analysis with the intonation variables as predictors. The result was significant for the paratone variable, indicating that the more a speaker increased their pitch range to mark a topic shift, the better the score they received for production.

Kang et al. (2010) explored relations between 29 suprasegmental measures in L2 speech and listener ratings of oral proficiency and comprehensibility. There were four paratone measures: number of low termination tones (a sum of all low terminations followed by high-key resets), average height of onset pitch, average height of termination pitch, and average paratone pause length. The authors conducted a hierarchical cluster analysis, which reduced the number of variables by grouping them into clusters, followed by a multiple regression. The findings revealed that the *boundary marking* cluster, which consisted of number of silent pauses and low termination tones, had positive relations to the comprehensibility and oral proficiency ratings. On

the other hand, the *pitch height* cluster, which included average height of paratone onsets and terminations, among other pitch variation parameters, was inversely related to the ratings. The authors explained this result by interpreting the cluster as referring to voice pitch (i.e., vocal "thinness") rather than pitch range *per se* or discourse intonation.

More research is warranted to further explore the impact that more or less target-like use of intonation to signal topic structure have on those who listen to nonnative speech. One of the research questions in this study asks about how use of paratones is related to listener evaluations of different speech dimensions (accentedness, comprehensibility, and fluency), topic structure clarity, and overall quality when listening to L2 graduate student presentations.

Naturalistic Improvement in L2 Graduate Students' Oral Performance

In the study by Wennerstrom (1998) reported above, it was observed that L2 graduate students from the same L1 background at about the same level of proficiency differed from each other in their mastery of discourse intonation, which was interpreted as evidence of interlanguage development in progress. It may well be the case, considering that the participants in her study had received focused instruction in intonation. However, another, perhaps more basic question that still has not been fully answered by research is whether and to what extent L2 graduate students in an English-medium university can improve their pronunciation in general, and particularly their discourse intonation, without focused instruction.

Flege (1988) suggests that the pronunciation of adults will tend to improve quickly within their first year of immersion in an L2 setting, even in the absence of instruction. However, according to Flege and Liu (2001), this will only happen if learners receive abundant input from native speakers of the language. Research on study-abroad experiences has in fact reported gains in L2 oral proficiency after short stays abroad, including improvement in fluency (Llanes & Muñoz, 2009; Segalowitz & Freed, 2004) and accentedness (Muñoz & LLanes, 2014). These gains are generally found to be superior to those observed in comparable L2 learners studying in their home countries, pointing to the benefit of immersion. A longitudinal study by Munro and Derwing (2008) found evidence of improvement in the vowel intelligibility of 44 recent arrivals in Canada who were taking general ESL classes. The speakers' vowels became significantly more intelligible to four phonetically-trained, native English-speaking listeners, and most

improvement was found to occur during the first 6-8 months of residence. In these studies, however, the participants were enrolled in intensive ESL courses or in undergraduate programs, which practically guarantees them several hours of L2 input per week and often more opportunities to speak the language than international graduate students typically have. Graduate students tend to spend fewer hours in the classroom and do more independent study and research. Moreover, for those studying in large cosmopolitan cities and in the fields of engineer or computer science, such as the participants in this study, L2 input often comes from teachers and peers who are nonnative speakers themselves, and English is often used as a lingua franca. In some cases, students may have frequent interactions in their L1 with peers and friends who share the same country of origin. The present study inquired whether it is possible for these learners to improve their L2 pronunciation, oral presentation skills, and more specifically their use of paratones without receiving instruction in speaking.

This mixed-method study is part of a larger project examining naturalistic longitudinal changes in the oral proficiency of L2 graduate students at an English-medium Canadian university located in an effectively bilingual (French-English) city. Previous preliminary studies of data from this project (Buss, Cardoso, & Kennedy, 2015a, 2015b) included descriptive analyses of longitudinal changes in the use of paratones in English by L1 Mandarin graduate students (a sub-set of the larger sample) at four time points during their first 6 months of study.[2] While in some cases there seemed to be change over time, it was impossible to determine significance in these preliminary studies due to the very small sample sizes.

Research Questions

In view of the gaps identified in the literature, the current study explored the following research questions:

1. Do the speaking and oral presentation skills of L2 graduate students improve after two terms of study in an English-medium university, as judged by listener ratings of

[2] Four participants from the previous studies are also featured in the current study (Participants 2, 3, 8, and 9), but their data have been reanalyzed using a modified method.

accentedness, comprehensibility, fluency, topic structure clarity, and presentation overall quality?

2. Do L2 graduate students improve in their ability to signal topic structure through intonation (i.e., in their use of paratones) after two terms of study in an English-medium university?
3. How does the L2 graduate students' use of paratones relate to the listener evaluations? Is more native-like intonation at topic shifts related to more positive evaluations?

Method

Participants

There were two groups of participants in the study: talkers and listeners. The talker participants were 10 international graduate students (six males and four females) in engineering and computer science at an English-medium university located in an effectively bilingual (French-English) Canadian city.[3] The students had different L1 backgrounds: Mandarin (6), Farsi (2), Arabic (1), and Tamil (1), and they had met the English proficiency requirement of the university, which was a minimum of 6.5 on each band of the International English Language Testing System (IELTS) or a minimum score of 80/120 on the Internet-based Test of English as a Foreign Language (TOEFL). They were between 22-28 years old (Median = 24) at the beginning of the study, had been in Canada for no more than two months, and had never lived in an English-speaking country. Eight of the 10 participants reported that they heard and spoke English for only 3 hours a day or less, and it is reasonable to assume that at least some of this input came from nonnative speakers, given the city and the programs they were in. Two NSs of English who were graduate students in engineering also participated as controls: one male aged 24 and one female aged 37. The listeners were 11 NSs (two males and nine females), nine of whom were graduate students in applied linguistics. The other two were second-year

[3] Six participants from this study are also featured in the second study of this dissertation.

undergraduate students in TESL who had taken a course in pronunciation pedagogy. All of them reported normal hearing.

Data Collection

The speakers completed a language background questionnaire at the beginning of the study. Data were collected at the beginning and at the end of the NNS participants' first 6 months (26 weeks, two terms) of study. At each data collection session, the participants were video-recorded giving a presentation of approximately five minutes in which they explained a key concept or term in their field to an imagined audience of first-year undergraduates. Similar data were collected from the two NSs, with the only difference that their two presentations were not 26 weeks, but only three weeks apart for the female speaker and seven weeks apart for the male speaker. These data were used for comparison purposes.

The audio of the 24 videos collected were converted to WAV files, normalized for volume, and divided into two sets of 12. The first file by each speaker was randomly assigned to one of the sets, then the second file was placed in the remaining set. This was done so that the raters did not listen to more than one presentation by each speaker within the same rating session. The two sets were evaluated by the listeners in separate sessions of about 100 minutes, approximately one week apart. In each session, the files were randomly presented to the listeners and rated using a graphical computer interface created with *Z-Lab* (Yao, Saito, Trofimovich, & Isaacs, 2013), a custom-designed MATLAB program. As shown in Figure 1, the interface contained five 1,000-point scales with short descriptors at the endpoints and a box for comments, in which the listeners were asked to explain their ratings for "overall quality".

Figure 1. Computer interface used for listener ratings.

The sessions were conducted one-on-one with the main researcher or a research assistant in a quiet room using a laptop PC and headphones. The listeners read the written instructions and descriptions of the constructs (see Appendix A), discussed them with the researcher, and practiced the rating procedure before rating the presentations. Topic structure clarity (henceforth also simply structure clarity) was conceptualized as how easy or difficult it was for the listener to identify when one topic ended and another one started. Overall quality was described as the rater's subjective overall impression of the presentation. For this construct, there were no pre-imposed evaluation criteria and the listeners were encouraged to comment on their ratings.

Unfortunately, due to a technical problem with the computer interface, some ratings from five of the 11 listeners were lost. Therefore, while not all of the presentations were rated by all of

the listeners, each presentation was rated by at least nine listeners. Cronbach's alpha reliability tests revealed high agreement among the raters on accentedness ($\alpha = .97$), comprehensibility ($\alpha = 0.90$), and fluency ($\alpha = 0.91$) ratings, and acceptable agreement on structure clarity ($\alpha = 0.75$) and overall quality ($\alpha = 0.80$).

Coding of Discourse Topics

All of the presentations were transcribed and segmented into discourse topics based primarily on semantic content. Generally speaking, each discourse topic represented "an aggregate of coherently related events, states, and referents" (Chafe, 1994, p. 121). The presence of "macro-markers" (Chaudron & Richards, 1986), such as "The other thing is that…" and "The second thing we need to do…", were also considered when segmenting the presentations, as they made the topic structure (i.e., the organization of the presentation into different topics) explicit. Phonetic cues were not used as a primary criterion for identifying discourse topics to avoid circularity as much as possible. However, it was sometimes the case that parts of a presentation could be segmented in different ways depending on the choice of the speaker. As pointed out by G. Brown and Yule (1983), while a text may have indications of its topic structure, ultimately "it is speakers and writers who have topics, not texts" (p. 68). Since we did not have any other way of determining the speakers' intentions, whenever a topic boundary was ambiguous from the transcript, the main researcher listened to the audio for prosodic cues that indicated the exact location of the boundary and, if necessary, watched the video to check for any corroborating visual cues, such as a change of PowerPoint slides.

The concept of *abstract*, borrowed by Wennerstrom (1998) from Labov and Waletsky (1967), proved useful for the analyses of the discourse topics and pitch measurements. The abstract is an optional component which is sometimes used for signposting at a rhetorical shift. As an example, this is an excerpt from the female NS participant's first presentation:

End of Topic 1	…and this is measured in joules per second which is equal to **WATTS [159 Hz]**.
Abstract to Topic 2	So NOW [312 Hz] that we understand the difference between energy and power, we wanna understand the difference between energy density and power density.

Beginning of Topic 2	**ENergy [431.38 Hz]** density is most simply understood as how much energy is in some given volume, okay?

As can be seen, the utterance that begins with “So NOW that we understand” is not part of Topic 1 and could have been considered the beginning of Topic 2. However, it was produced with a smaller F0 peak than the word “ENergy" that appears later on. Thus, “ENergy” was considered the beginning of Topic 2, while the previous utterance was an abstract to Topic 2, that is, an intermediate element used by the speaker to signpost the transition to the new topic.

Acoustic Analysis

In order to analyze the speakers’ use of paratones, the maximum F0 values on the first prominent syllable and on the final tonic syllable of each discourse topic (S. Thompson, 1994, 2003). Then, the F0 value found at the end of one discourse topic was subtracted from the value found at the beginning of the subsequent topic, resulting in measures of pitch increase at topic shifts (henceforth simply pitch increase). These measures were then converted into percentages and averaged for each participant (as in Wennerstrom, 1998). The pitch measurements were done by using the “Get maximum pitch” function on Praat version 6.0.39 (Boersma & Weenink, 2018). In the few cases in which Praat did not display the pitch contour or made measurement errors due to low volume or creakiness, pitch was calculated manually from the waveform. In one instance, an F0 measure could not be obtained due to vowel devoicing. Given that there were only four topic transitions in this particular presentation (Participant 6, Time 2) and this was an important transition, which the speaker marked very explicitly, we chose to use the closest measurable F0 value to the target segment, which was produced with slightly higher pitch, but still very low in the speaker’s range.

Statistical and Qualitative Analyses

The ratings given by the listeners were analyzed using a two-way repeated measures ANOVA with time (Time 1 and Time 2) and construct (accentedness, comprehensibility, fluency, structure clarity, and overall quality) as factors, followed by post hoc tests using the Bonferroni correction. Paired *t*-tests were conducted to compare the NNSs’ Time 1 and Time 2 mean percentages of pitch increase. Finally, bivariate Pearson correlations were conducted to

examine whether degree of pitch increase at transitions was related to any of the listener ratings. The raters' written comments on overall quality were used to complement the quantitative findings. Open coding was used to look for common evaluation criteria that emerged from the data (Thomas, 2006) and comparing Time 1 and Time 2 comments for extreme cases; that is, participants who exhibited large changes in their listener ratings and pitch measures.

Results

Listener Ratings

When running the ANOVA, Mauchly's test indicated that the assumption of sphericity had been violated for the main effect of construct, $\chi^2(9) = 25.257$, $p = .003$, and for the interaction, $\chi^2(9) = 19.276$, $p = .026$, therefore Greenhouse-Geisser corrected tests are reported ($\varepsilon = .670$ and .562, respectively). A significant interaction between time and construct was found, $F(2.249, 20.244) = 4.666$, $p = .019$. Post hoc tests using the Bonferroni correction revealed that the participants' accentedness and comprehensibility ratings were significantly higher at Time 2 with large effect sizes ($p = .005$, $d = 1.172$, and $p = .004$, $d = 1.206$, respectively), indicating that they became less accented and more comprehensible. However, no significant change was found for fluency, structure clarity, and overall quality ratings ($p = .132$, .296, .629, respectively). Figure 2 displays the mean listener ratings for each construct at Time 1 and Time 2. Higher mean values are an indication of more positive evaluations (see Figure 1). The mean ratings for each speaker are provided in Appendix B.

A follow-up analysis of the participants' individual performance revealed that half of the NNSs and one NS actually experienced reductions in their mean ratings of structure clarity and overall quality from Time 1 to Time 2, while the rest had only slight changes. On the other hand, half of the participant did show positive changes in fluency ratings, but one participant, Participant 4, exhibited a sharp 23.13% decrease in her mean fluency and pulled the mean down. She was the only participant to experience a drop in fluency besides a negligible 1.14% decrease in Participant 5's ratings. By looking back at Participant 4's presentations, it was noticed that at Time 1 her presentation was read out from her notes, while at Time 2 she relied only on PowerPoint slides, which was normal among the other presenters. At least two raters noticed this

and mentioned in their comments that it sounded like she was reading her first presentation. It is reasonable to suppose that this gave the speaker an advantage at Time 1 that resulted in higher fluency ratings. In fact, some comments suggested that the main fluency issue at Time 2 was that she made long pauses to think about what to say next: "Some long, dysfluent pauses" (R6), "she did make a few long pauses to figure out what she wanted to say which made the fluency a little less nativelike" (R7), "Good fluency when fluent, but long pauses" (R11). When running the ANOVA without Participant 4, it was found that the assumption of sphericity was not violated for the interaction between time and construct, $\chi^2(9) = 16.332$, $p = .068$, and that this interaction was still significant, $F(4, 32) = 6.164$, $p = .001$. This time, the post hoc tests also revealed significantly higher fluency ratings ($p = .007$, $d = 1.215$), in addition to better accentedness ($p = .004$, $d = 1.303$) and comprehensibility ratings ($p = .004$, $d = 1.327$), with large effect sizes. In further analyses and reporting of results, however, Participant 4 will be included.

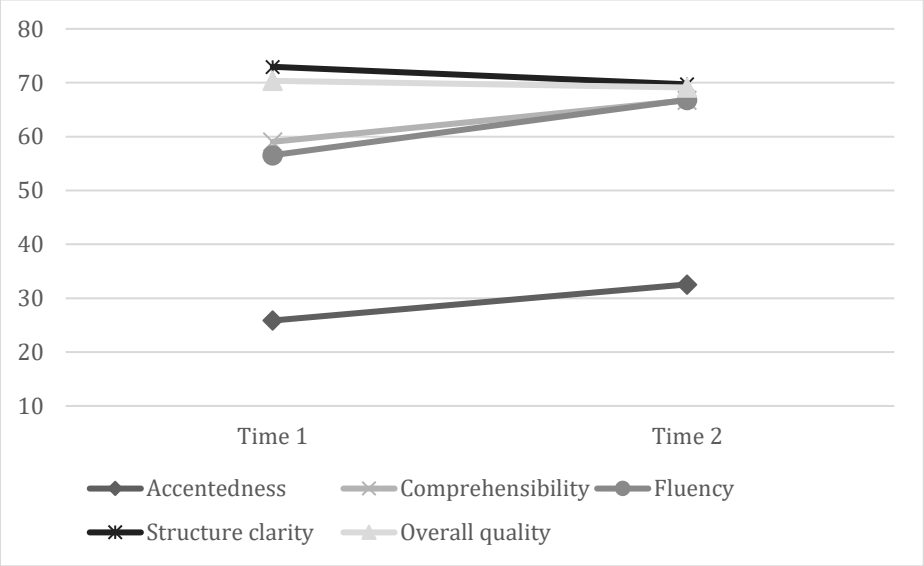

Figure 2. Mean listener ratings (0-100%) given to the NNS presentations.

Paratones

Table 1 below displays the average percentages of pitch increase at topic shifts for each participant's two presentations, as well as the number of topic shifts from each presentation. This number includes shifts from a general introduction to the first main topic. The coding of

discourse topics revealed relatively few rhetorical shifts in most presentations (Range: 1-8, Median = 3.5), due to their limited length. While all participants had been instructed to prepare five-minute presentations, Participant 4 at Time 1 and Participant 8 at Time 2 delivered very short presentations, only 2.5 and 2.8 minutes long, respectively. This resulted in their presentations having only one topic shift, which makes it difficult to gauge any potential changes in the speakers' command of paratones. Participant 12's second presentation also had only one transition, but this was due to the fact that most of her presentation involved deriving a single equation.

Table 1

Mean percentages of pitch increase at rhetorical shifts

	Gender	L1	Topic shifts		Mean pitch increase (%)	
			Time 1	Time 2	Time 1 (SD)	Time 2 (SD)
Participant 1	F	Farsi	4	5	52 (13)	88 (19)
Participant 2	M	Mandarin	5	5	68 (9)	59 (32)
Participant 3	F	Mandarin	3	4	42 (9)	40 (17)
Participant 4	F	Mandarin	1	5	16[a]	30 (11)
Participant 5	M	Farsi	5	4	54 (55)	40 (46)
Participant 6	M	Tamil	3	4	53 (21)	37 (50)
Participant 7	M	Arabic	3	3	54 (46)	50 (57)
Participant 8	M	Mandarin	2	1	32 (17)	14[a]
Participant 9	M	Mandarin	2	2	57 (14)	56 (13)
Participant 10	F	Mandarin	4	2	42 (25)	46 (24)
Participant 11	M	English	4	8	100 (61)	149 (45)
Participant 12	F	English	2	1	169 (3)	184[a]

[a] Only one topic shift; raw value for pitch increase; no mean or standard deviation.

Even within the same presentation, there was a fair amount of variability in the pitch levels used to signal a rhetorical shift. This appeared to be mainly due to different levels of semantic disjunction. Some topics that were introduced seemed to be "newer" than others, or

more distinct from the previous topic, thus beginning with a higher F0 peak. On other occasions, the last word of one topic was produced with special emphasis, which resulted in higher than normal pitch on that word and a smaller pitch increase when starting the next topic. This is the reason why there was a fairly large difference between the mean values found for control Participant 11's two presentations. On his first presentation, two topic-final words were emphasized, resulting in pitch increases of around 50%. If these two cases were to be ignored, Participant 11's mean value for Time 1 would go up to 152%, which is more consistent with his Time 2 value. Another issue was that some NNSs, especially Participants 5, 6, and 7, sometimes used rising intonation even when they were evidently ending a topic, which also resulted in small pitch increase values and contributed to very large standard deviations for these speakers.

Despite this variability, it is clear that NNSs produced considerably less marked rhetorical shifts than the NSs. NNS mean values of pitch increase ranged from 14-88% (Median = 48), whereas the NS mean values ranged from 100-184% (Median = 159). Furthermore, unlike the NNSs, the control participants appeared to have a minimum of 45% pitch increase to signal a new topic, and they usually increased pitch by over 100%, which was rare for NNSs. As in Wennerstrom (1998) and Buss et al. (2015a), the female NS had higher mean pitch increase values than the male NS, but this gender difference was not found for the NNSs.

Figure 3 shows examples of two rhetorical shifts in NNS presentations. In the first one, from Participant 2's second presentation, the percentage of pitch increase from "FUNCtional" to "microconTROLler" was 45%, close to the NNS median. The second example is from Participant 1's second presentation and shows a more pronounced pitch increase of 96% from "REgion" to "Other", which is closer to the values of the native speaker comparison participants.

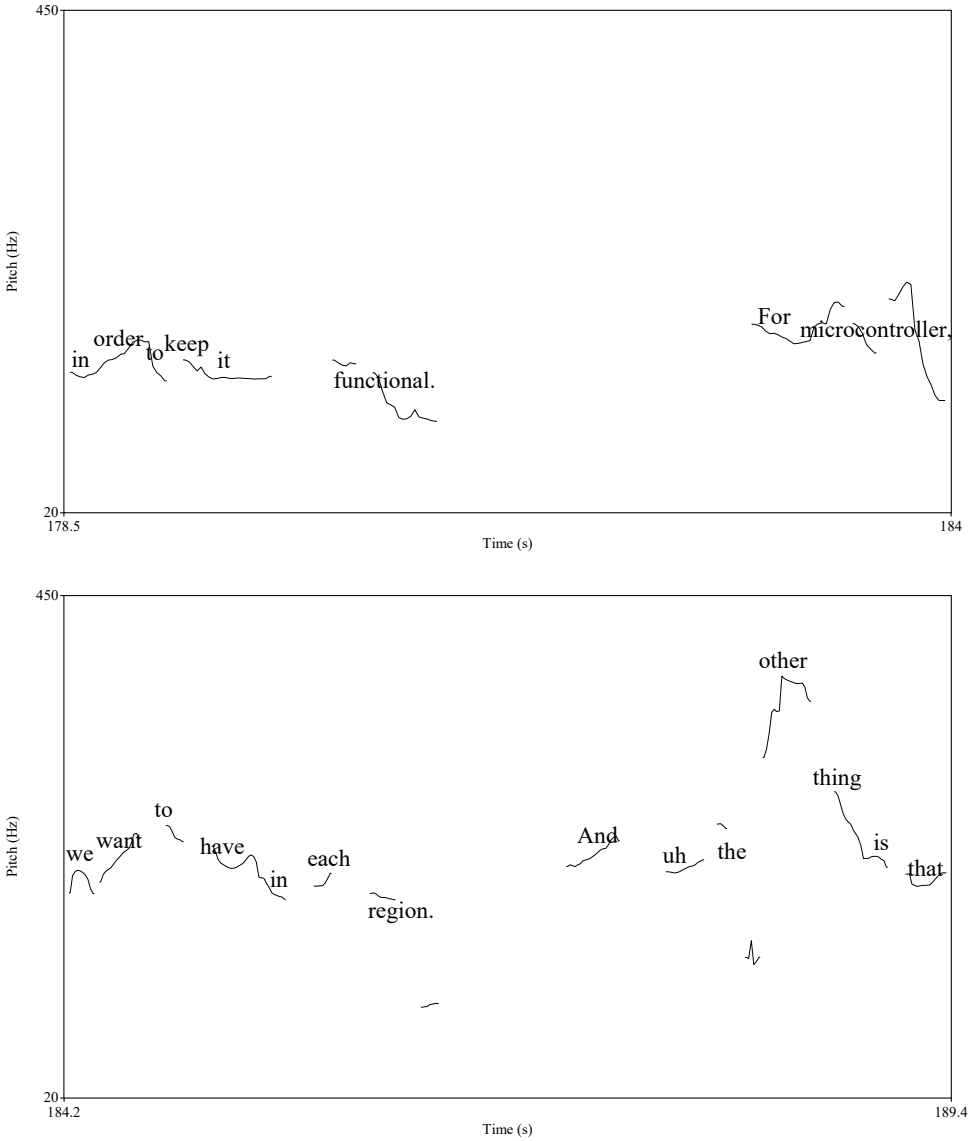

Figure 3. Examples of pitch increase at a topic shift in NNS presentations.

A paired *t*-test conducted with the NNS values of pitch increase revealed no significant change from Time 1 (M = 47.13, SD = 14.63) to Time 2 (M = 46.09, SD = 19.81), $t(9) = .201$, $p = .845$.

Relationships Between Variables

As shown in Tables 2 and 3, Pearson correlations with the NNS data revealed that structure clarity correlated very highly with overall quality: $r = .991$, $n = 10$, $p = .000$, at Time 1, and $r = .975$, $n = 10$, $p = .000$, at Time 2. This suggests the listeners did not distinguish much between the two constructs and that they may represent the same underlying construct. At the very least, it means that the clarity of the topic structure was a very important aspect when rating the overall quality of a presentation.

Table 2

Pearson Correlations between constructs at Time 1

	Accentedness	Comprehensibility	Fluency	Structure clarity	Overall quality	Pitch increase
Accentedness		.729*	.756*	.525	.533	.153
Comprehensibility			.892*	.896**	.911**	.224
Fluency				.896**	.899**	.097
Structure clarity					.991**	.069
Overall quality						.090

* $p < .05$, ** $p < .01$, two-tailed

Table 3

Pearson Correlations between constructs at Time 2

	Accentedness	Comprehensibility	Fluency	Structure clarity	Overall quality	Pitch increase
Accentedness		.742*	.626	.105	.240	-.008
Comprehensibility			.710*	.569	.668*	.264
Fluency				.483	.548	.209
Structure clarity					.975**	.522
Overall quality						.521

* $p < .05$, ** $p < .01$, two-tailed

Pitch increase at topic shifts did not correlate significantly with any of the listener ratings. However, at Time 2, it was moderately, albeit non-significantly, correlated with structure clarity (r = .522, n = 10, p = .122) and overall quality (r = .521, n = 10, p = .122). At Time 1, both structure clarity and overall quality correlated significantly and very highly with comprehensibility and fluency (see Table 2). At Time 2, following significant improvement in the NNS participants' accentedness and comprehensibility, structure clarity did not correlate significantly with any of the speech constructs, while overall quality only correlated significantly with comprehensibility, and less highly than before (see Table 3). It is possible that the participants' improved speaking skills meant fewer pronunciation and fluency issues detracting from other relevant factors.

Listener Comments

Open coding was used to identify criteria besides accentedness, comprehensibility, and fluency which were mentioned by the listeners as affecting overall quality. Table 4 lists aspects that were often said to positively affect quality when present or well done and to negatively affect quality when absent or poorly done, along with illustrative comments.

Table 4

Common Evaluation Criteria for Overall Quality of Presentation

Criteria	Examples (Rater number)
▪ Clear discourse structure or logical sequencing of ideas	"the structure seemed to follow in a logical way" (R4) "could have been structured better by using transition words" (R7)
▪ Quality of content or ideas	"Good ideas" (R1); "very informative" (R8) "His presentation was more interesting than the others, maybe because of the topic" (R10)
▪ Introduction, outline, or overview at the beginning	"Speaker provided an overview of the presentation, which helped with its clarity and quality" (R3)

	"did not provide an introduction, explain the importance, etc." (R2)
▪ Explanations, descriptions, and definitions	"the explanations did not seem complete" (R4) "Provides various definitions to gain a better understanding of 'the cloud'" (R6)
▪ Examples	"gave many useful examples" (R4) "She offers a lot of examples to clarify her points" (R7)
▪ Questions	"Asks rhetorical questions to get audience thinking" (R3) "raises a question to pique interest" (R6)
▪ Addressing of the audience	"Addresses his audience personally ("you may be wondering")" (R8) "Addressing audience in order to develop topics" (R11)
▪ Intonation and naturalness	"the speaker used many different tones in his speech, which elicited my attention" (R3) "Sounds like he's reading near the middle (interrupts natural sounding speech - makes for more monotone speech)" (R11)
▪ Confidence and knowledgeability	"He appeared to be extremely unsure of himself" (R1) "he seems knowledgeable" (R9) "Seems very confident in his knowledge" (R10)
▪ Conclusion or summary at the end	"concludes by summing up everything he has explained during his presentation" (R7) "He seems to be summarizing at the end, which helps to bring everything together" (R8)

Regarding the first item, almost half of the listeners' comments (117 out of 263) made some kind of reference to the structure of the presentation, which seems to corroborate the high correlation result between topic structure clarity and overall quality. In terms of factors that directly contributed to a clear structure, several comments referred to a speaker's explicit discourse structuring (or lack thereof) through lexical discourse markers, questions posed to

introduce a new topic, and statements indicating speaker intention (e.g., an overview of the presentation topics at the beginning). However, the raters also indicated that structure clarity was affected by whether the ideas in the presentation followed a logical and coherent sequence (e.g., became progressively more complex) or whether the speaker jumped around in their topics. Other items in Table 4 may also have indirectly contributed to making the discourse structure more explicit to the listener, such as a conclusion summarizing the topics addressed or examples after each point in a presentation.

Discourse structuring through pitch variations (i.e., use of paratones) was never mentioned by the raters. Even so, there was still evidence of a potential link between use of paratones and structure clarity. Participants 5 and 6, the two participants with the largest decreases in structure clarity from Time 1 to Time 2 (16% and 23%, respectively), also exhibited reductions in their mean percentages of pitch increase at topic shifts (14% and 17%) and in their overall quality ratings (13% and 21%), even thought their speech ratings for accentedness, comprehensibility, and fluency hardly changed. The comments confirmed that the second presentations by Participants 5 and 6 had less explicit topic structures and also indicated that they had fewer of the other aspects listed in Table 4. The most extreme case was Participant 6. Referring to his first presentation, the listeners wrote: “clearly explains what he is drawing (…) outlines the steps that need to be taken (…) explains the rationale” (R6), “had a clear introductory sentence” (R7), “addresses his audience personally” (R8), “asking questions in order to move to the next topic (…) clear structure” (R11), among other comments. On the other hand, his second presentation motivated opposite observations: “most of his speech was directed at the whiteboard” (R1), “did not provide an introduction, explain the importance” (R2), “there didn’t seem to be much structure” (R7), “it would be helpful if he stated a main goal and clear steps” (R10), “Very fast – no questions or pauses” (R11), among other similar remarks.

Moreover, as seen in Table 4, there was evidence that variations in intonation which characterize paratones may have influenced overall quality in other ways. Some raters referred to whether the presentation had good intonation, was natural or robotic, or sounded interesting, with tone variations, or monotone. Others referred to how confident or knowledgeable the speaker sounded, which may also have been influenced by intonation, in addition to fluency and other factors. One interesting case was Participant 1, who was the speaker with the largest positive

changes in fluency (22%), pitch increase values (36%), and overall quality rating (14%) from Time 1 to Time 2. At Time 2, she had pitch increase values which were most like the native comparison talkers (her mean was 88% of increase) and the highest ratings for overall quality among NNS participants (85%). While her presentations did not differ much in terms of structure clarity (according to ratings and comments), her first presentation, which was less fluent, less comprehensible, and more accented than her second, was considered "very monotone, which made the presentation a bit boring" (R3), with "kind of robotic intonation" (R7), and "very robotic" (R10). At Time 2, no such comments were found. Instead, there were the following positive comments: "She was fairly confident in her speech" (R1), "A clear and confident speaker, she seems very knowledgeable in her field" (R10). Similar words were used to describe the NS presenters: "confident" (R2, R10), "animated" (R3), "variation in tone that made it more interesting" (R4), "knowledgeable" (R8), "comfortable as a speaker" (R10), "energetic" (R10), "good variation in intonation" (R11).

Discussion

The first two research questions that guided this study asked about potential changes in the speaking and oral presentation skills of L2 graduate students, including their use of paratones, after two terms of study in an English-medium university. The findings revealed significantly higher listener ratings for accentedness and comprehensibility at Time 2. This supports Flege's (1988) prediction that the pronunciation of adult L2 speakers will improve naturalistically within their first year of residence in an L2 setting; however, it is unclear whether substantive amounts of native speaker input are in fact an absolute condition for this to occur (Flege & Liu, 2001). The graduate students in this research studied in English, but in a multilingual context, with many non-native English speakers in their environment. This suggests, happily for L2 graduate students, that even an immersion experience with potentially limited native-speaker input and no instruction in L2 speaking can result in pronunciation gains.

On the other hand, no overall improvement was observed in the L2 graduate students' ratings of fluency, structure clarity, and overall quality, or in their mean percentages of pitch increase at topic shifts. While half of the participants seemed to have improved their fluency at Time 2, four experienced little change, and one participant showed a sharp decrease which pulled the mean

down. For half of the participants, structure clarity, overall quality, and pitch increase were actually lower at Time 2, the largest drops being experienced by Participants 5 and 6. The listener comments on overall quality indicated that at Time 2 Participants 5 and 6 provided less explicit marking of the discourse structure and fewer elements that may contribute indirectly to a clearer topic structure, such as introductions, examples, questions, and conclusions.

It is possible that, instead of reflecting the participants' presentation skills or abilities, the results showing a decrease simply reflect a difference in planning, since there is evidence that amount of planning affects fluency (Ellis, 2009; Foster & Skehan, 1996) and the degree of discourse marking in L2 speech (Crookes, 1989; Pickering, 2004; Williams, 1992). Some participants may have prepared less for their second presentation because they were more confident in their speaking skills, more familiar with the study task, and possibly too busy, given that Time 2 coincided with the end of the winter term. Therefore, planning may have been another cause of variability in a speaker's degree of pitch variation at topic shifts, in addition to variability in the emphasis and level of semantic disjunction or "newness", which made it difficult to assess potential improvement in the participants' abilities to use paratones. It seems that one would need to average a fairly large number of pitch increase values in order to find consistent means for each speaker and be able to detect reliable changes over time. The presentations in this study, however, were found to have only a small number of rhetorical shifts. In spite of this variability, one consistent finding was that, overall, the NNSs were far from the native values, which confirms previous studies (Pickering, 2004; Wennerstrom, 1994, 1998).

The third and last research question inquired about the relationships between the L2 graduate students' use of paratones and the listener evaluations. Pitch increase at topic shifts was not significantly correlated with any of the ratings, although there was some indication that it may have become more important to overall quality at Time 2, following significant improvement in the participants' accentedness and comprehensibility. Topic structure clarity was a very important aspect to overall quality, which is in line with previous studies suggesting that degree of discourse marking predicts the effectiveness of L2 oral presentations and lectures (Tyler et al., 1988; Williams, 1992). However, native-like use of intonation to mark topic shifts was not found to be necessary for a presentation to be considered structurally clear. Participant 1's first presentation, for example, whose mean percentage of pitch increase was close to the NNS

median, was already considered to have a clear topic structure, based on the ratings and comments. At Time 2, her much improved use of paratones was not paralleled by a noticeable increase in structure clarity.

We propose three related explanations for these findings. First, given that L2 speakers tend to produce narrower pitch ranges overall (Mennen, 1998; Pickering, 2004; Zimmerer, Jügler, Adreeva, Möbius, & Trouvain, 2014), it is possible that the listeners "tuned in" to the NNSs' flatter intonation, so that, although their mean pitch increase values did not approximate those of native speakers, they may have still been salient enough in relation to the rest of the presentation to distinguish a new topic. In fact, in Kang et al. (2010), the number of low terminations followed by high-key resets (i.e., the number of identifiable paratones cues) was positively related to listener ratings of oral proficiency and comprehensibility, while the pitch levels themselves of paratone onsets and terminations appeared to be less important. Second, as indicated by the comments, other elements may have helped to signal the discourse structure to the listeners, including lexical discourse markers, statements indicating speaker intention, and a logical sequence of topics. Considering that L1 speakers also use a variety of devices to indicate the organization of their speech, and not necessarily all at the same time (S. Thompson, 1994; Tyler, 1992; Tyler et al., 1988), it makes sense that the presence of other cues may compensate for weaker intonational cues in L2 speech. Finally, because the listeners were asked to rate topic structure, they had to pay conscious attention to this aspect when evaluating the presentations. As a result, they may have needed fewer or less explicit organizational cues to perceive the topic structure than perhaps an unprompted audience would.

Despite these findings, there was still evidence that pitch increase at topic shifts was linked to other discourse structuring cues and to perceptions of topic structure clarity. In Participants 5 and 6's data, comments indicating a decrease in structure-signaling elements from Time 1 to Time 2 were accompanied by an empirical reduction of average pitch increase at topic shifts, which suggests an interconnection between different cues that serve the same purpose. For these participants, reduced effort to cater to their audience at Time 2 by providing an introduction, signposting, explaining, asking questions, etc., all of which may contribute to discourse structure salience, also meant less effort to make topic transitions explicit to the listeners by employing

pitch variations. The less distinct discourse structure resulted in lower topic structure clarity ratings and less favorable overall quality evaluations.

Going back to Participant 1's case, even though considerably higher pitch increase did not result in a much clearer discourse structure, it may have influenced her overall quality ratings in other ways. It is possible that Participant 1's more native-like (i.e., more varied, less flat) intonation at Time 2, as indicated by more marked paratone cues, combined with greater fluency, caused listeners to perceive the speaker as more interesting (see Hincks, 2005) and more confident. Indeed, the fact that people have narrower pitch ranges in their L2 has been attributed to lack of confidence (Mennen, 1998; Zimmerer et al., 2014), so it is possible that more native-like values in the range of intonation may be interpreted as a sign of confidence.

Finally, it should be noted that our results did not confirm those in Wennerstrom (1998), since there was no correlation between pitch increase at topic shifts and comprehensibility. In hindsight, this is not entirely surprising given that comprehensibility in this study was very precisely defined as how easy or difficult it was for the listener to identify the speaker's words (see Appendix C). Some comments appeared to explicitly differentiate between understanding the speech and being able to follow the ideas or the topics of the presentation themselves, and it seems that the latter, although not measured in this study, would be more affected by discourse marking. Another possibly relevant factor is that the participants in Wennerstrom (1998) apparently had lower proficiency in English than the speakers in this study, based on the minimum TOEFL scores reported and on the information that they had previously failed the SPEAK test (Educational Testing Service, 1985). Finally, lack of power due to the small sample size of this study may have contributed to the non-significant results.

Conclusion

In summary, the findings of this study indicate that it is possible for L2 graduate students studying in an English-medium university to improve their accentedness and comprehensibility fairly quickly, even without instruction in speaking and perhaps with more limited native-speaker input compared to previously studied immersion experiences. However, fluency, presentation structure clarity, presentation overall quality, and pitch increase measures appear to be more variable and possibly influenced by the amount of planning of the presentation. Mean

percentages of pitch increase at topic shifts were not linearly correlated with ratings of oral proficiency, structure clarity, or overall quality. There was evidence that other aspects of the presentations contributed to making the topic structure explicit, and that even smaller pitch increases than those typically produced by NSs may still have been salient enough to cue the listeners. One participants' increase to native-like levels at Time 2 was found to have little effect on structure clarity, although it may have contributed to enhanced listener perceptions of interestingness and speaker confidence. If in fact true, this means that L2 speakers do not need to have native-like intonation in order to signal topic structure effectively and that they may benefit from learning to use multiple structuring devices. They may also benefit from being made aware of the importance of structural clarity in oral presentations and lectures and of the potential negative effects of poor planning on structural clarity and overall quality. These conclusions, however, should be interpreted with caution, especially given the sample size and the small amount of data from each participant used for the paratone analysis. Future research needs to further explore this topic in order to establish the importance of paratones in L2 communication and the relative priority that they should be given in the ESL and EAP classroom.

Connecting Study 1 to Study 2

The results of Study 1 indicated that L2 graduate students can improve their pronunciation without instruction to some extent, and that native-like use of paratones may not be necessary for L2 presentations to be perceived as structurally clear and effective. It must be noted, however, that the raters in that study were mostly graduate students in applied linguistics who had an interest or conducted research in L2 pronunciation, and there is evidence that language experts may be more lenient in their evaluations of pronunciation (Saito et al., 2017; I. Thompson, 1991). Therefore, it is still unclear how the L2 graduate student presentations would be evaluated by listeners without a background in language teaching, and whether these listeners perceive naturalistic changes in the students` pronunciation. It is even more important to know how the students' potential audience (in this case, listeners with a background in engineering or computer science) evaluate their speech and presentations, and how the specialists' assessment criteria differ from those of non-specialists. This topic was addressed by Study 2, which also included two additional time points: approximately two months and four months after the first data collection session. No assessment of paratones or structure clarity was conducted, but the new construct *speaking style* was included for rating by the listeners. Six of the participants in Study 1 were also featured in Study 2: Participants 4, 6, 7, 8, 9 and 10 in Study 1 were Participants 1, 6, 8, 9, 10, and 11 in Study 2, respectively.

Chapter 3: Study 2

Oral performance in L2 academic presentations: Longitudinal changes and the role of subject-matter knowledge in listener evaluations

Abstract

The ability to speak English proficiently and give oral presentations in English is essential but challenging to international graduate students (Berman & Cheng, 2010). Yet they often go through their programs with little or no instruction in L2 pronunciation. This study had two overall aims. The first was to examine whether and how different aspects of these students' oral performance change over time. The second was to examine the aspects of oral performance that are most valued by these graduate students' target audience (subject-matter specialists) when listening to academic presentations, and how these differ from the criteria valued by naïve listeners. To answer these questions, academic presentations given by 11 L2 graduate students in engineering and computer science were recorded four times during their first two terms of study in an English-medium Canadian university. Speech samples taken from these presentations were evaluated on six constructs (accentedness, comprehensibility, content, fluency, organization, and speaking style) by two groups of listeners: 17 content specialists and 15 non-specialists. Overall, no significant change was found for any of the constructs, but an analysis of individual performance revealed that some speakers did seem to improve, particularly in fluency and speaking style. The main difference between the evaluations given by the two listener groups was that specialists appeared to value content and organization more than non-specialists, and their comments tended to be more specific and mention a wider range of assessment criteria.

Introduction

Every year, American and Canadian universities receive thousands of international students who speak English as a second language. Surveys conducted by Berman and Cheng (2010) and Ferris (1998) have shown that these students are concerned about their speaking skills and the speaking tasks that they are expected to perform at university, such as oral presentations and class discussions. Despite having met the English language standards of their universities, they feel that they need to improve their oral proficiency, but they often carry out their studies without receiving any specific instruction in L2 speaking or pronunciation. Given this situation, it is important to know whether and how L2 graduate students' oral performance develops in the absence of focused instruction and what aspects are most important for instructors to focus on in order to help these students become better communicators in their academic communities. The current study addressed these issues by examining longitudinal changes in the L2 speaking performance of 11 graduate students based on evaluations provided by subject-matter specialists and non-specialists, all of whom were native English speakers. Differences between the evaluations given by the two groups of raters were also examined in order to probe into the aspects of speech that are most valued by the speakers' target audience.

Naturalistic Development of Adult L2 Speech

The acquisition of L2 phonology by adult learners who are immersed in the target language environment has been the subject of many investigations. Some have looked at study-abroad experiences, often comparing them to "at home" settings, and found positive pre-post effects on foreign accent (Muñoz & LLanes, 2014) and especially fluency (Freed, 1995; Llanes & Muñoz, 2009; Segalowitz & Freed, 2004). Another body of research has employed cross-sectional designs to compare the production of L2 speakers with different lengths of residence (LOR) in the L2 environment. Findings from these studies have been mixed. While some have found longer LOR to be associated with better accent (Flege & Fletcher, 1992), vowel intelligibility (Flege, Bohn, & Jang, 1997), stress timing (Trofimovich & Baker, 2006), among other aspects, other studies have found no such effect (Flege, Munro, & Fox, 1994; Munro, 1993; I. Thompson, 1991).

One explanation for the conflicting results on the effect of LOR, mentioned by Piske, MacKay, and Flege (2001), is that LOR is only a rough index of L2 experience, since learners with similar LOR may differ in terms of actual quantity and quality of input received. Evidence for this comes from Flege and Liu (2001), who compared groups of Chinese immigrants to the United States that differed in LOR and occupational status: long and short LOR; students and non-students. The non-students had little or no education in the United States and had jobs that required little use of English, whereas the students had to use English to interact with professors and peers. The speakers' L2 proficiency was evaluated through three tests: a word-final consonant identification test, a grammaticality judgment test, and a listening comprehension test. In all of them, students with a longer LOR had significantly higher scores, but LOR had no effect on the non-students' scores. The authors concluded that the effect of LOR is only present when learners are immersed in an input-rich environment, that is, when they receive abundant input from native speakers of English.

Another possible explanation for the inconsistent findings on LOR is offered by Flege (1988), who suggests that adult L2 speakers go through a rapid phase of pronunciation improvement in their first year of residence, but tend to "fossilize" after that. Flege (1988) assessed the degree of perceived foreign accent in the speech of Chinese immigrants to the United States. The scores received by two groups who had lived in the L2 setting for one year and five years did not differ significantly. According to the researcher, it is possible that adult L2 pronunciation ceases to improve relatively early, so additional unaided experience with the language does not result in better pronunciation.

A third possible explanation, proposed by Munro and Derwing (2008), is that the cross-sectional design may be causing researchers to compare groups of learners that differ systematically on other aspects besides LOR. Thus, longitudinal studies can provide a better perspective on how language develops over time and at what point pronunciation starts to plateau (Munro & Derwing, 2008; Piske, Mackay, & Flege, 2001). Munro and Derwing (2008) conducted a longitudinal study of the L2 vowel production of 44 ESL speakers from different L1 backgrounds (Mandarin and Slavic languages) enrolled in a general ESL program. The participants were recorded repeating target words six times over their first year of residence in Canada. Overall, speakers from both linguistic backgrounds showed improvement in vowel

intelligibility, as judged by four phonetically trained Canadian listeners, but the largest changes occurred in the first half of the year. This was interpreted as support for Flege's (1988) hypothesis.

In another study that was part of the same longitudinal research project, Derwing and Munro (2013) investigated 11 Mandarin and 11 Slavic language speakers who received two years of basic ESL classes and were evaluated at different points in time by 44 untrained listeners (two months, two years, and seven years). By the end of the second year, the Slavic group showed improved ratings for fluency, comprehensibility, and accentedness. Five years later, the same group was perceived as more comprehensible and fluent than at Year 2, even though they had not received any further ESL instruction. Conversely, the Mandarin speakers, who were more reluctant to communicate in English, did not show significant improvement on any of the three dimensions across the seven-year study. The authors concluded that L2 learners may continue to develop their oral English skills after their formal ESL training. However, this development may be limited and dependent on a number of variables, including L1, age of arrival, L2 use and exposure, and willingness to communicate.

There is still a need for further longitudinal studies like the ones reviewed above, which investigate L2 speech development of adults who are immersed in a naturalistic setting (Munro & Derwing, 2015; Piske, Mackay, et al., 2001), particularly in the absence of L2 instruction. The findings of such studies may help to inform ESL and EAP courses about aspects that are more difficult to learn naturalistically and that therefore require more explicit attention from instructors. The current study addressed this important issue by investigating longitudinal changes in the oral skills of uninstructed L2 graduate students in the context of academic presentations. Knowing the aspects of speech that do and do not improve incidentally, however, is not enough to inform the selection of targets for instruction in ESL or EAP courses. International university students who wish to improve their academic speaking need to know the aspects of speech that their particular academic community takes into consideration when forming judgments about L2 speech. In other words, they need to be able to focus on the aspects that are valued by listeners in their fields of study, such as professors, colleagues, and peers.

L2 Speech Ratings by Subject-Matter Specialists and Non-Specialists

Several studies have examined the influence of subject-matter background on perceptions of L2 speech. Most of them have done so by comparing how ESL teachers and subject specialists evaluate authentic or simulated professional interactions. Lumley (1998) reported reasonable agreement between ESL-trained listeners and doctors in their ratings of clinically-based role-plays in terms of their "overall communicative effectiveness". However, studies investigating narrower constructs have found important differences between assessments by specialists and non-specialists. In Brown (1995), language teachers and raters with a background in tour guiding were asked to rate candidates for the Japanese Language Test for Tour Guides, which involves simulated interactions between tour guides and tourists. Although the overall grades awarded were similar, the two groups diverged in their perceptions of different aspects of speech. Teachers tended to be harsher on grammar and expression, vocabulary, and fluency, while industry raters marked pronunciation and comprehension more harshly.

Douglas and Myers (2000) investigated the comments made by applied linguists and veterinary professionals as they discussed videos of veterinary students interviewing clients about sick animals. While most of their evaluation criteria overlapped, the vets seemed to place greater emphasis on the students' relationship with the client and on their content knowledge. In Read and Wette (2009), role-plays performed by health professionals in a preparatory course for IELTS and the Occupational English Test received lower scores and less positive comments from a medical professional compared to the course tutor. The medical specialist's assessment scores and comments addressed weaknesses in areas such as knowledge of the medical content, fluency and intelligibility, questioning techniques, empathy, clarity of explanations, and ability to provide a coherent structure to the consultation.

Turning to an academic context, Elder (1993) compared how ESL teachers and subject specialists assessed videotaped or live classroom performances of graduates training as mathematics and science teachers. The raters used an observation schedule covering different dimensions of language use. When assessing the videos, the two groups of raters diverged in their ratings of "subject-specific language use", which included items such as knowledge of specialist terms, clear connections between ideas, and explanations of processes/concepts in ways appropriate to the audience. The evaluations of live performances revealed that the teachers

were more focused on traditional components of language proficiency in making their global assessments, while the specialists placed the greatest emphasis on the dimension of "classroom interaction", which included items such as posing questions to check understanding, dealing effectively with wrong answers, and adopting an appropriate level of formality.

In order to further explore what is valued by different professional and academic communities, it is helpful to look at studies in English for Specific Purposes (ESP) investigating "indigenous assessment criteria" (Jacoby, 1998), i.e., criteria used by subject specialists to evaluate the communicative performances of novices and colleagues. Two studies involving professionals from very different fields have shown little difference in how native and nonnative speakers are evaluated (Elder et al., 2012; Jacoby & McNamara, 1999). In Elder et al. (2012), three groups of health professionals were asked to provide general feedback on instances of trainee-patient communication. There were both native and nonnative English-speaking trainees. The professionals were found to have similar assessment criteria, which were more focused on health-specific aspects (approach, content, organization, techniques, terminology, among others) than on the speakers' language skills. Jacoby and McNamara (1999) discussed the findings of Jacoby (1998), which analyzed the feedback provided by members of a university physics research group to native and nonnative speakers' pre-conference dry runs. The evaluation criteria employed by the physicists included the ability to articulate the motivation/importance of the research, clarity, economy of expression (i.e., concision), persuasion, content accuracy, and overall quality. Comments addressing linguistic errors were few and only made when the errors were observed in the visual aids. Except for linguistic error, all the other criteria were applied equally to all presenters, regardless of their native/nonnative status.

Knoch (2014) conducted focus group interviews in which 10 pilots were asked to evaluate the communicative effectiveness of L2 speakers taking aviation English tests. The main criteria mentioned by the pilots were technical knowledge and pronunciation, and the speakers' perceived technical knowledge was found to outweigh their linguistic shortcomings to some extent. Finally, Pill (2016) investigated the indigenous assessment practices of healthcare professionals when evaluating consultations between trainees and their patients. The participants were found to value the trainees' knowledge and the language they used. They were particularly

concerned about language behaviors that could facilitate or hinder interaction, such as word choice (use of everyday language or medical jargon), paraphrasing, and signposting.

The studies above suggest that subject specialists place more emphasis on content, organization, and skills that are specific to the oral genre under assessment than on general linguistic skills. It would appear that they are more concerned with the accomplishment of communicative goals than with the quality of language *per se* (Elder, 1993; Elder, Mcnamara, Kim, Pill, & Sato, 2017). However, the studies reviewed above seem to be at opposite extremes. They either provided specialists with inflexible quantitative instruments with no room for comments (e.g., Brown, 1995) or they asked them to provide general feedback, without probing the discussion of specific constructs (e.g., Jacoby, 1998). In the latter studies, it is quite possible that the raters were unwilling to comment on pronunciation because they did not feel competent to judge in this area (Elder et al., 2012). Thus, still little is known about how subject-matter specialists interpret and evaluate different dimensions of pronunciation (accentedness, comprehensibility, etc.). The present study helps to fill this gap by comparing the ratings and written comments provided by subject-matter specialists and non-specialists when evaluating the oral performance of the L2 graduate students on six constructs: accentedness, comprehensibility, content, organization, fluency, and speaking style.[4]

The Current Study

To sum up, there is evidence that L2 oral skills can develop in the absence of focused instruction, but this may be more likely to happen with some aspects of speech than others. Identifying the aspects that are more prone to incidental development and those that may need to be taught can help to define pedagogical priorities for ESL and EAP instruction at universities. However, it is also important to determine the aspects of oral performance that are most valued by the listeners that L2 university students will encounter in their professional and academic careers. Prior studies point to differences between how content specialists and non-specialists

[4] "Speaking style" is here defined as the degree to which the speaker uses pitch changes which make their speech sound dynamic and interesting, as opposed to monotone and boring (see Hincks, 2005). Definitions for the other constructs can be found in the Appendix.

assess oral performance, but it is still unclear how the former evaluate and interpret different dimensions of pronunciation. In light of these gaps in the existing literature, the current study addresses two main research questions:

1. Do evaluations of L2 graduate students' oral performance improve over time in the absence of explicit instruction? If so, which of the following aspects show the most improvement: accentedness, comprehensibility, content, organization, fluency, or speaking style?
2. How do evaluations of oral performance differ when provided by content specialists and non-specialists? Which aspects (accentedness, comprehensibility, etc.) are most valued by the specialists?

Method

Participants

There were three groups of participants in this study: one group of talkers and two groups of listeners. The talker participants were 11 international graduate students (five males and six females) starting their first year of studies at an English-medium university in an effectively bilingual (French-English) Canadian city. All of them had had multiple years of English study in their home countries (Range: 8-17 years, Median = 13 years), starting at childhood or puberty (Range: 4-13 years old, Median = 9 years old). Only one participant (Participant 5) had lived in an English-speaking country before her arrival in Canada, and all but two participants (Participant 2 and 3) had been in Canada for one month or less (Age of arrival range: 20-34, Median = 25). Most of them spoke Mandarin as their L1 and studied engineering or computer science. All of them had a minimum of 6.5 on each band of the International English Language Testing System (IELTS) or a minimum score of 80/120 on the Internet-based Test of English as a Foreign Language (TOEFL), as required by the university. Based on their test scores and self-ratings of their English-speaking skills, five participants were considered upper-intermediate-level speakers, with IELTS 6.5 or TOEFL 93 or lower (Participants 1, 2, 9, 10, and 11), while the remaining six participants had higher scores and were considered advanced-level speakers

(Participants 3, 4, 5, 6, 7, and 8). The participants did not receive any instruction in pronunciation or academic speaking during the six months of the study.

The listeners were all native speakers of English with self-reported normal hearing. One group of listeners, referred to as "non-specialists", consisted of 15 full-time undergraduate students in programs unrelated to engineering or computer science. The second group, referred to as "subject-matter specialists" consisted of 17 graduate students in different branches of engineering and computer science: 13 Masters and four PhD students; 10 in engineering and seven in computer science.

Talker data collection

Speech tasks and materials. Talker data were collected at four time points over two consecutive academic terms (26 weeks): Time 1 was at the beginning of the participants' first semester of studies (September/October 2011), Time 2 was at the end of their first semester (November/December 2011), Time 3 was at the beginning of their second semester (January 2012), and Time 4 was at the end of their second semester (March 2012). At each data collection session, the participants were video-recorded giving a five-minute presentation where they explained a key concept or term in their field to an imagined audience of first-year undergraduates. This task was chosen because oral presentations are common communicative events for graduate students.

Speech sample preparation. A total of 44 presentations were video-recorded (four for each of the 11 participants), but only their audio component was used for analysis. Samples of approximately one minute were taken from the beginning of each audio file. In order to control for the effects of topic familiarity on listener ratings, a short phrase stating the topic of the presentation was recorded by a female native speaker and added to the beginning of each sample. The pool of 44 samples was equally divided into two sets of 22. Two samples by each talker were randomly assigned to each set, with both sets being presented to the raters in alternating order (i.e., rater 1 heard set 1 before set 2, rater 2 heard set 2 before set 1, and so on). This was done to reduce the likelihood of raters listening to two or more consecutive samples by the same talker. A graphical computer interface was created for the rating sessions with *Z-Lab* (Yao et al.,

2013), a custom-designed MATLAB program. As shown in Figure 4, the interface contained six continuous scales, one for each target construct, with short descriptors at the endpoints.

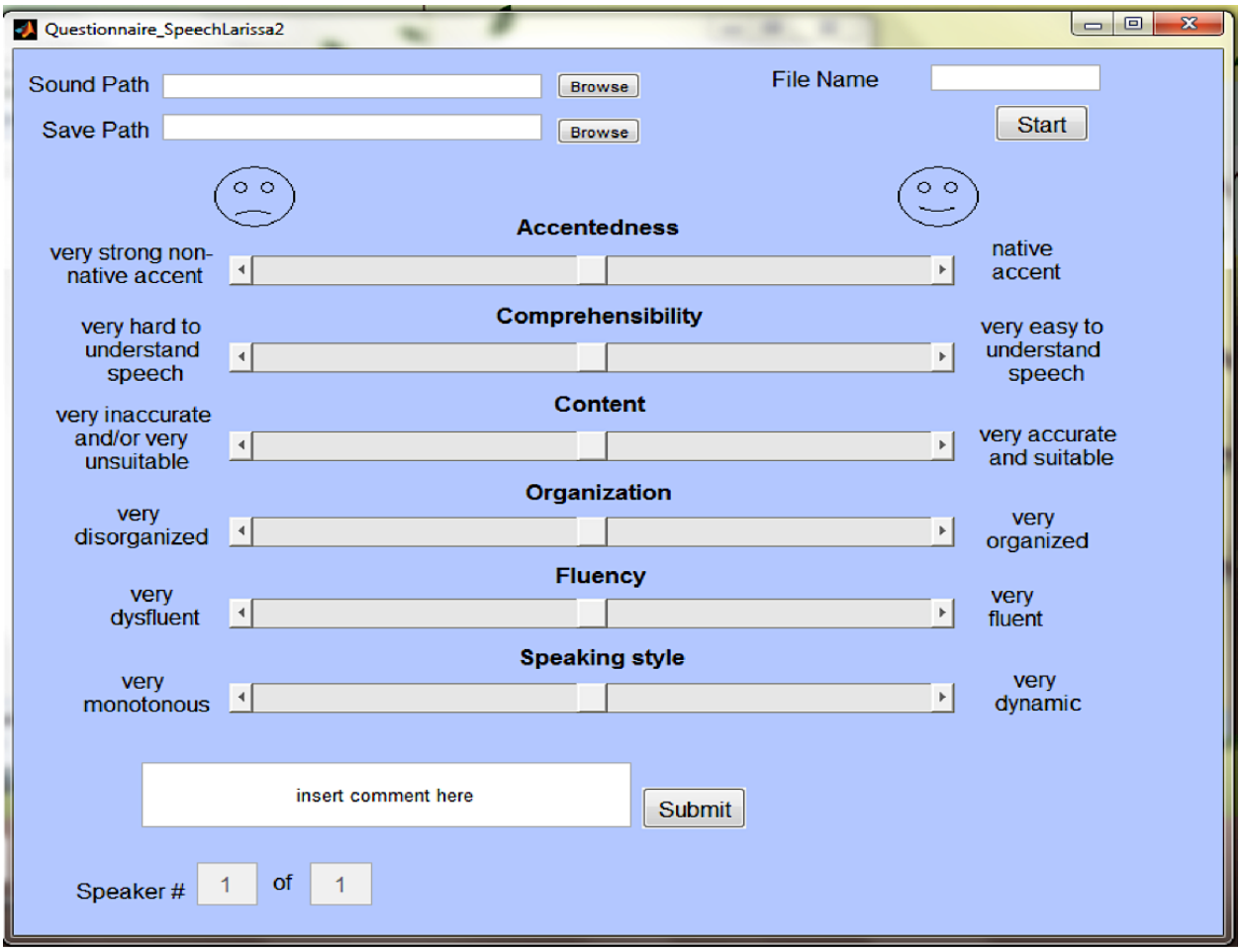

Figure 4. Computer interface used for listener ratings.

Rating sessions

The ratings were conducted in one-on-one sessions in a quiet room using a laptop PC and headphones. Before the sessions, the listeners were asked to fill out a background questionnaire. At the beginning of each session, they received written instructions with information about the speech samples, the rating scales, and the rating procedure. The instruction sheet also contained a table with brief descriptions of each rating category, as well as instructions on how to use the interface (see the Appendix C). The listeners were asked to write comments about reasons for

their ratings or anything they found noticeable about the speaker or the sample. After the instructions had been read, the author or trained research assistant would review orally important information and ask the listeners to explain their understanding of each rating category, in order to clarify any potential misconceptions. The participants practiced the rating procedure with two samples from presentations not included in the study and discussed their ratings with the researcher. The 22 samples within each set were automatically randomized by MATLAB and could be heard only once. After a sample finished playing, a "Next" button would appear on the screen to allow the listener to move on to the next sample. The rating sessions lasted approximately 100 minutes, and the participants were able to take a short break between the two sets of 22 samples. Cronbach's alpha reliability coefficients were high for all constructs and ranged from 0.85-0.98.

Analysis

The quantitative data in the study (i.e., the ratings given by the listeners) were analyzed using a three-way repeated measures ANOVA with *time* (Time 1 to 4), *construct* (accentedness, comprehensibility, etc.), and *listener* (specialists and non-specialists) as factors. The data were then plotted as line graphs of mean ratings over time for both listener groups together, then for specialists and non-specialists separately. This was followed by a speaker-by-speaker analysis in which each speaker's ratings were plotted separately and compared to others in order to identify potential patterns of change. As a final step, Pearson correlations were conducted between the ratings and potentially relevant data from the speakers' language background questionnaires: age of learning, age of arrival, amount of English use, and amount of prior ESL instruction.

The qualitative data in the study (i.e., the listeners' written comments) were subjected to a three-phase thematic analysis (Gibson & Brown, 2009) conducted by the first author, which is detailed below. The first phase involved reading all comments written by each rater in order to gain a general sense of the data. Then, the comments were coded using the six constructs (accentedness, etc.) as a priori codes, and they were also subclassified as either "positive" or "negative". In most cases, this coding was fairly straightforward. First, because the raters were encouraged to explicitly indicate the construct they were referring to in their comments (e.g., by preceding it with an abbreviation such as "Cont." for content). Second, because they often used

the names of the constructs in their comments, stating, for example, that the speaker was "accented" or "comprehensible". When this was not the case, four main strategies were used in order to identify the construct referred to: (1) searching for key words that were present in the guidelines or synonyms (e.g., "dynamic" and "monotonous" for speaking style), (2) examining other entries by the same rater to see if the same comment was linked more explicitly to a construct elsewhere, (3) examining the construct to which other raters linked similar comments, and (4) looking at the rater's quantitative ratings and trying to match them to the comments. The few comments that still could not be unambiguously assigned to a construct were coded as "other". In the third phase of the analysis, the data organized by construct were read multiple times in order to generate empirical codes representing the criteria used by the raters when evaluating the speakers on each of the six constructs or just overall (e.g., "speech rate" may be a criterion for fluency). Sixteen percent of the already coded dataset was randomly selected and that coded data underwent verification by a PhD student with research experience in L2 pronunciation. The verifier read written instructions which contained the explanations of the constructs in Appendix C and a description of the three-phase thematic analysis. He agreed with 98.7%, so no changes were made to the original coding. Finally, a tabulation was made of the number of comments and listeners that referred to each code and the number of times each code was used in a positive and in a negative comment. Differences in the evaluation criteria used by the two groups of listeners were identified by looking at the tabulation.

Results

Quantitative Ratings

Overall analysis. The results of the ANOVA revealed that there was no three-way interaction, no interactions between *time* and *construct*, and no main effect for *time*, which indicates that the talkers' ratings did not change significantly over time for any construct and for either group of raters. In other words, there appeared to be no significant changes in the talkers' oral performance from Time 1 to Time 4. Furthermore, no interaction was found between *construct* and *group*, which suggests that the two groups of listeners did not rate the constructs differently from one another when disregarding time. However, a significant interaction was

found between *time* and *group*, $F(3,30) = 6.466$, $p = .002$, as well as significant main effects for *construct*, $F(5,50) = 20.870$, $p = .000$, and *group*, $F(1,10) = 6.363$, $p = .030$. This indicates that both specialists and non-specialists differentiated among the six constructs, but there were significant differences in the way specialists and non-specialists rated the constructs over time.

Figure 5 displays the mean ratings given by non-specialists and specialists side by side. The ratings provided using the continuous scales were saved by the computer interface as numbers between 0 and 1 and were converted to percentages (0-100%) for ease of interpretation. For both groups, content and organization were rated significantly higher than the speech-related constructs (i.e., accentedness, fluency, comprehensibility, and speaking style), but there was no significant difference between them. Accentedness had the lowest ratings (where lower represents stronger non-native accent), while comprehensibility had the highest ratings among the speech-related constructs. Fluency and speaking style were in the middle and were not significantly different from each other for either group. As the chart shows, the two groups' ratings followed a similar pattern of change up until Time 4, when non-specialists' ratings go up and specialists' ratings go down. At Time 4, specialists were apparently harsher than non-specialists in their ratings. Post hoc tests using the Bonferroni correction revealed that this difference was significant for organization ($p = .038$), fluency ($p = .014$), and speaking style ($p = .010$). There were also significant differences between the two groups' ratings of speaking style at Time 1 ($p = .027$) and organization at Time 3 ($p = .023$). The means for these cases are shown in Figure 5.

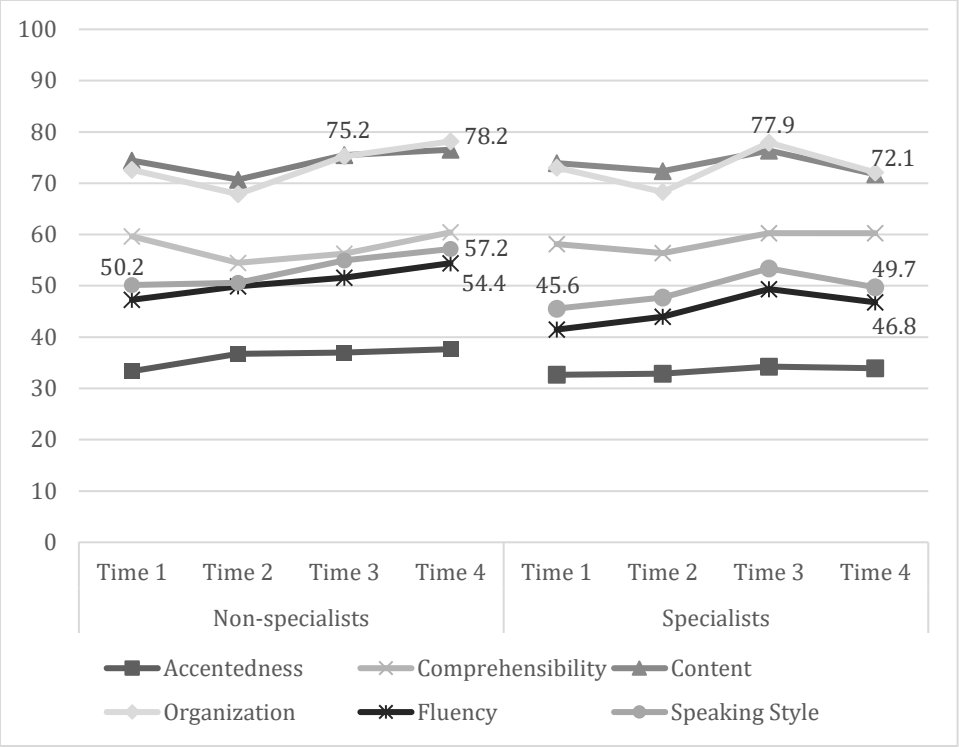

Figure 5. Mean ratings by non-specialists and specialists over time.

Case-by-case analysis. A follow-up speaker-by-speaker analysis suggested that, in some cases, change did seem to occur, but the lumping of all speakers together obscured different trends. A speaker's constructs that were already rated highly at Time 1 (73% or more) were found to remain roughly the same or show a downward trend. Such high ratings were more frequently observed for organization (eight cases), content (seven cases), and comprehensibility (five cases), and high Time 1 ratings for speech-related constructs occurred only with advanced speakers. Constructs rated less highly at Time 1 tended to either remain roughly the same (most cases) or display an upward trend in ratings.[5] These patterns are illustrated in Figure 6, which shows the average ratings given by both listener groups together to the least and the most proficient speakers, as indicated by their TOEFL scores and self-reported English-speaking

[5] The only exceptions were Participant 1's comprehensibility and fluency and Participant 4's accentedness, which were not rated highly at Time 1 and yet displayed downward trends.

skills. Participant 10 rated his English-speaking skills a 3 out of 9 and had a TOEFL score of 83, whereas Participant 6 rated his English-speaking skills an 8 out of 9 and had an IELTS score of 8.0. As can be seen from the chart, Participant 10 showed an upward trend in his ratings of organization, content, comprehensibility, and speaking style, but almost no change in fluency and accentedness. As for Participant 6, only accentedness (rated at 58.9% at Time 1) showed an upward trend, while the other constructs (initially rated at 73% or higher) remained approximately the same or showed a downward trend.

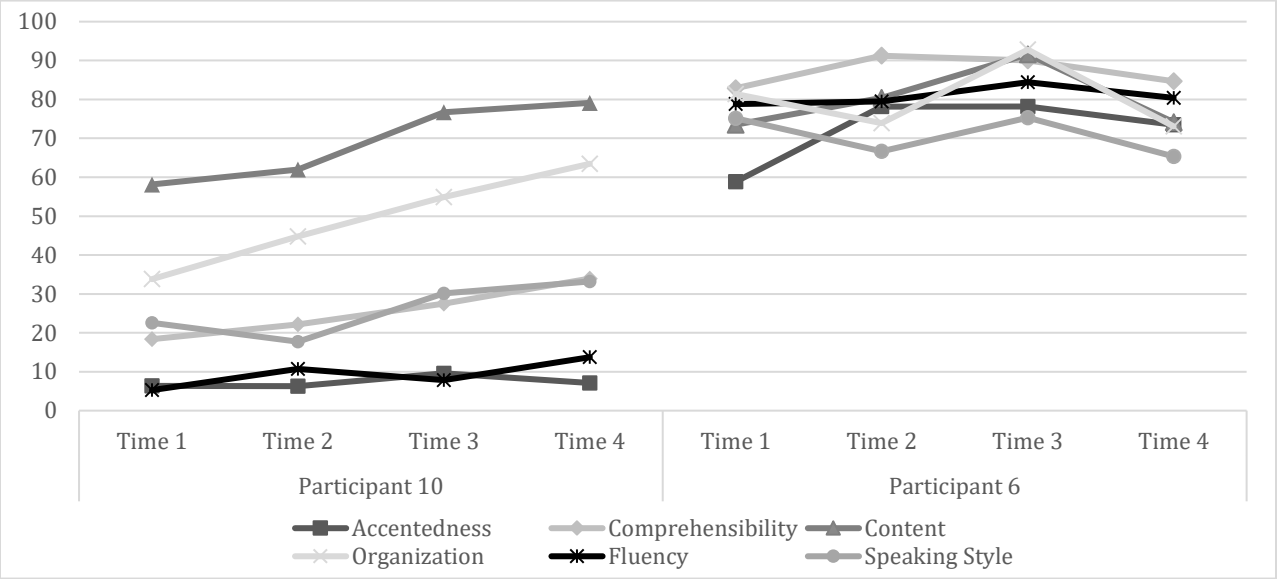

Figure 6. Mean ratings given to the least and the most proficient participants, respectively.

Figure 7 shows the constructs that displayed an upward trend, suggesting that improvement over time may have occurred. An increase of 10% or more in the trend line from Time 1 to Time 4 was considered an upward trend. Each chart line corresponds to a different speaker participant (e.g., P6 stands for Participant 6). Out of 17 cases of mean ratings exhibiting upward trends (i.e., 17 lines in Figure 7), 10 were given to the upper-intermediate speakers and seven were given to the advanced speakers. Fluency and speaking style were found to display upward trends more frequently than other constructs (four times each). While there were cases of gradual increase over time, change was usually non-linear, and there was often a decrease in ratings at Time 4. Participant 10 and 11, both with upper-intermediate speakers, experienced the

most improvement. They had the highest number of constructs with upward trends (four each) and steep trend lines.

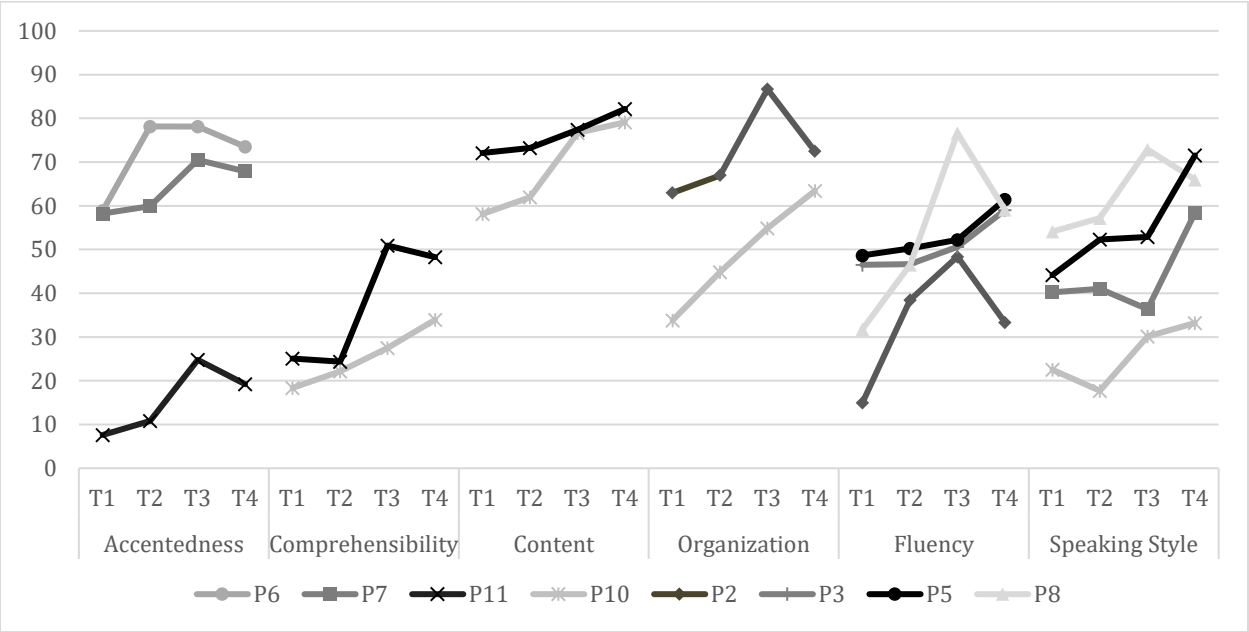

Figure 7. Constructs showing upward trends over time.

Effects of age of learning, age of arrival, amount of English use, and amount of prior instruction. The Pearson correlations revealed that the speakers' ratings were not significantly correlated with age of arrival (AOA), amount of English use (reported as hours per week), or amount of English instruction before coming to Canada (reported as years and hours per week). However, age of learning (AOL) onset in their home countries did correlate negatively with accentedness at Time 3 (r = -.622, p = .041) and Time 4 (r = -.621, p = .042), with comprehensibility at Time 2 (r = -.623, p = .041), Time 3 (r = -.691, p = .018), and Time 4 (r = -.612, p = 045), and with fluency at Time 2 (r = -.650, p = .030) and Time 3 (r = -.604, p = .049). This suggests that, with the exception of fluency at Time 4, the younger the participants were when they began studying English in their home countries, the better their speaking was perceived to be a few months into their programs.

Raters' Comments

The comments written by the non-specialists came to 12,963 words, and those written by the specialists amounted to 15,406 words, for a total of 28,369 words of data. As can be seen from Table 5, the speech-related constructs which were the most and the least frequently addressed were comprehensibility and accentedness, respectively. When addressing such constructs, there was not a considerable difference between the average number of words used by specialists and non-specialists. However, on average, specialists wrote considerably more than non-specialists on content and organization: 31.1% and 59.5% more, respectively. Specialists also wrote more on these two constructs than on any other construct.

Table 5

Total and Average Words Written on Each Construct by Rater Group

Construct	Total words (Mean per rater)		Mean difference
	N-S	S	
Accentedness	1,447 (96.5)	1,576 (92.7)	3.8
Comprehensibility	2,701 (180.1)	2,972 (174.8)	5.3
Fluency	2,009 (133.9)	2,303 (135.5)	1.6
Speaking style	1,654 (110.3)	1,731 (101.8)	8.5
Content	2,645 (176.3)	3,930 (231.2)	54.9
Organization	1,683 (112.2)	3,042 (178.9)	66.7

Note: "N-S" stands for *non-specialists* and "S" stands for *specialists*.

The following subsections will present the results of the qualitative analysis by construct, with a focus on the raters' main assessment criteria and differences between the comments made by specialists and non-specialists. In the interest of concision and focus, only criteria mentioned by more than five raters will be reported.

Accentedness. For both specialists and non-specialists, most of the comments about accentedness mentioned how strong or light the speaker's foreign accent was, often with short phrases such as "strong accent", "very little accent", "good accent", etc. Several comments

mentioned that a speaker's accent would waver throughout the presentation, usually becoming more accented as they went on. In some comments, the raters tried to guess where they thought the speaker's accent was from and identified specific pronunciation problems. Table 6 displays the number of comments and raters mentioning each of the main evaluation criteria for accentedness. The comments are organized by type (negative or positive) and group (non-specialists or specialists).

The main difference observed between specialists and non-specialists was the number of comments stating that the accent was *not strong*. The 15 non-specialists wrote a total of 119 comments describing the accent heard as *strong* ("thick", "heavy", "obvious", "highly accented", etc.), while there were only 56 comments, written by 11 non-specialists, describing it as *not strong* ("slight", "little", "moderate", "not very strong", etc.). Specialists, on the other hand, wrote a similar number of comments of each type: 105 comments by 16 specialists mentioning that the accent was *strong* and 91 comments by 15 specialists describing it as *not strong*. Furthermore, specialists pointed out mispronunciations of words, including key words and domain-specific terms, more often than the non-specialists.

Table 6

Frequency of Coded Criteria for Accentedness

Aspect	Type	Comments (Raters)		Examples
		N-S	S	
Strong	(-)	**119 (15)**	**105 (16)**	"The accent was pretty thick" (R15)
Not strong	(+)	**56 (11)**	**91 (15)**	"Her accent is not that strong" (R1)
Mispronunciation	(-)	7 (5)	21 (7)	"A: isn't it Cerburus pronounced Sir-Bur-
of words	(+)	0 (0)	2 (1)	Us, not Cer-Bur-Us?" (R16)
	Total	**7 (5)**	**23 (7)**	
Specific problems	(-)	7 (3)	6 (4)	"Very accented with rolled r's" (R28)
	(+)	2 (1)	0 (0)	
	Total	**9 (3)**	**6 (4)**	

Note: "A" is an abbreviation for "accentedness". Illustrative comments are followed by the rater number in brackets. Type of comment is negative (-) or positive (+).

Comprehensibility. In their comments about comprehensibility, the raters usually stated how much they understood from the speech and how easy or difficult it was to understand. Many of them provided reasons for their difficulty or ease in understanding the speech, which are displayed in Table 7. As can be seen, the most common criterion mentioned by the raters was the presence, nature, or strength of the speaker's accent. Quality of enunciation was the second most common reason given by the raters. This category refers to how distinctly the speakers articulated their words, including general comments about whether the speech was produced clearly or not and references to muffled speech, soft speech, slurring, and mumbling. Other reasons, provided by a smaller number of raters, were fluency, speed, and mispronunciations. One negative comment about the speaker speaking "too slowly" was included under fluency. The category "pace of speech" refers to negative comments about speech that was rushed or too fast and comments that mentioned slow speech as positive.

Even though accentedness was the most frequently mentioned aspect to affect comprehensibility, there were several comments that indirectly acknowledged the independence between these two constructs. As many as 78 comments by 22 raters (36 comments by 11 non-specialists and 42 comments by 11 specialists) expressed that the speaker's foreign accent, even when strong, was fully or mostly comprehensible: "Understood mostly everything even though her accent is strong" (R4), "despite accent, very easy to understand" (R6), "good comprehensibility even with accent" (R18). In terms of differences between the two groups of raters, accent, fluency, and especially mispronunciations were more frequently mentioned by specialists, and some of their comments pointed out key words from the presentation that the speaker did not pronounce correctly and that interfered with comprehensibility.

Table 7

Frequency of Coded Criteria for Comprehensibility

Aspect	Type	Comments (Raters)		Examples
		N-S	S	
Accent	(-)	29 (7)	58 (14)	"Accent makes some words hard to understand" (R6)
	(+)	2 (1)	0 (0)	"very easy to understand since she does not have a
	Total	**31 (7)**	**58 (14)**	heavy accent" (R4)
Enunciation	(-)	22 (8)	18 (9)	"hard to understand because he was not enunciating"
	(+)	8 (6)	2 (2)	(R27) "Lots of enunciation made the speaker a lot
	Total	**30 (11)**	**20 (10)**	easier to comprehend" (R9)
Fluency	(-)	5 (3)	12 (5)	"poor fluency led to low comprehensibility" (R25)
	(+)	0 (0)	3 (2)	"Speaker was easily understood despite strong accent
	Total	**5 (3)**	**15 (6)**	because they (…) were fluent" (R29)
Pace of speech	(-)	4 (4)	6 (4)	"The speaker speaks fast –making it hard to
	(+)	0 (0)	2 (2)	understand" (R3)
	Total	**4 (4)**	**8 (5)**	"slow speech allows for more comprehension" (R25)
Mispronunciation	(-)	**3 (1)**	**12 (6)**	"Comp: not pronouncing some words correctly" (R21)

Note: "Comp" is an abbreviation for "comprehensibility". Illustrative comments are followed by the rater number in brackets. Type of comment is negative (-) or positive (+).

Fluency. As expected, based on the guidelines given to the raters (see Appendix C), pausing was the main aspect considered by raters when judging fluency (see Table 8). Pauses, also referred to as "breaks", "stops", "interruptions", and "disruptions", were mentioned by the raters as problematic, especially when they were frequent, long, or used in inappropriate places. Their absence or limited presence was considered a sign of fluency. The second most frequently mentioned aspect was the use of fillers, which are sounds and words often used when speakers are thinking of what to say next (e.g., "uh", "um", "OK", "like"). They are considered "filled pauses", thus being directly related to the first aspect. Other aspects mentioned by the listeners are shown in Table 4. When comparing the two groups, it is possible to notice that, overall, non-specialists referred to hesitations more often, whereas specialists wrote more frequently about the speaker's use of fillers and self-repetitions.

Table 8

Frequency of Coded Criteria for Fluency

Aspect	Type	Comments (Raters)		Examples
		N-S	S	
Pausing	(-)	79 (11)	109 (16)	"Some awkward pauses" (R2)
	(+)	13 (6)	8 (6)	"F: not too many inappropriate pauses (R5)"
	Total	**92 (12)**	**117 (16)**	
Use of fillers	(-)	19 (7)	87 (13)	"F: there are a lot of "uuhh"-s in her speech" (R12)
	(+)	1 (1)	5 (4)	"Flu: not too many umms and awws" (R22)
	Total	**20 (7)**	**92 (14)**	
Hesitations	(-)	62 (10)	35 (4)	"Fluent except for a few hesitations" (R6)
	(+)	8 (3)	1 (1)	"There were no hesitations" (R15)
	Total	**70 (10)**	**36 (4)**	
Struggling to speak	(-)	**10 (7)**	**19 (7)**	"struggles with every word" (R13)
Speed	(-)	9 (5)	9 (5)	"Slow speech lead to a lower fluency score" (R27)
	(+)	3 (2)	3 (2)	"was fluent, thus fast" (R12)
	Total	**12 (5)**	**12 (7)**	
Stumbling	(-)	7 (4)	12 (5)	"Flu: stumbling on words" (R21)
	(+)	1 (1)	0 (0)	"doesn't stumble, showing confidence" (R13)
	Total	**8 (5)**	**12 (5)**	
Stuttering	(-)	8 (4)	29 (3)	"stutters repetitively" (R18)
	(+)	1 (1)	3 (3)	"He is very fluent and doesn't stutter throughout the
	Total	**9 (4)**	**32 (5)**	speech" (R4)
Self-repetition	(-)	**2 (2)**	**16 (4)**	"Flu: repeating words" (R21)

Note: "F" and "Flu" are abbreviations for "fluency". Illustrative comments are followed by the rater number in brackets. Type of comment is negative (-) or positive (+).

Speaking Style. Following the guidelines provided, most raters from both groups described the speaking style of a speaker by the adjectives "dynamic", "monotone", "interesting", "boring", and synonyms. Many would also justify their ratings based on whether they heard a large or small amount of changes in pitch. One aspect that was not included in the description of the construct in Appendix C but was mentioned by a small number of raters as

negatively affecting speaking style was a slow speech rate (4 comments by 2 non-specialists; 4 comments by 4 specialists). No noticeable differences were observed in how the two groups of raters interpreted and judged speaking style.

Organization. The aspect that was most frequently mentioned in the comments about organization was whether it sounded like the speaker had planned the presentation well as opposed to making things up on the spot (see Table 9). The second most frequently mentioned aspect, which is closely related to the description in Appendix C, referred to the connection or "flow" between ideas. Speakers who presented their ideas in a random or disconnected way were judged less favorably than those who established links and a nice flow between parts, particularly between the introduction and the rest of the presentation. The next most common aspect mentioned by the raters was the presence and quality of an introduction, which could be an overview of the talk, a statement of the topic or problem, a real-life example, or other forms of opening. Speakers who jumped directly into the details of the subject were considered less organized than those who provided some kind of introduction. In examining differences between the groups of raters, the first thing that stood out was the fact that the specialists wrote approximately 59.5% more than the non-specialists on organization (see Table 5). Moreover, specialists placed considerably greater importance than non-specialists on the presence of an introduction and the speaker's ability to stay on track during the presentation. They also wrote more than twice as many comments about establishing connections between ideas.

Content. The raters were asked to evaluate content for accuracy and suitability (see Appendix C). The vast majority of the comments about content provided reasons for the suitability or, more frequently, for the unsuitability of the content, which are listed in Table 10 below. The most recurrent reason (mentioned by 21 raters) was content that was too advanced or too difficult for first-year undergraduate students, often requiring background knowledge that they were not likely to have. The next most frequent criteria, mentioned by 15 raters, were the language or terminology used in the presentation and the explanations provided or omitted by the speakers. Language was considered unsuitable particularly when technical terms and acronyms were used without definitions. Further aspects mentioned by the raters are shown in Table 10. A noticeable difference between the two groups is that a considerably larger number of specialists commented on advanced level, introduction, and language/terminology.

Table 9

Frequency of Coded Criteria for Organization

Aspect	Type	Comments (Raters)		Examples
		N-S	S	
Planning	(-)	12 (8)	13 (5)	"Sounded somewhat improvised" (R28)
	(+)	10 (7)	14 (6)	"org: Obviously has a plan" (R16)
	Total	**22 (9)**	**27 (8)**	
Connection	(-)	4 (4)	23 (9)	"org – ideas are not well tied together" (R19)
between ideas	(+)	14 (5)	21 (4)	"Org: good flow from example to topic" (R21)
	Total	**18 (7)**	**44 (9)**	
Introduction	(-)	2 (2)	21 (5)	"His talk had not introduction" (R17)
	(+)	5 (3)	41 (6)	"org: Leading with a good example indicated
	Total	**7 (4)**	**62 (8)**	organization" (R16)"
Explanations	(-)	2 (2)	5 (4)	"poorly organized, little explanation" (R19)
	(+)	3 (2)	4 (2)	"[organization] well explained" (R10)
	Total	**5 (4)**	**9 (5)**	
Staying on	(-)	0 (0)	13 (7)	"Org: has a plan but deviated a bit" (R21)
track	(+)	1 (1)	2 (1)	"Maybe not pre-organized but still stays on track"
	Total	**1 (1)**	**15 (7)**	(R13)
Repetition	(-)	3 (2)	7 (6)	"org: went over the same thing twice" (R16)
	(+)	0 (0)	2 (2)	"She repeats her main point to drive home the
	Total	**3 (2)**	**9 (6)**	message (+organization)" (R30)
Dysfluencies	(-)	**6 (5)**	**9 (3)**	"Seems disorganized due to pauses" (R13)
Logical order	(-)	1 (1)	7 (4)	"Organization seemed a little 'backwards'" (R31)
	(+)	1 (1)	2 (2)	"There was a logical progression of the
	Total	**2 (2)**	**9 (5)**	presentation" (R31)

Note: "Org" is an abbreviation for "organization". Illustrative comments are followed by the rater number in brackets. Type of comment is negative (-) or positive (+).

Table 10

Frequency of Coded Criteria for Content Suitability

Aspect	Type	Comments (Raters)		Examples
		N-S	S	
Advanced level	(-)	**26 (6)**	**84 (15)**	"cont: quite advanced for 1st year I think" (R16)
Language /	(-)	14 (5)	26 (10)	"some unclear terminology (content)" (R28)
terminology	(+)	0 (0)	2 (1)	Language (…) is appropriate for the
	Total	**14 (5)**	**28 (10)**	audience" (R32)
Explanations	(-)	5 (4)	9 (6)	"Content: could have explained materials
	(+)	10 (6)	3 (3)	better" (R21)
	Total	**15 (7)**	**12 (7)**	"good explanation of content" (R18)
Introduction	(-)	2 (2)	7 (3)	"need more 'why' and context" (R25)
	(+)	9 (4)	20 (7)	"Content: well introduced for audience" (R21)
	Total	**11 (4)**	**27 (9)**	
Confidence /	(-)	7 (4)	3 (3)	"Content seems poorly known (…)" (R13)
Knowledge	(+)	17 (6)	7 (4)	"The speaker seems to know what he is talking
	Total	**24 (7)**	**4 (6)**	about but I don't" (R3)
Relevance /	(-)	3 (3)	5 (2)	"cont: 1st years might get this, but I'm not sure
Interestingness	(+)	5 (4)	7 (3)	how interesting it would be to them" (R16)
	Total	**8 (7)**	**12 (4)**	"Content; it is relevant and basic content" (R12)
Basic level	(-)	**12 (4)**	**5 (4)**	"Seemed too basic for Uni" (R8)
Use of examples	(-)	1 (1)	0 (0)	"not engaging with examples" (R13)
	(+)	3 (2)	8 (4)	"She gives some basic examples to clarify her
	Total	**4 (2)**	**8 (4)**	abstract idea (+content)" (R26)

Note: "Cont" is an abbreviation for "content". Illustrative comments are followed by the rater number in brackets. Type of comment is negative (-), positive (+), or unsure (?).

Besides writing, on average, 31.1% more about content than the non-specialists, unsurprisingly, the specialist raters also addressed accuracy much more frequently (92 mentions of accuracy by 13 specialists against only 30 mentions by eight non-specialists). Most comments (and all comments made by non-specialists) were positive, stating that the content was accurate.

There were only 10 negative comments, by four specialists, and 22 comments in which seven specialists said that they were unsure or that they could not comment on accuracy. Usually, the raters would simply state how accurate the content was or seemed ("inaccurate", "fairly accurate", or, most often, "accurate"). The only five comments that provided further information on accuracy were made by three specialists (R19, R23, and R26) and pointed out specific inaccuracies in the samples.

Other comments. Comments that could not be unambiguously assigned to one of the six constructs above were coded as "Other". The most common themes found among these comments were: **English skills** (23 comments by 5 non-specialists; 29 comments by 5 specialists): "Great English" (R7), "Poor mastery of language though" (R13); **nervousness** (22 comments by 6 non-specialists; 4 comments by 3 specialists): "Seemed extremely nervous" (R8), "she sounded calm" (R7); **reading** (7 comments by 4 specialists; 6 comments by 3 non-specialists): "suspect she is reading from script" (R18), "Sounded like she was reading" (R27). As can be seen, non-specialists commented more on whether the speaker seemed nervous or at ease.

Discussion

The first research question inquired whether evaluations of L2 graduate students' oral performance improved over time in the absence of explicit instruction and which aspects showed the most improvement: accentedness, comprehensibility, content, organization, fluency, or speaking style. The statistical analysis suggested that the L2 graduate students' evaluations did not improve significantly on any of the six aspects during their first six months of study in an English-medium university. However, a speaker-by-speaker analysis revealed different patterns of change caused by differences in baseline performance at Time 1 and possibly other aspects. Construct ratings which were very high at Time 1 remained the same or showed a downward trend over time, and this occurred more often with advanced-level participants. Lower initial ratings usually remained approximately the same, but sometimes displayed upward trends. Fluency and speaking style had the highest number of upward trends, and the participants who experienced the most improvement in their ratings were Participants 10 and 11, both upper-intermediate speakers. Yet there were two upper-intermediate speakers who showed no upward

changes in their ratings (Participants 1 and 9) and two advanced speakers who had two speech-related constructs that showed upward trends (Participants 7 and 8).

Previous research has indeed reported greater gains for lower-proficiency learners than for advanced learners after a stay abroad (Freed, 1995; Lapkin, Hart, & Swain, 1995; Llanes & Muñoz, 2009), often due to a ceiling effect, since advanced learners tend to have higher scores to begin with. However, assuming there is room for improvement in some of their language skills, more advanced learners may have reached a plateau, and significant improvement would require focused instruction. In this study, although there were two different proficiency levels, there were no beginner or intermediate-level speakers. All of the participants were already proficient in English, as they were pursuing graduate degrees at an English-medium university, which may partially account for the overall lack of statistically significant change in ratings over time.

Another important factor, which likely contributes to different outcomes for participants at the same proficiency level, is quantity and quality of input. Even though most participants had similar LORs, some may have had more opportunities to communicate in English and better-quality input than others. As in Flege and Liu (2001), there was no relationship between the ratings received by the speakers and their reported frequencies of English use at the beginning of the data collection. It is possible that these frequencies were wrong (see Bernard, Killworth, Kronenfeld, & Sailer, 1984) or that they changed over the six-month period. However, assuming that they were correct and did not change significantly, some speakers may have had more native-speaker input than others (Flege & Liu, 2001). Input quality may also account for the little overall change observed, given that the speakers were studying in a city and in programs with very high rates of non-native English speakers, among both professors and peers.

The different patterns of change could also have been influenced by individual differences among the participants. In addition to amount of L2 use, this study looked at AOA in Canada, AOL, and amount of English instruction prior to arriving in Canada. AOA and amount of English instruction did not correlate with any of the ratings. In the literature, an early AOA has been associated with better L2 pronunciation (Flege & Fletcher, 1992; Flege, Munro, & Mackay, 1995), and some more recent studies have found age effects even in adulthood (W. Baker, 2010; Derwing & Munro, 2013; Trofimovich & Baker, 2006). If age effects in L2 learning truly continue to exist in adulthood, the lack of AOA effects in this study may be caused

by the very small sample size. Alternatively, it could be because at Time 4 most participants had been in Canada for only about 7 months, and perhaps they had not had sufficient opportunity to benefit from greater language exposure and use. Another issue, and possibly the most relevant, is that AOA did not coincide with AOL for our sample, and the latter may be a more important factor. Most previous studies have used AOA as an index of AOL, as their samples involved learners who arrived in the L2 country with little or no knowledge of the target language, often as immigrants (Piske, MacKay, et al., 2001). The participants in this study, however, had achieved upper-intermediate or advanced proficiency while still in their countries of origin, as is the case of many international students. AOL was found to correlate with accentedness, comprehensibility, and fluency ratings at Times 2-4. This suggests that, even when AOL does not coincide with arrival in an L2 setting and consists simply of the onset of formal English instruction in a non-English-speaking environment, it may still contribute to better L2 pronunciation after immersion. On the other hand, amount of English instruction prior to arriving in Canada was not related to any of the ratings, which is somewhat surprising, but similar to what Derwing and Munro (2013) found. This may be due to the sample size of this study or to common inaccuracies in self-reported retrospective data (Bernard et al., 1984). Arguably, data such as AOA and AOL are more likely to be precise than estimates of English use or years of English study. Besides AOL, other individual differences may have also played a role in the results, such as language learning aptitude (DeKeyser, 2000) and willingness to communicate (Derwing & Munro, 2013; Macintyre, Dörnyei, Clément, & Noels, 1998), but these were not measured.

Finallly, shortcomings of the listening and rating tasks might have played a role in the results. First, it could be that the talkers' speaking became better (less accented, more comprehensible, etc.) the further they got into the presentation, but the one-minute excerpts were too short for the listeners to perceive any improvement. Similarly, the full extent of the content and organization of the presentations could not be rated through the excerpts. It should also be considered that the presentations were not rated in a real-life communicative situation, but as part of an isolated research task, and that the listeners did not have access to visuals that the speakers relied on to get their message across (e.g., PowerPoint and whiteboard).

In response to the second part of the first research question, the constructs that showed the most improvement were fluency and speaking style. The fact that accentedness showed less change is not surprising, since there is some evidence to suggest that accent stabilizes early (Flege, 1988), and the participants were already fairly proficient at Time 1. In fact, almost all late-onset L2 speakers, even highly proficient ones, speak with a foreign accent (Ioup, 2008), and this does not necessarily affect their intelligibility or comprehensibility (Munro & Derwing, 1995). Thus, accentedness seems to be more resistant to naturalistic improvement than other features (see Derwing & Munro, 2013), and the participants would probably need focused instruction in order to notice specific discrepancies between their production and a more target-like version and make significant changes (Derwing & Munro, 2005; Schmidt, 1990). While comprehensibility has high communicative value and is probably more prone to improve without instruction (Derwing & Munro, 2013), several participants already had very high comprehensibility ratings at Time 1, so there was not as much room for change. The same was true for content and organization. Initial ratings of fluency and speaking style, on the other hand, were not as high, so more change was expected. Study-abroad research has indeed found that immersion can have positive effects on fluency (Freed, 1995; Llanes & Muñoz, 2009; Segalowitz & Freed, 2004). As for speaking style, there is evidence that speakers produce a narrower pitch range in their L2 than in their L1, probably because they lack confidence and need to concentrate more on other aspects of speech (Mennen, 1998; Pickering, 2004; Zimmerer et al., 2014). It is possible that, as the speakers in this study gained more experience with the L2 and with academic presentations, they became more confident and started to produce more pitch changes.

Strangely, however, improvement over time was found to be non-linear in most cases, and there was often a decrease in ratings at Time 4. It is difficult to explain this pattern, although we suspect that it is due to external circumstances or a task effect. Time 4 coincided with the end of the winter term, when the speakers were probably busy with school work and final exams, so they might not have prepared as much for their presentations. Lack of preparation may also have resulted from the speakers' heightened confidence in their abilities, familiarity with the task, or decreased motivation, reflecting in their ratings. It is worth noting that lower ratings at Time 4 were given mostly by specialist listeners, which brings us to the second research question: How

do evaluations of oral performance differ when given by content specialists and non-specialists? Which aspects (accentedness, comprehensibility, etc.) are most valued by the specialists?

Overall, the quantitative ratings of the two listener groups (specialists and non-specialists) were found to follow a similar pattern of change up to Time 4, when the specialist ratings of fluency, speaking style, and organization became significantly harsher than the non-specialist ratings. It is possible that the task effect mentioned above was picked up by specialists, but not by non-specialists. In other words, the fact that the speakers prepared less at Time 4 (assuming this was the case) may have been more evident to specialists because they value organization more and have a clearer schema of what constitutes an organized presentation. In fact, the qualitative analysis showed that specialists wrote 59.5% more on organization (the largest difference between the groups) and that they commented more on the presence of an introduction, whether the presenter stayed on track, and whether they established connections between ideas, which are aspects that would likely distinguish a well-prepared presentation. Less planning may have also contributed to poorer fluency and speaking style, since the speakers' cognitive resources became overloaded with what they had to say next and they paid less attention to their speech (see Ellis, 2009; Foster & Skehan, 1996). Alternatively, a lower organization rating may have created a negative "halo effect", which led listeners to rate fluency and speaking style (which came right below organization in the computer interface – Figure 4) less favorably as well. These hypotheses would need to be verified by a detailed analysis which objectively measured and compared features of organization, fluency, and speaking style (e.g., introduction, pauses, pitch variations, etc.) in the presentations over time.

Besides organization, specialists also wrote more about content than non-specialists, which suggests that they placed greater value in it or that they felt more competent to address it than the novice raters. Content and organization were also highly valued in expert evaluations reported in the ESP literature (Douglas & Myers, 2000; Elder et al., 2012; Jacoby, 1998; Knoch, 2014). In their comments, specialists focused more on essential features of academic presentations, such as an introduction or motivation to the talk and clear connections between its parts. They also commented more frequently on the level of difficulty or technicality of a presentation. This reflects their higher awareness of the characteristics of the genre under evaluation and its communicative goals (Elder, 1993; Elder et al., 2017; Jacoby, 1998) and a

greater concern for the accessibility of the talk to the audience. Content specialists in the ESP literature have indeed been found to value aspects related to the professional's relationship, empathy, or interaction with their clients (Douglas & Myers, 2000), patients (Pill, 2016; Read & Wette, 2009), or students (Elder, 1993). Interestingly, whether the speaker was nervous or appeared to be confident or knowledgeable was more important to non-specialists than to specialists, which may indicate that specialists are better able to see through smaller performance issues and focus on the essence of the presentation.

In terms of how specialists and non-specialists evaluated the speech-related constructs, little insight was gained from the comments on speaking style. The other comments revealed key similarities and differences between the groups. Among the similarities were their awareness of the independence between accentedness and comprehensibility, as was found in Kennedy, Foote, and Buss (2015), and main evaluation criteria which were equally valued by them: quality of enunciation for comprehensibility and pausing for fluency. As for the differences, specialists commented more on accent, fluency, and mispronunciations as affecting comprehensibility, sometimes mentioning specific key words that had been mispronounced by the speakers. Mispronunciations were also mentioned more frequently in the specialists' comments on accentedness. When evaluating fluency, they addressed the use of fillers and self-repetition more frequently, whereas non-specialists referred generically to "hesitations" more often. Despite the fact that the two groups wrote approximately the same average number of words on comprehensibility and fluency, specialists provided more comments in which they stated their evaluation criteria for these constructs, and their criteria appeared to be more specific and varied than those used by the non-specialists. Perhaps greater familiarity with the subject and genre under evaluation allowed them to go into more detail even when commenting on speech-related constructs, because they did not need to expend as many attentional resources to follow the presentations. On the other hand, this could simply reflect a difference in the raters' level of maturity and education, given that the specialists were all graduate students, whereas the non-specialists were undergraduate students.

Pedagogical Implications and Limitations

When combined with the literature, the findings of this study suggest that once adult learners achieve a threshold of L2 competence in their countries of origin, naturalistic change in an L2 setting may be more difficult and dependant on several variables, such as quantity and quality of input, age of learning, and other individual differences. Yet some speakers may exhibit improvement, and fluency and speaking style (i.e., pitch variations) seem to be more prone to change than accentedness and comprehensibility. This is possibly because, at the upper-intermediate or advanced level, some learners' L2 accents may have stabilized, and their comprehensibility may already be quite high.

Overall, it appears that the participants in this study and learners like them would need specific instruction in order to show substantial changes in their oral performance, particularly in accentedness, which could sometimes interfere with comprehensibility. Subject-matter specialists seemed to think that the speakers' L2 accent affected their comprehensibility more often than non-specialists, and some of them pointed out mispronunciations of key words in the presentations as problematic. ESL or EAP teachers could help international graduate students by working on high functional load errors in their speech (Munro & Derwing, 2006), by addressing problematic words through awareness-raising activities, perception exercises, and repetition, for example (see Szpyra-Kozłowska & Stasiak, 2010), and by teaching them strategies in order to learn the correct pronunciation of important words in their fields of study, such as checking the phonetic transcription in a dictionary. Working on reducing pausing and fillers could also help them sound more fluent more quickly than they would without instruction. Importantly, instructors should be mindful of the specificities of the genres their students need to master and their communicative goals. When teaching or evaluating academic presentations, the findings of this study suggest that a focus on organization would be helpful. Students should be taught culturally- and language-appropriate ways of providing a good introduction or motivation to their presentations, establishing clear connections between its parts, and staying on track. They should also be made aware of the ways in which lack of planning could affect their oral performance. Finally, they should be able to simplify their language and explain technical terms when appropriate for their audience.

The findings and pedagogical implications discussed above should be interpreted in light of the limitations of this study, which include the small sample size and the fact that two of the talker participants were not in programs related to engineering and computer science. When it comes to the listeners, the specialists were graduate students from different levels (Masters and PhD) and different years, so their level of expertise in their fields certainly varied. Also, they are probably better representations of the perceptions of the speakers' peers than of faculty members, professionals, or experienced researchers. Another limitation to consider is that the raters differed in more than just subject-matter knowledge, given that the non-specialists were all undergraduate students. Despite these shortcomings, we believe that this study makes a unique contribution to our understanding of important issues related to adult L2 acquisition and oral performance in academic presentations.

Connecting Studies 1 and 2 to Study 3

The findings of Studies 1 and 2 combined suggest that naturalistic improvement in the speech of L2 graduate students within their first six months of residence and study in an L2 context may be more evident to language experts than to other listeners, including both subject-matter specialists and non-specialists. Although subject-matter specialists seemed to pay more attention to the content and organization of the presentations than non-specialists, they were not any more lenient than the latter in their pronunciation ratings, and they were particularly bothered by mispronunciations of words in the presentations. These results pointed to the need for explicit pronunciation instruction to foster improvement in the speech of L2 graduate students that is perceptible to subject-matter specialists and non-specialists. Unfortunately, however, ESL and EAP teachers may sometimes be reluctant to teach speaking because they lack the confidence and skills to do so (e.g., Baker, 2011; Macdonald, 2002). By fostering positive cognition changes, teacher training may help to bridge this gap between research findings in L2 pronunciation teaching and learning and actual classroom practices. Study 3 investigates the unique contribution of teacher training in pronunciation pedagogy to pre-service teacher cognition related to L2 pronunciation. Specifically, I hoped to determine if training in the field might promote favorable views of explicit pronunciation teaching and increase student teachers' self-efficacy when it comes to teaching pronunciation, thus increasing the chances that they will include pronunciation instruction when teaching ESL.

Chapter 4: Study 3

The Role of Training in Shaping Pre-Service Teacher Cognition Related to L2 Pronunciation

This manuscript has been published in:
Buss, L. (2017). The role of training in shaping pre-service teacher cognition related to L2 pronunciation. *Ilha Do Desterro*, *70*(3), 201–226.

Abstract

This study analyzed changes in the pronunciation-related cognitions of undergraduate TESL students from two Canadian universities during their first term in their programs. The students from one university attended a 13-week course in phonology and pronunciation teaching, while those from the other university received no specific training in pronunciation. Towards the end of the term, the participants who received specialized training had more favorable views of explicit pronunciation teaching and became more confident in their ability to teach pronunciation than the comparison group. The course also helped the participants increase in awareness of their own speech and limitations. The findings point to the importance of native-nonnative speaker interactions in shaping cognitions and suggest that more support to nonnative-speaking TESL students and more training in the use of communicative activities would be helpful.

Introduction

Several researchers have identified a need among ESL teachers for more specialized training in pronunciation pedagogy (Derwing & Munro, 2005; Foote, Holtby, & Derwing, 2011). However, the effects of this type of training on pre-service and in-service teachers' beliefs and practices have not been sufficiently explored. Thus, we still lack knowledge or insight into what constitutes effective teacher training in L2 pronunciation. The present study addressed this gap in the literature by investigating changes in the cognitions of pre-service ESL teachers after taking an undergraduate course in phonology and pronunciation teaching. Teacher cognition (TC) is here defined as the "cognitive dimension of teaching – what teachers know, believe, and think" (Borg, 2003, p. 81). This study focuses specifically on teachers' *beliefs*, as there is compelling evidence to suggest that what teachers believe can affect their behavior in class and learner outcomes. Self-efficacy, for example, which is defined as teachers' belief in their ability to influence students' performance (Bandura, 1977; Tschannen-Moran, Hoy, & Hoy, 1998), has been found to positively correlate with willingness to try out different practices (Ghaith & Yaghi, 1997; Smylie, 1988) and with learning outcomes (Ashton & Webb, 1986).

L2 Teacher Cognition

According to Borg (2006), many factors can influence TC, including personal experiences, professional coursework, and classroom practice. The importance of these aspects has been underscored by several TC studies involving pre-service and in-service L2 teachers. Teachers having learned the language as nonnative speakers, for example, is one factor that has been found to potentially influence their TC. The vast majority of nonnative-speaking teachers surveyed by Reves and Medgyes (1994) reported that they had language difficulties, especially with vocabulary and speaking, and that they believed these difficulties affected their teaching to some extent. Similar results were found by Samimy and Brutt-Griffler (1999), who analyzed classroom discussions, interviews, and autobiographical accounts by 17 students (all NNSs) enrolled in a graduate-level Teaching English to Speakers of Other Languages (TESOL) seminar in the United States. The participants reported having language difficulties in several areas, and most of them believed that these difficulties hindered their teaching.

Another factor that can influence TC is professional coursework. Mattheoudakis (2007) administered the Beliefs about Language Learning Inventory (BALLI) (Horwitz, 1985) questionnaire to a group of 36 pre-service English as a Foreign Language (EFL) teachers at three time points over the course of a teacher education program in Greece. She observed several changes in the participants' responses, including weaker beliefs in the primacy of vocabulary and grammar teaching and in the importance of correct pronunciation, possibly resulting from the communicative training they received. Similarly, MacDonald, Badger, and White (2001) reported changes in the beliefs of 55 TESOL students in Scotland after a semester of courses in SLA. These changes were reflected in the participants' responses to a questionnaire on language learning based on Lightbown and Spada (1995, p. xv). After studying SLA, many of the participants' initial common-sense beliefs became more informed by research, while no significant changes were observed in the beliefs of a comparison group of 25 undergraduates in EFL and Initial Teacher Education programs that did not include instruction in SLA.

Finally, classroom practice can change TC (Borg, 2006). Evidence of this was found by Bateman (2008), who observed that 10 student teachers of Spanish as a second language experienced changes in their beliefs about target-language use in the classroom after a semester of practicum. As revealed by their responses to a pre/post questionnaire, the students lowered their expectations regarding the use of the target language to give instructions for assignments and projects and to provide cultural information. The above-mentioned study by Mattheoudakis (2007) also analyzed the beliefs of a second group of 30 student teachers from the same program in Greece who had chosen to do a teaching practicum in their final year of studies. The beliefs of this group (called "practice group") were assessed only at the end of the program, i.e., Time 3. Compared to the first group of 36 student teachers (the "non-practice group"), they seemed to be more critical of their EFL teachers and the teaching they had received as language learners. At the same time, several of their beliefs were traditional and more in tune with their grammar-based learning experience than with the communicative training program they had completed. According to Mattheoudakis, the classroom reality might have caused them to re-evaluate the theoretical knowledge acquired during the program.

Clearly, TC is a complex phenomenon affected by several factors. Research on this topic can support pre-service and in-service teacher education by investigating how training may shape

TC, possibly helping to change misinformed, incomplete, or unconstructive cognitions. However, very little research in the field of L2 pronunciation has explored this issue.

Teacher Cognition in L2 Pronunciation

Almost all known studies on TC related to pronunciation have employed cross-sectional designs to explore the current beliefs and practices of in-service and often experienced ESL teachers who were not undergoing training. These studies have explored teachers' level of training in pronunciation pedagogy, activities and techniques used to teach pronunciation, and the most frequently taught pronunciation features, as well as teachers' beliefs about pronunciation teaching and learning and about their own ability to teach pronunciation. Research in ESL contexts includes surveys of teachers and program coordinators conducted in the UK by Burgess and Spencer (2000), in Australia by Burns (2006), and in Canada by Breitkreutz, Derwing, and Rossiter (2001) and Foote, Holtby, and Derwing (2011), as well as interview studies conducted by Macdonald (2002) in Australia and by Baker (2011b) in the United States. More survey and interview research has been done with teachers from EFL settings, including Greece (Sifakis & Sougari, 2005), Northern Cyprus (Hismanoglu & Hismanoglu, 2010), Finland (Tergujeff, 2012a, 2012b), Brazil (Albini & Kluge, 2011; Buss, 2016), Uruguay (Couper, 2016), and several others (Henderson, 2013; Jenkins, 2005; Timmis, 2002). A common finding in many of these studies is that ESL/EFL teachers lack confidence in teaching pronunciation and would like more training in the field.

Buss (2013) adopted a different method from the above studies in that it sought to understand the pronunciation-related cognitions of pre-service EFL teachers in Brazil through an analysis of their teaching practicum reports. However, given that the reports did not deal exclusively with pronunciation, a relatively small amount of data was obtained. More importantly, the student teachers had received training in phonology and pronunciation pedagogy years before their practicum, and there was no pre-measure of their cognitions or control group for comparison. Therefore, potential links between training and TC related to pronunciation were not clear-cut.

Teacher Training and Teacher Cognition in L2 Pronunciation

To date, few studies in L2 pronunciation have researched the influence of training on TC. This is an important topic of investigation because specialized training in the field has often been called for as an answer to teachers' self-reported lack of knowledge or confidence to teach pronunciation (Couper, 2016; Foote et al., 2011; Macdonald, 2002). Nonetheless, how and to what extent training can actually promote positive changes to these teachers' cognitions about pronunciation is not fully understood. Baker (2011a, 2014) indirectly addressed this issue in her investigation of the cognitions and classroom practices of five experienced ESL teachers. The earlier study (2011a) sought to determine the source of these teachers' knowledge of pronunciation pedagogy. When asked during interviews, three of the five teachers identified an MA-level course in pronunciation pedagogy as having the greatest impact on their teaching of pronunciation. Classroom observations confirmed this finding, as many of the techniques and activities that they reported learning in the graduate course were used in their classes. The other two teachers, who had not taken a graduate course specifically devoted to pronunciation pedagogy, identified other factors as playing a prominent role. The only one of the five who had little or no training in pronunciation pedagogy also mentioned that she felt insecure teaching pronunciation. In Baker (2014), the same three teachers who had completed a graduate-level course in pronunciation pedagogy were found to use a larger range of pronunciation teaching techniques than the other two. Still, most of the techniques used were controlled ones, which are less communicative and more carefully constrained by the instructor. Although interesting, Baker's conclusions were drawn long after training, so the development of TC throughout the course, as well as more specific links between the kind of training received and cognition changes, could not be fully explored.

Burri (2015a, 2015b) and Burri, Baker, and Chen (2017) used a longitudinal case study design in order to look more directly at the role played by teacher education in shaping TC. They explored the cognition development of 15 ESL student teachers, including NSs and NNSs of English, during a 13-week graduate course in pronunciation pedagogy. Data were collected from interviews, focus groups, questionnaires, and classroom observations. Some changes were observed in the cognitions of the participants, especially a heightened awareness of the importance of suprasegmentals and of NNSs' ability to teach pronunciation (Burri, 2015b). It

was found that the cognition development of NNSs was stimulated by an increased awareness of their English speech and the perception that their own pronunciation improved with the course. NSs, on the other hand, benefited more from their interactions with NNS classmates, which helped them realize that NNSs could be effective pronunciation teachers. In Burri (2015a), the same participants were reported to have become more aware of nonnative English varieties and accents, which in turn led them to believe that the goal of pronunciation teaching should not be accent elimination. A later analysis by Burri et al. (2017) indicated that the participants became more aware of the benefits of kinesthetic/tactile teaching techniques. However, they still favored controlled activities – the kind they had experienced as L2 learners – when writing an end-of-course assignment in which they provided recommendations on how to address certain pronunciation problems.

Burri's research is the first to investigate the longitudinal development of pronunciation TC in a teacher education program. However, several of its participants had years of teaching experience and some had previous training in TESL or related areas. It is necessary to determine whether similar outcomes would be observed for student teachers in undergraduate programs who have little or no teaching experience. Considering many ESL teachers working in schools do not hold a graduate degree, it is important to know whether pronunciation training at the undergraduate level, given to less experienced or less educated student teachers, can also inspire positive changes in TC. Furthermore, a weakness of Burri's study is that it does not attempt to sort out the influence of specific pronunciation training from that of other TESL courses on student teachers' cognitions. In other words, the pronunciation-related cognition development observed by Burri may have also been influenced by other courses in the TESL program, and not only by the specific course in pronunciation pedagogy. Thus, it is helpful to have a control group of TESL students who are not undergoing pronunciation training, so as to compare their cognitions over time to those of the treatment participants.

The Current Study

In this longitudinal study, the gaps identified in the literature review were addressed by exploring the developing cognitions of pre-service ESL teachers in an undergraduate university course in phonology and pronunciation teaching. The course consisted of a taught portion on

phonology and pronunciation teaching and a short teaching practicum at the end, where student teachers prepared and taught pronunciation tutorials to ESL learners. The participants' cognition changes were compared to those of a control group of undergraduate pre-service TESL teachers who had received no specific training in phonology or pronunciation pedagogy.

The main research questions addressed by the study were the following:

1. How do TESL student teachers' cognitions about pronunciation change after the taught portion of an undergraduate course in phonology and pronunciation teaching, and how are the identified changes related to the course content and activities?
2. How do these changes compare to those experienced by TESL student teachers who do not receive any specific training in phonology and pronunciation teaching?
3. What further cognition changes (if any) take place after the pronunciation teaching practicum, and how are these changes linked to the course or to the practicum?

Method

Participants

The treatment group participants in the study were first-year undergraduate TESL students at a Canadian university who were enrolled in a phonology and pronunciation teaching course offered in the fall term. The same instructor taught two groups: one had classes in the afternoon twice a week (1h15 each) and the other had classes in the evening once a week (2h15). Both groups included students in a four-year Bachelor of Education program and students in a one-year TESL certificate program. Comparable TESL students from another university who received no instruction on phonology and pronunciation teaching in their first year were recruited to participate as a comparison group. Both NSs and NNSs of English were included in the sample.

In total, there were 18 students in the treatment group and 15 students in the comparison group. Most of the treatment participants (13) were from the afternoon group, with only five being from the evening group. Part of the comparison data was collected in the fall of one year (six participants). The rest were collected the following fall (nine participants), along with the treatment group data. The 15 comparison participants were all from the same university. No

changes were made to the curriculum of the program from the first to the second year of data collection. Table 11 shows the TESL courses that were taken by the participants. All of the participants took a grammar course and almost all of them received some instruction in general L2 teaching, but only the treatment group took a course in English phonology and pronunciation teaching.

Table 11

Treatment and Comparison Groups' TESL Courses

	Treatment Group		Comparison Group
Courses	BEd	Certificate	BEd
English phonology and/or pronunciation	✓ (3 credits)	✓ (3 credits)	✗
English grammar	✓ (3 credits)	✓ (3 credits)	✓ (3 credits)
General L2 teaching	✓ TESL pedagogy (6 credits):* observation and assistance in ESL classes, examination of L2 teaching approaches and methodologies, and discussion of relevant issues to the ESL classroom	✓ TESL methodology (3 credits): review of theory in applied linguistics directly related to L2 teaching and discussion of teaching methods and techniques	✓ L2 education (3 credits): focus on L2 teaching methods and approaches from a historical and analytical perspective ✓ Classroom observation (2 credits): "participant observer" field experience at a school ✓ Professional seminar (1 credit): how to observe L2 classrooms and reflection on the classroom observation experience

*This course was not required in the first term, so it was not taken by all participants from the BEd program.

Of all participants (n = 33), most were females (n = 29) in their 20s (n = 19) from the Greater Montreal area (n = 18). The majority considered themselves to be native speakers of English (n = 25), although some reported having more than one native language, with English and French being the most common combination (n = 9). With few exceptions, the participants had no prior training in TESL or pronunciation pedagogy and little or no experience teaching ESL at the start of the study (see Table 12). The treatment group differed from the comparison group in that the treatment group had males and a few mature students. Four of the treatment participants were in their 30s and three were over 50 years old. There were also two participants who had extensive TESL experience in their non-English-speaking home countries: one had taught for seven and a half years and the other for more than 10 years.

Table 12

Participant Demographics

	Treatment	Comparison
N	18	15
Sex		
Male	4	0
Female	14	15
Age (yrs)		
Median	24.5	20
Range	18-63	18-23
Birthplace		
Greater Montreal	10	8
Other Quebec	2	1
Other Canada	1	0
Abroad	5	6
Native language(s)		
English only	7	4
English and other(s)	6	8
Other(s)	5	3

TESL or pronunciation pedagogy training*		
No	15	13
Yes	2	0
ESL teaching experience		
None	11	13
Less than 6 months	5	0
1 year or more	0	2
7 years or more	2	0
L2 pronunciation teaching experience		
No	16	14
Yes	2	1

*No answer given by three participants.

The subset of participants who were interviewed represented different age groups and birthplaces and included both NSs and NNSs of English (see Table 13). It was representative of the sample in that it had mostly female participants with little or no experience teaching ESL.

Table 13

Interviewed Treatment Participants

Name*	Age	Gender	Birthplace	Declared English-speaking status	ESL teaching experience
Margaret	63	F	USA	Native	None
Amanda	29	F	Peru	Nonnative	None
Kassim	21	M	Canada (Montreal)	Native	None
Christine	21	F	Canada (Quebec)	Nonnative	None
Parvin	33	F	Iran	Nonnative	10+ years
Jennifer	24	F	Canada (Montreal)	Native	None
Leslie	25	F	Canada (Montreal)	Native	None
Lynn	52	F	Netherlands	Native	3 months

*Names have been changed to preserve confidentiality.

Data Collection

Phase 1: Questionnaires. There were two phases to the study. The first phase used a web-based questionnaire hosted on Survey Monkey to explore students' cognitions before and after the theoretical portion of the TESL course in phonology and pronunciation teaching. Both treatment and comparison groups participated in the first phase. Recruitment for the pre-treatment questionnaire took place on the second week of class (mid-September – Time 1), per the two TESL instructors' requests, and the participants were given approximately two weeks to complete the questionnaire. The post-questionnaire was completed in mid-November (Time 2), before the treatment group taught their first pronunciation tutorial.

The questionnaires were fairly long (76 items) and covered several topics, most of which were related to the participants' background, beliefs about pronunciation teaching and learning, and beliefs about their self-efficacy as pronunciation teachers. The majority of the questionnaire items were worded as statements with which the participants indicated their level of agreement on a 7-point scale (from "0" = strongly disagree to "6" = strongly agree). Some of the items were written by the researcher drawing on topics from the literature on pronunciation and TC, others were taken from the instrument developed by Foote et al. (2011), and yet others were adapted from the Self-efficacy Teaching and Knowledge Instrument for Science Teachers (SETAKIST), designed by Roberts and Henson (2000). The complete questionnaire given to NSs from the treatment group at Time 1 can be found in Appendix D. The main difference between the questionnaire version given to NSs and the one given to NNSs was that latter featured questions about their experiences learning ESL rather than other second languages. The version given to the comparison group was slightly different in that it did not refer specifically to the pronunciation tutorials taught as part of the pronunciation course, but to teaching pronunciation in general in the future. The only difference between the Time 1 and the Time 2 questionnaires was that the latter did not repeat certain questions background questions (e.g., gender, date of birth, native language, etc.).[6]

[6] Please contact the author to receive the other versions of the questionnaire.

The internal consistency of the questionnaire items designed to assess the construct of self-efficacy was good, with a Cronbach's alpha of .82 at pretest and .90 at posttest. The other constructs addressed in the questionnaire were not considered for reliability analyses because they were measured by too few items (usually just one or two) or by open-ended questions. The questionnaire response rates were approximately 26% for the pool of treatment group students and 20% for the pool of comparison group students. During Phase 1, the researcher also contacted the instructor responsible for the TESL course in phonology and pronunciation teaching and was given access to the Moodle sites for the course. Each of the two groups taught by the instructor (afternoon and evening) had their own Moodle site, which contained the materials used in class, as well as the course outline, homework, and assignment guidelines. The researcher was able to download all of these documents for analysis.

Phase 2: Interviews. The second phase of the study took place after the treatment participants finished their teaching practicum in early December and involved a smaller sample of eight participants from the treatment group. They volunteered to participate in the interviews following an email invitation sent by the researcher to the whole group. As a final project for their TESL course, the students in the treatment group were required to put together a pronunciation teaching portfolio, which included their individual lesson plans for the teaching practicum, descriptions of procedures, materials used, and reflections on the experience. Those who agreed to participate in the second phase of the study gave the researcher access to their portfolios and were interviewed individually, in English, a few days later. This interview was semi-structured and probed more deeply into cognition changes observed in the pre/post questionnaires and any further changes resulting from the practicum experience. The researcher asked general questions about what the participants thought of the teaching experience and whether they thought something had changed in their knowledge and beliefs related to pronunciation after it. All of them were asked whether they felt more or less confident in teaching pronunciation in the future. Specific questions were created based on the participants' answers to the questionnaires and based on what they had written in their pronunciation teaching portfolios.

Analysis

Materials. The analysis of the TESL course materials was based on the contents of the Moodle sites. These sites were well organized and constantly updated by the professor with PowerPoint slides and handouts used in class, activities and/or answer keys to activities done in class, extra resources, homework for upcoming classes, as well as the course outline and guidelines for assignments. The first step in the analysis was to read the course outlines and the dashboard pages of the Moodle sites in order to have an overview of the course, especially the topics covered, their distribution over the 13 weeks, and the types of activities used. The second step was to classify the resources according to the topics they addressed and their types – "PowerPoint files", "handouts", "activities", or "homework" – and count them. Then, the researcher looked for potential differences between the teaching received by the afternoon group and the evening group by comparing the dashboard pages of the two Moodle sites. An overall percentage of difference between the two sites was calculated.

Questionnaires. The analysis of the questionnaires focused on the findings related to the main topics of interest in this study: the participants' beliefs about pronunciation teaching and learning (addressed in Section II, no. 2, of the questionnaire in Appendix D) and beliefs about their self-efficacy as pronunciation teachers (addressed in Section III except letters "b", "e", and "f"). The first step in the analysis was to develop a demographic profile of the participants. Then, percentage tables were created to summarize the participants' responses to the 22 questionnaire items that addressed cognitions about pronunciation teaching and learning. For some of the statements, higher agreement (i.e., a higher response number) indicated higher self-efficacy, such as "I know what to do to teach pronunciation effectively". For others, such as "I feel anxious about having to teach English pronunciation", higher agreement suggested lower self-efficacy. Thus, some responses needed to be recoded so that higher numbers always reflected higher self-efficacy. Self-efficacy means were then calculated for each participant at Time 1 and Time 2 by averaging their responses to the 19 questionnaire items that addressed this construct. Finally, statistical analyses were conducted in order to compare the two groups and test for any significant changes from Time 1 and Time 2.

Interviews. The interviews were first transcribed and read by the researcher. Relevant data were organized using three *a priori* categories based on the research questions (Miles,

Huberman, & Saldaña, 2014): “TESL course influence on cognitions”, “Practicum influence on cognitions” and “Other”, which was reserved for any additional information on the participants’ cognitions which did not fit into the first two categories. The organized data were read again and coded for recurring topics or ideas which indicated cognition changes and experiences that were common across participants.

Pronunciation Teaching Course

During the course of the study, the treatment group received specialized training in the form of a 13-week undergraduate TESL course in phonology and pronunciation teaching. The course was taught by a PhD in applied linguistics who was a researcher in L2 pronunciation and had more than 10 years of experience teaching higher education. Because the instructor posted all of the course materials on Moodle, it was possible to have a fairly good overview of each class and the course as a whole. This section will provide a description of the TESL course, based on the analysis of the course materials. This description is needed in order to understand how the course and practicum may have influenced TC.

Goals and Contents

According to the outline, the course had four instructional goals:

- to provide students with an overview of the sound system of English organized around three main topics—sound-, word-, and phrase-level phonology;
- to demonstrate to students how theoretical information about the sound system of English can be applied to classroom teaching of pronunciation;
- to help students create materials for teaching pronunciation to learners of English;
- to provide students with an opportunity to practice teaching pronunciation to learners of English.

The required text for the course was Lane’s (2010) *Tips for Teaching Pronunciation*. A schedule of topics in the course outline indicated the pages or chapters of the book that the students were expected to read before each class. The first week of class was an introduction to the course and included a lecture on basic issues in the field and a group activity (see details in Table 14). After the first week, the course was organized into three four-week blocks, covering

segmentals, suprasegmentals, and pronunciation teaching. The first two blocks of the course were theoretical, and classes consisted mainly of lectures to introduce new topics followed by practice activities. Each of these blocks ended with an exam.

Table 14

Pronunciation Teaching Course Overview

Week	Content	Details
1	**Introduction**	- **Topics**: Definition, scope, and importance of phonology; the role of age and L1 in L2 pronunciation; the relationship between perception and production; the goals of pronunciation teaching - **Group activity**: students listened to recordings of L2 speakers, identified pronunciation "problems", and gave examples of how they could help the speakers
2-5	**Individual sounds**	- **Consonants**: how consonants are produced; the vocal tract; articulators; description of consonants (place of articulation, manner of articulation, and voicing); phonetic symbols - **Vowels**: description (height, backness, lip rounding, and tenseness); the schwa sound; phonetic symbols - **Phonemes and allophones**: definition and importance of contrastive sounds; minimal pairs; examples of positional variation (aspiration, flapping, glottal stop, dark and light /l/, /r/ coloring) - **Exam 1**
6-9	**Suprasegmental features**	- **Word stress**: definition; levels (major, minor, and no stress); full vs. reduced vowels - **Rhythm**: definition; properties (intensity, pitch height, and vowel duration); stress-timed vs. syllable-timed; importance; characteristics (alternation of peaks and valleys and regular timing); content vs. function words; natural speech phenomena - **Sentence stress**: definition; characteristics; functions; importance; placement rules

		- **Exam 2**
10-13	**Pronunciation teaching**	- **In class**: how to administer a diagnostic test; pronunciation teaching; how to write pronunciation activities; examples and analyses of pronunciation activities - **Outside class**: prepare and teach four pronunciation tutorial classes to an L2 speaker - Tutorial 1: diagnostic test - Tutorials 2-4: total 90 minutes of pronunciation teaching - **Pronunciation teaching workshop**: students gave a short in-class presentation about their tutorial experience. Suggested topics: most or least successful activity, what they learned about pronunciation teaching or English phonology, and their L2 learner's progress. **- Teaching portfolio** submission: copies of lesson plans and teaching materials; brief self-evaluations reflecting on their teaching; peer-evaluations reflecting on their partner's teaching

The final block of the course started with a demonstration by the instructor of how to administer a pronunciation diagnostic test. The following classes included short lectures and examples of pronunciation activities. Some of the guidelines for writing pronunciation activities and many of the sample activities were bound together in a course pack. During this block, the students were also required to prepare and teach four pronunciation tutorial classes to an L2 speaker. The TESL students were paired up by the professor and each pair was assigned one L2 learner. On their first meeting with the learner, they administered a diagnostic test and each TESL student chose a small number of pronunciation features (usually one or two) to teach for approximately 30 minutes during each of the next three tutorials (for a total of 90 minutes of

teaching time for each TESL student). Most of the preparation and all of the teaching took place outside of class time. This pronunciation teaching block ended with a pronunciation teaching workshop and with the submission of a teaching portfolio (see Table 14 for details).

Materials

The course materials were of four types: PowerPoint slides, handouts, activities and their answer keys, and homework. PowerPoint slides were used by the instructor to facilitate his lectures, which introduced new topics, usually at the beginning of the class. Handouts were either looked over in class or provided as extra resources for the students, including sound charts, diagrams, explanations, examples, sample activities or lesson plans, the course outline, and guidelines for evaluations. The activities were done in class and gave the students a chance to practice whatever was discussed in the slides. They included tasks such as identifying sounds, matching, transcribing, explaining, etc. The instructor would check the answers in class or post the answer key on Moodle. All activities but one involved only written words, sentences or texts, with no audio. There were, however, a few audio examples in the PowerPoint slides. Homework consisted of one to three exercises that were usually similar to the activities done in class. They gave the students extra practice on the topic discussed in the previous class and counted towards final grades.

The course materials were almost the same for the two TESL groups. The only differences were that the afternoon group did four additional activities in class (out of a total of 27 activities) and that each of the groups received two handouts that the other did not (out of a total of 41 handouts). Thus, out of a total of 84 resources posted on Moodle (not including answer keys of already posted activities), only eight differed between the groups, which is a 9.5% difference. Table 15 shows how many of the total number of resources (i.e., files) posted on Moodle address each course topic. Overall, the resources were evenly divided between the two theoretical blocks of the course (segments and suprasegmentals), except for the greater number of handouts on segments, which was due to the several sound charts and sets of phonetic symbols provided to the students. A large number of handouts were also given in the last block, pronunciation teaching, and consisted mostly of sample pronunciation activities. Compared to consonants and vowels, the students did not get as much practice on phonemes and allophones

(only one activity in class), as shown in Table 15. Word stress was the most practiced suprasegmental feature (six activities and three pieces of homework). As for pronunciation teaching, although there were almost no in-class activities (only one of the two groups did an activity where they analyzed and evaluated pronunciation activities), the students practiced choosing and writing pronunciation activities in preparation for their tutorials. They also had the opportunity to critique pronunciation activities (their own and their tutorial partner's) in their portfolios.

Table 15

Course Resources by Topic

	PPT files	**Handouts**	**Activities**	**Homework**
Total	**10**	**41**	**27**	**6**
Introduction and course guidelines	**1**	**10**	**1**	**0**
Segments	**4**	**9-10**	**11-13**	**3**
Consonants	1	7	6-8	2
Vowels	1	2-3	9-10	1
Phonemes and allophones	2	0	1	1
Suprasegmentals	**3**	**2-3**	**10-11**	**3**
Word stress	1	0	6	3
Rhythm	1	1	2	0
Natural speech phenomena	1	1	2	0
Sentence stress	1	0-1	3-4	1
Pronunciation teaching	**2**	**18**	**0-1**	**0**

Note. Main topics are shown in bold, with subtopics listed below. Some files addressed more than one subtopic (e.g., consonants and vowels), thus totals for the topics do not equal the sums of the subtopics. Ranges (e.g., 9-10 handouts on segments) mean that the number differed between the afternoon and the evening groups.

Results

Cognition Changes After Theory

This part of the results section will address changes in the participants' cognitions from the first questionnaire, which was given at the beginning of their first term in the TESL program, to the second questionnaire, given two months later. This coincides with the end of the theoretical blocks of the treatment group's phonology course, just before the start of the pronunciation teaching tutorials.

Changes in beliefs about pronunciation teaching and learning. Changes for each group over time and differences between groups on the questionnaire items that addressed cognitions about pronunciation teaching and learning were analyzed using Wilcoxon signed-rank tests and Mann-Whitney tests. Parametric statistics could not be used because the data violated the normality assumption. The data used were mean values of participants' responses. The complete results (raw values) for all of the items can be found in Appendix E. Table 16 below displays only those items for which significant differences within and/or between participant groups were found. For ease of comparison and due to space constraints, the 7-point responses were reduced to three: "disagree" (including responses 1-3: "strongly disagree", "disagree", and "tend to disagree"), "unsure or neutral" (response 4), and "agree" (including responses 5-6: "tend to agree", "agree", and "strongly agree"). As can be seen from the running of Wilcoxon signed-rank tests, a significant change over time was only found in the comparison group's response to item "s" ($Z = -2.233$, $p = 0.026$). That is, two months into their first term of studies, the comparison group agreed significantly less that ESL learners benefit from paying conscious attention to the input and becoming aware of how different features are produced. On the other hand, agreement with this statement increased for the treatment group; Mann-Whitney U tests showed that the two groups became significantly different from each other at Time 2 ($U = 70.5$, $p = 0.018$).

Despite the fact that most of the response changes over time were not significant, some of them resulted in between-group differences at Time 2. As was the case for item "s", the differences found for items "a", "m", and "o" suggest that, compared to the comparison group, at Time 2 the treatment group held significantly more favorable views of explicit pronunciation

teaching and learning. They believed significantly more than the comparison group that pronunciation teaching was not dispensable ($U = 61.5$, $p = 0.011$), that it was important to know phonology to teach pronunciation ($U = 58$, $p = 0.008$), and that it was not boring to teach it ($U = 48$, $p = 0.001$). Even at Time 1, the vast majority of treatment participants disagreed that teaching pronunciation was boring (94.4%), as opposed to only 66.7% of the comparison group, which was also a significant difference ($U = 53.5$, $p = 0.002$). The two groups' responses to item "k" were significantly different at Time 1 ($U = 80$, $p = 0.048$), but not at Time 2, when there was an increase in the percentage of comparison group participants who agreed that some individuals resist changing their pronunciation to maintain their identity (from 40% to 53%). The treatment group's agreement with this statement was high at both time points (72.2%).

The phonology course had an important role in shaping the treatment group's cognitions about pronunciation teaching and learning, as suggested by several comments made by the eight participants who were interviewed at the end of the term. The interviewees often made direct reference to the course and to the professor, Peter (not his real name), when explaining their views on a particular subject or when asked why certain answers had changed over time, as in the following examples:

"Um… but I think like Peter said at the end of the semester, he said, like, try to do just uh… little pronunciation things, like, try to just insert them in your lessons..." (Christine talking about teaching pronunciation in general ESL classes)

"Well, because I have learned a lot about how to teach the pronunciation, you know, developed a material, and uh… I thought that, well, OK, if a teacher can do that much, if a teacher can have all these resources and use them in a positive, useful way in a class, then teaching pronunciation is seriously important." (Parvin referring to questionnaire item "a" from Section II, no. 2)

"No, I mean, from a common-sense point of view I would have said no, it's not… I tend to think that it's not a good idea to teach to get rid of the accent completely. But now, from having done the course, of course it's not! (…) I know one thing Peter said, that sometimes a strong accent, even if you can understand, sometimes there's gonna be more prejudice." (Lynn referring to item "f" from Section II, no. 2)

Changes in self-efficacy. Nineteen of the questionnaire items were related to the construct of self-efficacy, addressing the participants' beliefs about their knowledge of the subject matter and their beliefs about their skills as pronunciation teachers. The response data met the assumptions for parametric tests, so a mixed ANOVA was conducted with time and group as factors. The output revealed significant main effects for time, $F(1,31) = 6.308$, $p = .017$, and group, $F(1,31) = 5.770$, $p = .022$, but no interaction effect. As shown in Figure 8, the treatment group started out with a higher self-efficacy mean (M = 3.81, SD = 0.81) than the comparison group and increased more over time (M = 4.15, SD = 0.73). The comparison group had a less marked increase in self-efficacy means from Time 1 (M = 3.40, SD = .50) to Time 2 (M = 3.51, SD = .57). The effect size of the change was greater for the treatment group, $d = -0.640$, than the comparison group, $d = -0.225$. The first effect size can be considered moderate, while the second one is small (Cohen, 1988).

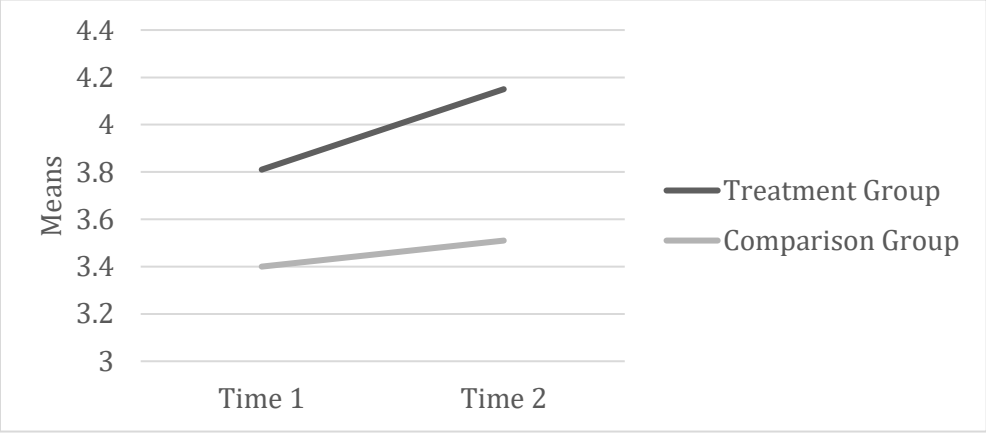

Figure 8. Self-efficacy changes from Time 1 to Time 2.

The interview data suggest that the phonology course also helped the participants become more aware of their limitations. For example, native English speaker Margaret confessed she initially felt "a bit arrogant" because she had been "blessed with well-spoken parents". With the course, she realized that was not enough and there was much more to teaching pronunciation than being able to model it properly:

> I became so aware that unless you just happen to be a good mimic, in which case your accent's probably going to be pretty good anyway, that somebody needs to teach you

what your mouth is doing, which most people are *completely* unaware of (…). And you don't have to be a native speaker to do that.

Table 16

Beliefs about Pronunciation Teaching and Learning with Significant Results

	Treatment Group								Comparison Group							
	Time 1				Time 2				Time 1				Time 2			
	D	U	A	M	D	U	A	M	D	U	A	M	D	U	A	M
a. Teaching pronunciation is often unnecessary, as most learners are able to pick up on pronunciation when frequently exposed to good language input.[a]	14 82.4	0	3 17.6	2	14 77.8	2 11.1	2 11.1	1	11 73.3	0	4 26.7	2	5 33.3	5 33.3	5 33.3	3
k. Some individuals resist changing their pronunciation in order to maintain their identity.[b]	3 16.7	2 11.1	13 72.2	5	1 5.6	4 22.2	13 72.2	4	3 20	6 40	6 40	3	3 20	4 26.7	8 53.3	4
m. It is important to know phonology (i.e., the rules of pronunciation) in order to teach English pronunciation.[a]	0	1 5.6	17 94.4	5	1 5.9	1 5.9	15 88.2	6	2 13.3	3 20	10 66.7	4	2 13.3	4 26.7	9 60	4
o. Teaching pronunciation is boring.[a b]	17 94.4	0	1 5.6	1	16 88.9	1 5.6	1 5.6	1	10 66.7	3 20	2 13.3	2	7 46.7	3 20	5 33.3	3
s. Learners benefit from paying conscious attention to the input and becoming aware of how different features are produced.[a c]	0	4 22.2	14 77.8	4	1 5.6	2 11.1	15 83.3	5	0	4 26.7	11 73.3	5	1 6.7	4 26.7	10 66.7	4

Note. D = Disagree; U = Unsure or neutral; A = Agree; M = Median

Content of unshaded cells = Count and percentage

[a] Significant difference between groups at Time 2 ($p < .05$)

[b] Significant difference between groups at Time 1 ($p < .05$)

[c] Significant difference from Time 1 to Time 2 for the comparison group ($p < .05$)

Leslie and Lynn, also NSs, expressed some concern that they were not always able to consciously identify pronunciation sounds and patterns. For example, they had a hard time identifying stress in an utterance, even though they could produce it. NNSs Amanda and Parvin gained awareness of foreign aspects of their own pronunciation. Amanda, whom the researcher thought had the most noticeable accent, was the only interviewed participant who believed the course made her less confident in her ability to teach pronunciation in the future. She explained that she became aware of everything she needs to improve in her own pronunciation. She also said during the interview that she wished the phonology course had been structured in two sessions and that NNSs had been given the chance to work on their own pronunciation problems first. The other two NNSs, Parvin and Christine, had generally native-like pronunciation, yet Parvin said the course made her realize that she had some "occasional slips", as she put it. She added, "I did become more conscious and more aware of maybe the slight differences in pronunciation that I have with the native speakers…" "You know, (the course) made me self-aware. It was total self-discovery."

Almost all of the interview participants reported feeling anxious or worried at some point during the course about having to teach the pronunciation tutorials. However, several of them also mentioned aspects of the course that helped prepare them for the tutorials and increase their confidence. Christine said it was "very beneficial" and "eye-opening" to have the professor explain to them what to do in a lesson, give them the course pack, and demonstrate the diagnostic test. Kassim also referred to the sample pronunciation activities given by the professor as "very helpful". For Leslie, it was learning the theory that helped her feel more confident: "I felt more confident (…) than I thought I would going into it, because of the course, because I've been shown that there were issues that could be caught and stuff like that". Similarly, Jennifer reported increasing self-efficacy as she went through the course:

> At the beginning of the class when he mentioned about, you know, the portfolio, oh my God, the stress, the nerves, I was like, "Ugh! How am I gonna do this? I don't know anything about this course". And so I was like, "You know, I think I can do it, but we'll see, we'll see". And then by the end I was like, "No, I got this" (laughs).

Further findings. It was not only what was taught in the course that influenced the participants' cognitions, however. For the NS participants, interactions with NNSs throughout the course also seemed to have an important role. Margaret reported being very positively impressed with the teaching skills of her tutorial partner Parvin, which served as confirmation for her that being a NS was not necessary to teach pronunciation. Leslie said that she noticed from the beginning of the term that her NNS classmates had learned more about English pronunciation in school, whereas she was learning it for the first time. This led her to have feelings of inferiority. Similarly, Kassim talked about a NNS classmate who impressed him at first with his technical knowledge of pronunciation, because he had already learned it in a classroom setting. Over time, however, Kassim noticed that this classmate had difficulty applying what he knew to his own pronunciation, which led him to believe that NSs are better suited to teach pronunciation. Lynn and Jennifer also talked about classmates whose pronunciation was difficult to understand and wondered how that would affect their teaching. Even so, at Time 2 Jennifer disagreed more that NSs are better suited to teach pronunciation. The fact that three out of four of her professors that term were NNSs caused this change, as she was very impressed with their level of English and the quality of their teaching.

Cognition Changes After Teaching

The third research question inquired about further cognition changes that may have taken place after the pronunciation teaching practicum (Time 3) and how they are connected to the course or the practicum. This question will be answered from the interview data exclusively.

Changes in beliefs about pronunciation teaching and learning. The cognitions of the interviewees about pronunciation teaching and learning did not seem to change much after the practicum. In most cases, they would mention experiences from the tutorials to back up opinions they had already expressed in the questionnaires. Yet most participants mentioned at least one change that took place during or after the tutorials. In response to the statement "English sounds are easier to teach than global aspects like stress, rhythm, and intonation", Amanda first agreed (Time 1), then disagreed (Time 2). When asked about this change in her interview, she mentioned that she agreed with the statement again after the tutorials (Time 3). Kassim said that before the tutorials he thought pronunciation was too complicated to be taught, but he realized

afterwards that it is very teachable. Christine was unsure or neutral at first, but after the tutorials she said she "tended to agree" that knowledge of phonology is important in order to teach pronunciation. She explained that it is easier to know the rules to be able to explain them to students. Furthermore, her response to the statement "Teaching pronunciation is boring", which had gone from "tend to agree" (Time 1) to "unsure or neutral" (Time 2), changed again at Time 3. She said teaching the tutorials was not boring, but interesting and enjoyable. Similarly, Parvin was surprised at how exciting it was to teach pronunciation, even though she did not find it particularly boring before. Finally, interacting with her student made Leslie realize how important it can be for ESL learners to sound native-like, something she was not aware of. As can be seen, the few cognition changes related to pronunciation teaching and learning that took place after the tutorials varied from participant to participant.

Changes in self-efficacy. The interviewed participants changed their answers to several questions related to self-efficacy after the practicum. In all but one case, their new answers revealed increased feelings of self-efficacy. Kassim went from agreeing with the statement "I find pronunciation a difficult topic to teach" at Times 1 and 2 to disagreeing with it completely. Likewise, Margaret said teaching pronunciation was not as difficult as she thought it would be before trying it. Margaret, Kassim, and Christine reported becoming more confident or less anxious about teaching pronunciation in the future after the experience:

"If I had to go teach right now I think I would feel more confident about teaching pronunciation (than grammar), just because I've done it now." (Margaret)

"I'm not afraid of having to teach pronunciation. In fact, I'm excited that that's something that I'm able to do." (Kassim)

"…if I had to do what we did for this class again, then I would probably feel more comfortable. Just because I've done it once." (Christine)

Like Margaret and Christine, Jennifer also felt more confident about teaching pronunciation in the future because it would not be her first time:

> (The tutorials) definitely helped in giving me confidence in giving some kind of pronunciation-related lesson in future classes. (…) I have the practice, if you will. I got the chance to, like, test the waters. (…) I'm always shy and anxious the first time I do something, so it, you know, broke the ice.

Five participants (Amanda, Kassim, Christine, Parvin, and Leslie) mentioned that they noticed improvement in their learners' speech throughout the tutorials, which indicated to them that they were doing well as pronunciation instructors. The only instance in which self-efficacy seemed to decrease because of the tutorial experience was Christine's response to the statement "I am afraid that students might 'catch' me making pronunciation mistakes when I teach", which had changed from "tend to disagree" to "unsure or neutral" from the first to the second questionnaire. In her interview, Christine mentioned that if she had to answer the question again she would say "tend to agree". She explained that she made a mistake in one of her tutorial lessons on vowels and that this made her realize mistakes could happen.

Further findings. The interview data indicated that the phonology course might not have focused enough on the use of communicative activities to teach pronunciation. During the last block of the course, the instructor did talk about the importance of moving from more controlled to less controlled pronunciation practice (following Celce-Murcia, Brinton, & Goodwin, 2010), and he also provided samples of guided and communicative activities. However, it seemed like the participants needed more examples and guidance on how to use these types of activities. Amanda observed that "when you look at the course pack, it's a lot of drillings and repetitions", which led her to believe that communicative activities were not that helpful for teaching pronunciation. Similarly, Parvin said that her preparation for the tutorials "was all about explicit phonology teaching" and not about meaningful communication.

The participants also mentioned a few challenges that they faced when trying to implement more communicative activities. Christine, Jennifer, and Lynn reported being surprised by their learners' lack of vocabulary. In addition, Lynn said she found it hard to get her learner to talk and felt she needed more teaching experience with communicative activities. Finally, pronunciation mistakes during authentic communication posed a challenge for Kassim and Lynn, as they did not know how to address them with their learners.

In sum, the findings suggested that general TESL training and specialized training in phonology and pronunciation pedagogy affect TC differently. While most of the participants' pronunciation-related cognitions did not change significantly over time, two months of specialized theoretical training led the treatment group to value explicit pronunciation teaching significantly more than the comparison group and was more effective at increasing the

participants' self-efficacy as pronunciation teachers. A short practicum at the end of the course further increased self-efficacy, but did not seem to generate considerable changes in other cognitions. Further findings pointed to the relevant role of NS-NNS interactions in shaping TC and to the need for more training on guided and communicative pronunciation teaching techniques.

Discussion

The first two research questions inquired about potential changes to pronunciation-related TC after the taught portion of an undergraduate course in phonology and pronunciation teaching, how these changes related to the course content and activities, and how they compared to those experienced by student teachers who did not receive specialized training in the area. The analysis of the questionnaire data revealed that the treatment group had significantly more favorable views of explicit pronunciation teaching and learning than the comparison group at Time 2. Moreover, the treatment group experienced a larger increase in self-efficacy from Time 1 to Time 2. The interview data indicated that the treatment participants' awareness of their limitations also increased because of the course.

The different kinds of instruction each group received may explain the cognition differences observed between the groups towards the end of the term. The treatment group had an entire course dedicated to phonology and pronunciation pedagogy, which emphasized the importance of explicit teaching and learning of pronunciation and provided them with guidance, techniques, and examples on how to teach pronunciation. The comparison group, on the other hand, only received general instruction on L2 education. The course they did on L2 education analyzed ESL teaching methods and approaches from a historical and analytical perspective, so there might have been an emphasis on communicative language teaching (CLT) as the most current approach, and it is possible that the students connected explicit teaching of pronunciation to past and limited approaches such as the audiolingual method. Lack of specific training may also be the reason the comparison group's self-efficacy as prospective pronunciation teachers did not improve as much. The interview data seemed to confirm the connection between the treatment participants' cognitions and the instruction received, as suggested by the many

references made to the phonology course and the professor. The cognitions of NS teachers were also influenced by their interactions with and observations of the NNSs in the course.

These results are generally in line with those of Baker (2011a), Burri (2015a, 2015b), and Burri et al. (2017), who also found that a specialized course in pronunciation pedagogy had an important role in shaping TC. However, some noticeable differences from the findings of Burri's study should be mentioned. While in Burri (2015b) NS teachers came to have more positive views of NNS teachers, that was not always the case in this study. It is possible that the NNSs in Burri's study, who were graduate students and more experienced ESL teachers, simply had higher proficiency levels or even higher confidence in speaking English than some of the students in the undergraduate course here studied. Most likely, however, Burri's participants became more open to nonnative models of pronunciation because one of the course objectives was to develop an appreciation of different English varieties and accents. The course included regular comparisons of accents and discussions on World Englishes (Burri, 2015a), something that was not present in the pronunciation course here analyzed.

Another difference is that Burri (2015b) suggested links between NNSs' cognition development and their sense of pronunciation improvement. Although none of the three NNSs interviewed in this study reported a sense of improvement, they still became more aware of their own speech and changed several of their pronunciation-related cognitions. Nonetheless, the only NNS who struggled with some aspects of English pronunciation did not experience an increase in self-efficacy and wished she had had the opportunity to work on her own pronunciation during the course. It is therefore possible that, for her and other NNSs like her, experiencing pronunciation improvement or at least having pedagogical attention given to their pronunciation may be necessary for confidence to increase.

The fact that the NNSs in Burri (2015b) believed they improved their pronunciation while those in this study did not may be attributed to the different class structures. In this study, most of the training in pronunciation teaching activities and techniques was received at the end of the course. In Burri's study, each of the instructor's weekly lectures was followed by a training session in a pronunciation teaching technique. This may have given NNSs a chance not only to become aware of the pronunciation features covered in the lectures, but also to practice them by applying the teaching techniques learned throughout the course. Future research should

address how courses in pronunciation pedagogy can provide better support to the growing number of ESL/EFL teachers who are NNSs, given that language difficulties negatively affect these teachers' self-efficacy (Reves & Medgyes, 1994; Samimy & Brutt-Griffler, 1999) and that insecurity about the quality of their own pronunciation might make them reluctant to teach pronunciation (Murphy, 2014).

The third research question was "What further cognition changes (if any) take place after the pronunciation teaching practicum, and how are these changes linked to the course or the practicum?" The findings of the interviews indicated that the practicum, albeit short, contributed to further increase the self-efficacy of most of the interviewed students. Simply having tried to teach pronunciation for the first time and seeing their learners' progress were important factors. However, in contrast to what was found by Bateman (2008) and Mattheoudakis (2007), very few changes were observed in the participants' beliefs about teaching and learning from Time 2 to Time 3. Given that the phonology course was very teaching-oriented and evidence-based, it is possible that what they learned was generally in line with the teaching reality they experienced and that the practicum solidified their beliefs. Another possibility is that the practicum was too short to generate substantial changes to these beliefs.

While the phonology course was found to increase the participants' overall confidence in their ability to teach pronunciation, the data indicated that they might not have been fully prepared to use more communicative pronunciation activities in their lessons. This problem may be common to other phonology and pronunciation pedagogy courses, as Baker (2014) noticed that even experienced teachers who had undergone specialized training in the field tended to use only controlled activities in class. Similarly, Burri et al. (2017) found that student teachers had a preference for controlled activities at the end of a graduate course in pronunciation pedagogy. This is concerning because ESL students who learn target sounds or patterns in controlled activities (e.g., reading, repetition, mechanical drills) may not transfer what they learned to their communicative production (Elliott, 1997). Although the phonology course here studied did not neglect guided and communicative teaching techniques (as was also the case in Burri et al., 2017), the use of such techniques in class may be more challenging and less predictable than the use of controlled ones, thus requiring more attention during training.

Conclusion and Future Directions

Overall, the findings of this study suggest that specific training that combines an introduction to English phonology and guidelines for pronunciation teaching has a positive impact on pre-service ESL teachers' cognitions. It is considerably more effective than general TESL instruction in helping teachers develop favorable views of explicit pronunciation teaching and confidence in their ability to teach pronunciation. Including a short practicum as part of the training has the potential of further enhancing self-efficacy. These results are promising, given that lack of confidence in teaching pronunciation is a common problem reported by many ESL/EFL teachers which can lead to avoidance of pronunciation instruction (Couper, 2016; Foote et al., 2011; Macdonald, 2002). To the author's knowledge, this was the first study on pronunciation teacher education to include a comparison group, making it easier to attribute the differences found in TC to the pronunciation training.

The findings here presented should be interpreted in light of the limitations of the study, which can be addressed in future research. First, it would be helpful if future studies included a measure of participants' cognitions *before* starting training. In this study, the first questionnaire was completed two to four weeks into the term, so it is not clear which cognitions expressed in the pre-questionnaire (and how much of them) had been influenced by the participants' first weeks of TESL instruction. Second, the use of random sampling of participants would be ideal, as there is always a possibility that the students who self-selected for this study differed systematically from those who did not (e.g., they may have been more interested in pronunciation than the average TESL student). Furthermore, interviewing all treatment participants rather than a subset of them would control for the possibility that the interview findings are not representative of the whole sample. Besides addressing these limitations, future research can also include delayed post-measures that investigate whether and how pronunciation-related cognitions change long after training (e.g., by the end of the TESL program). Finally, it would be interesting to explore how the teachers' beliefs play out in the classroom once they start teaching. Given that TC affects behavior in class, it is reasonable to believe that trained teachers will be more likely than non-trained teachers to integrate pronunciation instruction into their general ESL classes in the future. However, this is an empirical question that requires further investigation.

Chapter 5: General Conclusion

Introduction

The studies that make up this dissertation have their own specific objectives and were intended to stand alone. Nonetheless, taken together, their findings can provide deeper insight into issues that are relevant to the selection of instructional targets for L2 pronunciation teaching and teacher training. Specifically, Studies 1 and 2 provide evidence related to the under-researched topic of naturalistic acquisition of pronunciation by adult L2 speakers and offer a novel contribution to the body of literature dealing with the various factors that feed into listener evaluations of pronunciation. Study 3 addresses the effects of training in pronunciation pedagogy on the cognitions of pre-service ESL teachers, providing insights related to the student teachers' needs and to the factors that may contribute to the effectiveness of training in the field. In this chapter, I summarize and discuss key findings from the three studies regarding the general themes identified in Chapter 1 and draw pedagogical implications for pronunciation and EAP instruction and for teacher training in pronunciation pedagogy.

Summary and Discussion of Key Findings

Naturalistic Development of L2 Oral Performance

Review of the study methods. Before discussing the findings from Studies 1 and 2, it is pertinent to remind the reader of the main similarities and differences between the studies. Both Studies 1 and 2 investigated native-speaker listener evaluations of oral presentations made by international graduate students during their first two terms (six months) in an English-medium university located in a French-English bilingual and multicultural Canadian city. Almost all of the talker participants had been in Canada for one month or less and had never lived in an English-speaking country before. They did not receive instruction in speaking during the course of the study. Study 1 featured 10 NNS talker participants and two NS control participants, while Study 2 featured 11 talker participants, who were classified as either upper-intermediate or advanced speakers. There was an overlap of six talker participants between the two studies. The oral presentations in Study 1 were evaluated in their entirety for accentedness, comprehensibility, fluency, topic structure clarity, and overall quality at two time points (Month 1 and Month 7) by

listeners with a background in language teaching (mostly graduate students in applied linguistics). Also, the speakers' mean percentage of pitch increase to signal topic shifts in each presentation was calculated based on acoustic measurements. In Study 2, one-minute samples taken from oral presentations delivered at four time points (Month 1, Month 3, Month 5, and Month 7) were rated for accentedness, comprehensibility, fluency, speaking style, content, and organization by two groups of listeners: subject-matter specialists (graduate students in engineering and computer science) and non-specialists (undergraduate students in other fields).

Changes in speech and oral presentation skills. The findings from Study 1 showed that accentedness and comprehensibility ratings were significantly higher at Month 7 compared to Month 1. Half of the sample also had notably higher fluency ratings at Month 7, although the overall increase was not statistically significant. No significant changes were found for topic structure clarity, overall quality, and pitch increase at topic shifts, and these values actually decreased for half of the sample. The findings related to the pronunciation dimensions were interpreted as support for Flege's (1988) hypothesis that recent arrivals can indeed experience incidental improvement in pronunciation within their first year of immersion, as previous studies have found (Derwing & Munro, 2013; Munro & Derwing, 2008). Further, the findings suggest that this can occur even in a potentially less favorable situation, where the L2 speakers: (1) were already proficient in English; (2) were not taking ESL speaking classes, but subject-matter classes in English; (3) had potentially limited class time and did more independent work; and (4) were surrounded by NNSs of English and sometimes speakers of their own L1s, which may have meant less interaction with NSs.

On the other hand, these limiting factors may have contributed to the lack of overall improvement for the sample in Study 2, since no significant changes were found in any of the ratings given by either of the listener groups. An analysis of individual performance, however, revealed that most participants from Study 2 had one or more constructs showing upward trends from Month 1 to Month 7, although the changes were usually non-linear. Fluency and speaking style showed more upward trends than the other constructs (four cases each). Ratings that were already high at Month 1 (73% or more) tended to remain roughly the same, but sometimes displayed downward trends. The constructs with most cases of high ratings were organization (eight cases) and content (seven cases). High initial ratings were more common among advanced

speakers, while upward trends were more common among upper-intermediate speakers. This is in line with other studies showing that lower-proficiency speakers benefit more from study-abroad experiences (Freed, 1995; Lapkin et al., 1995; Llanes & Muñoz, 2009) and might indicate that some advanced learners had reached a plateau in their oral performance. Another difference between participants that may have played a role in the results was age of learning (AOL) onset in the speakers' home countries, which was found to correlate negatively with the pronunciation ratings, especially at Months 5 and 7. Other possible explanations for the non-significant results in Study 2 include the possibility that the speakers experienced more improvement in the second half of their first year or that the speech samples were too short for the listeners to perceive improvement in some cases. Finally, listener-related variables may have come into play, but these will be discussed later in the chapter.

Fluency in both studies and speaking style in Study 2 became notably higher for a few participants, but not for others. The fact that these two constructs behaved differently than accentedness and comprehensibility may suggest that they are more subject to situational factors, such as how nervous the speaker feels or how prepared he/she is for the presentation. Lennon (1990) described fluency as a "performance phenomenon" (p. 391) and connected dysfluencies with a speaker's effort to plan and produce the intended message. Logically, this effort will be greater in a less planned presentation, and previous research has confirmed that amount of planning can affect fluency in L2 speech (Ellis, 2009; Foster & Skehan, 1996). Regarding speaking style (changes in pitch that make the speech dynamic), it is reasonable to believe that amount of planning can influence the speaker's level of confidence, which in turn may result in more or less variations in pitch (Mennen, 1998; Zimmerer et al., 2014). Furthermore, some studies have found a link between planning and degree of discourse marking, which may include the use of pitch variations to signal topic structure (Crookes, 1989; Pickering, 2004; Williams, 1992). In fact, degree of pitch increase at topic shifts in Study 1 is also speculated to have been affected by planning. Lastly, this variable may have affected the constructs designed to evaluate oral presentations themselves (topic structure clarity and overall quality in Study 1; content and organization in Study 2), which tended to either remain the same or decrease from Month 1 to Month 7. It is unlikely that the participants who showed reductions in these scores actually became worse at their oral presentation skills as they progressed with their graduate studies.

What Underlies Evaluations of L2 Speech and Oral Presentations

Evaluations of oral presentations. In Study 1, one finding that stood out was that the ratings of overall quality were highly correlated with ratings of topic structure clarity, and the listeners often mentioned the structure of the presentation when explaining their evaluations of overall quality. Aspects said to contribute to a clearer discourse structure were lexical discourse markers, use of questions to introduce a new topic, statements indicating speaker intention, and a logical sequencing of ideas. Besides discourse marking, the listeners' comments mentioned other aspects related to what Rounds (1987) referred to as "classroom discourse techniques", which are typically used by instructors to create an interactive atmosphere and make the content accessible to the listeners. These aspects included providing introductions, examples, and conclusions, explaining concepts, defining terms, asking questions, and addressing the audience directly. Some of them can also contribute to signaling the structure of the presentation, such as an introduction that includes an overview of the topics to be covered and examples that are provided to illustrate main points or subpoints in a presentation, thus reinforcing them. These findings point to the importance of a clear discourse structure to evaluations of L2 oral presentations, confirming prior research (Rounds, 1987; Tyler, 1992; Tyler et al., 1988; Williams, 1992).

In Table 17, the main evaluation criteria for overall quality in Study 1 and for content and organization in Study 2 are organized into larger categories and displayed side by side. The last line of the table shows "Other" criteria that did not overlap. As can be observed, evaluation criteria for overall quality in Study 1 included criteria mentioned for both content and organization in Study 2, and also featured interactive elements (questions and addressing of the audience), intonation and naturalness, and conclusion or summary. While this last item is probably important to the organization (and perhaps the content) of a presentation, it could not have been mentioned by the raters in Study 2 because they only listened to one-minute samples from the beginning of the presentations. Despite a few differences between the evaluation criteria for content suitability and those for organization, several criteria overlapped, and there was no significant difference between the mean ratings given on these two constructs in Study 2. This could be interpreted as further indication of the importance of a clear topic structure for evaluations of oral presentations, since organization was defined as "the degree to which a speaker's presentation has a clear structure and connections between ideas" (Appendix C).

Table 17

Listener Evaluation Criteria for Oral Presentations from Studies 1 and 2

Categories	Study 1	Study 2	
	Overall quality	Content Suitability	Organization
Discourse structuring and organization	▪ Discourse structuring or logical sequencing of ideas	-	▪ Planning ▪ Connection between ideas ▪ Staying on track ▪ Logical order
Content	▪ Quality of content or ideas	▪ Advanced level ▪ Relevance / interestingness ▪ Basic level	-
Introduction	▪ Introduction, outline, or overview	▪ Introduction	▪ Introduction
Explanations and definitions	▪ Explanations, descriptions, and definitions	▪ Language / terminology ▪ Explanations	▪ Explanations
Examples	▪ Examples	▪ Use of examples	-
Confidence	▪ Confidence and knowledgeability	▪ Confidence / knowledge	-
Other	▪ Questions ▪ Addressing of the audience ▪ Intonation and naturalness ▪ Conclusion or summary	-	▪ Repetition ▪ Dysfluencies

The influence of paratones. Although topic structure clarity was considered an important aspect of oral presentations, Study 1 did not find that more native-like use of intonation (i.e., larger pitch increase) to signal topic shifts was related to topic structure clarity or overall quality. In fact, this variable was not significantly related to any of the listener ratings. The NNSs were found to produce much more restricted pitch increase values than the NSs comparison participants, but their presentations were usually rated highly on topic structure clarity. This was the case for Participant 1's first presentation (at Month 1), which listener comments also confirmed had a clear topic structure. Although her second presentation (Month 7) had a pitch increase value that was much higher and close to the NS values, there was not much change in structure clarity, as indicated by ratings and comments. There was, however, an increase in overall quality, and some comments suggested that her more varied intonation may have contributed to listener impressions of interestingness and confidence.

It is possible that other factors may have contributed to the salience of the topic structures of the presentations, compensating for the apparently weak intonation cues at topic shifts. Yet there was also evidence that intonation cues may be intertwined at some level with other forms of discourse marking and other "classroom discourse elements" (Rounds, 1987), so that reduced effort on the speaker's part to make his/her presentation accessible to the audience through lexical markers, introductions, examples, and questions, may also mean less marked intonation cues. As already mentioned, this weaker discourse structuring may stem from lack of preparation (Pickering, 2004; Williams, 1992), since a limited processing capacity may restrict the degree of discourse structuring that can take place on the spot.

Study 1's findings contrast with Wennerstrom's (1998), which found a relationship between pitch range expansion at rhetorical transitions and comprehensibility ratings. The participants in Wennerstrom's study, however, may have had lower proficiency in English, based on the minimum TOEFL scores reported and on the fact that they had all failed the SPEAK test. If this was truly the case, paratone measures may correlate with comprehensibility only up to a certain level of L2 proficiency. More advanced speakers may provide listeners with enough cues at other levels to help them identify the discourse structure, so that their non-native-like levels of pitch increase at topic shifts may not be problematic. Further, it is also possible that in a context of flatter pitch range overall (Mennen, 1998; Pickering, 2004; Zimmerer et al., 2014), the pitch

increase values produced by the L2 speakers, although lower than those of L1 speakers, may still be salient enough to distinguish new topics. A follow-up analysis comparing topic-initial pitch peaks to other peaks in the presentation could lend support to this hypothesis. If true, this may help to explain the findings in Kang et al. (2010), where the frequency of identifiable paratones (i.e., low terminations followed by high-key resets) appeared to be more important to listener ratings of oral proficiency and comprehensibility than the onset and termination pitch values themselves.

When discussing the use of acoustic measurements to assess pronunciation, Derwing and Munro (2015) underscore the fact that, even when these measurements are noticeably different from the target, the context of production may secure intelligibility. "Because human listeners take into account context at multiple levels, their perceptions may accommodate deviations from an expected target. These deviations might be very apparent in acoustic data, yet have little or no importance from the standpoint of intelligibility" (p. 9). This may have been the case for the NNS paratone measures in Study 1.

The influence of subject-matter knowledge. Study 2 examined the role of subject-matter knowledge in listener ratings of oral presentations. Significant differences between the quantitative ratings given by the two groups were found for organization and speaking style at two time points, and for fluency at one time point. At Month 7, all three of these constructs were judged significantly more harshly by the specialists. The analysis of the qualitative data revealed similarities and differences between the groups in terms of evaluation criteria, which are displayed in Table 18 below. To further summarize the main differences, it is possible to say that the specialists showed: (1) more detailed and specific comments; (2) more frequent references to mispronounced words in comments about accentedness and comprehensibility; (3) more mentions of accent as affecting comprehensibility; (4) more mentions of fillers as affecting fluency; (5) more words written per rater on content and organization; (6) more references to level of difficulty when judging content and to connections between ideas and introductions when judging organization; and (7) more comments on content accuracy.

Table 18

Comparison of Non-Specialist and Specialist Comments from Study 2

Construct	Non-specialists	Specialists
Accentedness	▪ More comments on "strong" than "not strong" accents	▪ More comments on **mispronounced words**
Comprehensibility	▪ Quality of enunciation was most important	▪ **Accent**, **fluency**, and **mispronunciations** were also important
Fluency*	▪ Pausing and **hesitations** were most important ▪ Much fewer comments on fillers	▪ Pausing and **fillers** were most important ▪ More references to **self-repetition**
Speaking style*	▪ No specific criteria besides a few comments on **slow speech rate**.	
Content	▪ 176 words/rater ▪ More comments on speaker's **level of confidence** and **knowledge** ▪ Fewer comments on **accuracy** ▪ Comments on accuracy were **brief** and **all positive**	▪ 231 words/rater (**31% more**) ▪ More comments on **level of difficulty** and **language technicality** ▪ More comments on **introductions** ▪ More comments on **accuracy**
Organization*	▪ 112 words/rater ▪ **Planning** was the most frequently mentioned aspect, followed by **connections between ideas**	▪ 179 words/rater (**59% more**) ▪ Planning was third, after **connections between ideas** and **introductions** ▪ More comments about **staying on track** and **repetition**
Other/Overall	▪ **Fewer** and **more generic** evaluation criteria ▪ More comments on **nervousness**	▪ **More specific** and **varied** evaluation criteria; **more detailed** comments

*Significant differences in ratings found at one or more time points.

The results related to content and organization corroborate previous findings from the ESP literature suggesting that these are important aspects in subject-matter specialists' evaluations of academic and professional tasks (Douglas & Myers, 2000; Elder et al., 2012; Jacoby, 1998; Knoch, 2014). Further, the fact that specialists were more concerned about the level of difficulty of a presentation, including the use of overly technical language and undefined terms, and placed greater value on the presence of an introduction and connections between ideas, suggests that they are more aware of the characteristics and communicative goals of the genre of academic presentations, which is also in line with prior studies (Elder, 1993; Elder et al., 2017; Jacoby, 1998). The specialists were also found to provide more detailed comments, using more varied evaluation criteria, even when commenting on the speech dimensions. It is possible that their familiarity with the subject and genre under assessment meant that their processing resources were less strained with trying to follow the presentations, so they were able to pay more attention to the speakers' pronunciation and write better comments. Alternatively, this might be accounted for by the fact that the specialists were graduate students and therefore were more educated than the non-specialists, who were all undergraduate students.

The influence of language teaching knowledge. When comparing findings from Studies 1 and 2, the influence of language teaching knowledge on listener ratings can also be discussed. Notably, it was observed that the raters with a background in TESL or applied linguistics were able to perceive improvement in the pronunciation of the L2 graduate students in Study 1, whereas the raters from Study 2 were not. Naturally, this may have been caused by an objective difference in performance between the non-overlapping participants in the two studies and several other factors discussed in the previous section of this chapter. Still, it is reasonable to suppose that the raters' backgrounds in Study 1 made them more attuned to changes in pronunciation than the raters in Study 2. Previous research has indeed found that raters with linguistic and pedagogical experience may give more positive ratings on pronunciation dimensions (Saito et al., 2017; I. Thompson, 1991). Therefore, taken together, the insights provided by Studies 1 and 2 suggest that, when assessing L2 speech and specific oral tasks, it is important to elicit the perspective of listeners who resemble the speakers' potential audience or interlocutors, since their views may not coincide with those of other listeners.

Teacher Training in L2 Pronunciation

So far, this chapter has addressed key findings from Studies 1 and 2 related to acquisition and assessment of L2 speech and to evaluations of academic presentations, which can help to inform pedagogical practice in L2 pronunciation and EAP. Study 3 sought to provide insights for evidence-based teacher training in L2 pronunciation. The design of the study compared changes in the cognitions related to L2 pronunciation of two groups of TESL students. The first group consisted of 18 students who took a 13-week undergraduate course in English phonology and pronunciation pedagogy. The second group consisted of 15 students from another university who did not receive specific training in these areas, but had otherwise similar training to that received by the first group, including a grammar course and general TESL pedagogy classes. The course in phonology and pronunciation pedagogy included theoretical blocks on English segmentals, suprasegmentals, and pronunciation teaching followed by a short practicum.

Findings from a pre-post questionnaire revealed that, towards the end of the term, the treatment group believed significantly more than the control group that pronunciation teaching was not dispensable, that it was important to know phonology to teach pronunciation, and that teaching pronunciation was not boring. This suggests that specific training in pronunciation pedagogy promoted more favorable views of explicit pronunciation teaching than general training in TESL. Furthermore, a comparison between self-efficacy means, calculated from the participants' responses to 19 questionnaire items, revealed that both groups became more confident in their abilities to teach pronunciation, but the effect size of the change was greater for the treatment group. Data from interviews conducted with a subset of eight treatment participants revealed changes in NS participants' views of NNSs' suitability to teach pronunciation. While in most cases the participants developed positive views regarding the suitability of their NNS peers, as was the case in Burri (2015b), some of them expressed concern that NNS teachers' pronunciation might negatively affect their teaching. Another change in TC identified from the interviews was that some NNS participants became more aware of differences between their own pronunciations and the native model used in the course, while some NSs became aware of limitations in their explicit knowledge *about* pronunciation.

The interview data confirmed the role of the course in the cognition changes identified above. Specific aspects of the course that helped to promote positive cognition changes were:

theory related to English phonology; detailed explanations and demonstrations of what to do in a pronunciation lesson; access to sample pronunciation activities; for NSs, interactions with NNS peers and instructors; and the short practicum that gave the student teachers a chance to teach pronunciation for the first time. Study 3 therefore confirms previous studies which have also identified positive effects of specialized training in pronunciation pedagogy on TC (Baker, 2011a; Burri, 2015a, 2015b; Burri et al., 2017), and it also suggests that general TESL training does not produce the same effects. There was evidence, however, that self-efficacy could have increased even more had the course featured more training in the use of communicative activities to teach pronunciation. Pedagogical attention to the pronunciation of NNS pre-service teachers might have been helpful as well. All in all, considering the positive results of Study 3, the findings from Baker (2011a), and the major influence of TC on teaching practices (e.g., Ghaith & Yaghi, 1997; Smylie, 1988), it is reasonable to believe that specific training of the kind reported in Study 3 increases the likelihood that pre-service teachers will address L2 pronunciation in their ESL classes. This is, however, a topic for future research to confirm.

Pedagogical Implications

The first chapter of this dissertation started and concluded with references to the important task of establishing instructional priorities for pronunciation teaching and for teacher training in pronunciation pedagogy. It was said that this decision should consider the needs of the students (or student teachers) and research evidence on the features or topics that may be more relevant to achieving the desired goals: improve communication, for pronunciation teaching, and improve teacher cognitions and practices, for teacher training. This section begins by discussing pedagogical implications from Studies 1 and 2, especially for pronunciation and EAP teaching to L2 graduate students, focusing on what the studies have to say about "what to teach" and providing brief suggestions on "how to teach it". It ends with evidence from Study 3 that is relevant to those involved in teacher training in pronunciation pedagogy.

One finding from Studies 1 and 2 is that naturalistic improvement of L2 pronunciation, particularly accentedness and comprehensibility, is possible for international graduate students during their early period of residence and study in an L2 environment. However, improvement is not guaranteed, and it may be too subtle to be perceived by these students' regular interlocutors.

Thus, it would be helpful for them to receive focused instruction on speaking upon entering their programs, so as to take advantage of the "window of maximal opportunity" (Derwing & Munro, 2015) to promote further pronunciation development. Instruction may be particularly necessary for advanced speakers who wish to improve their pronunciation, given that their accents may have plateaued and may be more resistant to naturalistic change than that of lower-proficiency speakers.

As for fluency and speaking style (i.e., variations in pitch that make the presentation more interesting and dynamic), there was evidence that performance on these dimensions may be more volatile and susceptible to situational variables, such as amount of planning. The same appeared to be true for discourse structuring, including the use of intonation to signal topic shifts in a presentation. ESL teachers can help their learners by raising their awareness of how planning can affect their oral performance, especially in monologue. In EAP classes, in particular, students can be encouraged to plan and rehearse academic presentations, and they can be taught outlining skills and how to make the best use of visual aids, such as PowerPoint slides.

Studies 1 and 2 also provided insights into aspects that underlie listener evaluations of L2 speech and oral presentations. One main finding was that the salience of the discourse structure speaks to the quality of a presentation. Therefore, it is important for EAP learners to learn and practice using discourse structuring devices that highlight the macro-organization of a presentation, which include paratones. Although, in Study 1, degree of pitch increase at topic shifts was not directly related to listener ratings, this does not preclude the role of paratones in discourse structuring, and it is probably beneficial (and fun) for learners to be made aware of paratones and practice producing them, along with other structuring cues. Read-aloud activities and pitch visualizing software may be useful for this purpose (see Levis & Pickering, 2004). However, teachers are reminded that evidence regarding the role of paratones in listener evaluations is still limited, and that native patterns may not be necessary for effective use of paratones.

EAP learners may also benefit from practicing different ways of introducing and concluding a presentation, summarizing information, giving examples to illustrate points, creating an interactive atmosphere with the audience, and sounding more confident, interesting, and natural, all of which were mentioned in listener evaluations of overall quality in Study 1. To

improve in the last point, the activities suggested in Levis and Pickering (2004) and others that encourage learners to vary their pitch in English, such as drama exercises, may be effective.

Findings from Study 2 indicated that subject-matter specialists mentioned mispronounced words, including technical terms, more frequently than non-specialists when evaluating L2 presentations for accentedness and comprehensibility. While these mispronunciations may stem from segmental and prosodic difficulties due to L1 interference, it is often the case that they are idiosyncratically deviant words (see Szpyra-Kozłowska, 2013), whose erroneous production may be influenced by orthography. As suggested in Szpyra-Kozłowska and Stasiak (2010), these problematic words can be tackled through awareness-raising activities, perception exercises, and repetition. Learners can also be taught strategies to learn the pronunciation of key words in their fields, such as checking phonetic transcriptions in a dictionary.

Finally, the results of Study 3 provide important insights for course design in L2 pronunciation pedagogy. Although generalizations about pronunciation training cannot be made, the results may be considered transferable to similar training contexts; that is, undergraduate courses within TESL programs. The findings pointed to the importance of theory in phonology and especially practical knowledge about pronunciation teaching. Regarding this last point, teachers in training may benefit from: detailed explanations and demonstrations of what to do in a pronunciation lesson, examples of pronunciation activities, and the opportunity to teach pronunciation lessons during a short practicum. Interactions between NSs and NNSs in the class may be beneficial as well, for example, during pair and group work in class or as co-teachers for a practicum. Plenty of guidance seems to be needed regarding the use of communicative activities in pronunciation teaching, including many sample activities, with step-by-step instructions, and strategies for correcting pronunciation issues in spontaneous, authentic speech. Finally, some pedagogical attention to the pronunciation of NNSs may be important for them to develop confidence. If not possible, a course structure that integrates training in pronunciation teaching techniques with the theory learned throughout the course may give NNSs a chance to apply these techniques to their own pronunciation and have a sense of pronunciation improvement by the end, as was the case in Burri (2015b).

References

Albini, A., & Kluge, D. (2011). Professores de inglês da rede pública paranaense e o ensino da pronúncia [Public school English teachers from Paraná and the teaching of pronunciation]. *Revista de Letras DACEX/UTFPR*, *1*, 1–12.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*(4), 529–555.

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*. New York: Longman.

Atkinson, M. (2005). *Lend me your ears: All you need to know about making speeches and presentations*. New York, NY: Oxford University Press.

Baker, A. (2011a). ESL teachers and pronunciation pedagogy: Exploring the development of teachers' cognitions and classroom practices. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference* (pp. 82–94). Aimes, IA: Iowa State University.

Baker, A. (2011b). *Pronunciation Pedagogy: Second Language Teacher Cognition and Practice*. (Unpublished doctoral dissertation). Georgia State University, Atlanta.

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly*, *48*(1), 136–163.

Baker, W. (2010). Effects of age and experience on the production of English word-final stops by Korean speakers. *Bilingualism: Language and Cognition*, *13*(3), 263–278.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215.

Barr, P. (1990). The role of discourse intonation in lecture comprehension. In M. Hewings (Ed.), *Papers in Discourse Intonation. Discourse Analysis Monographs 16* (pp. 5–21). Birmingham: ELR/University of Birmingham.

Bateman, B. (2008). Student teachers' attitudes and beliefs about using the target language in the classroom. *Foreign Language Annals*, *41*(1), 11–28.

Berman, R., & Cheng, L. (2010). English academic language skills: Perceived difficulties by

undergraduate and graduate students, and their academic achievement. *Canadian Journal of Applied Linguistics*, *4*(1–2), 25–40.

Bernard, H. R., Killworth, P., Kronenfeld, D., & Sailer, L. (1984). The problem of informant accuracy: The validity of retrospective data. *Annual Review of Anthropology*, *13*, 495–517.

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer [Computer program]. Version 5.3.77, retrieved 18 May 2014 from http://www.praat.org/.

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, *36*(2), 81–109.

Borg, S. (2006). *Teacher cognition and language education: Research and practice*. London: Continuum.

Brazil, D., Coulthard, M., & Johns, C. (1980). *Discourse intonation and language teaching*. London: Longman.

Breitkreutz, J., Derwing, T. M., & Rossiter, M. J. (2001). Pronunciation teaching practices in Canada. *TESL Canada Journal*, *19*(1), 51–61.

Brown, A. (1995). The effect of rater variables in the development of an occupation-specific language performance test. *Language Testing*, *12*(1), 1–15.

Brown, G. (1977). *Listening to spoken English*. London: Longman.

Brown, G., Currie, K. L., & Kenworthy, J. (1980). *Questions of intonation*. London: Croom Helm.

Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.

Burgess, J., & Spencer, S. (2000). Phonology and pronunciation in integrated language teaching and teacher education. *System*, *28*(2), 191–215.

Burns, A. (2006). Integrating research and professional development on pronunciation teaching in a national adult ESL program. *TESL Reporter*, *39*(2), 34–41.

Burri, M. (2015a). "My perspective changed dramatically": A case for preparing L2 instructors to teach pronunciation. *English Australia Journal*, *31*(1), 19–37.

Burri, M. (2015b). Student teachers' cognition about L2 pronunciation instruction: A case study. *Australian Journal of Teacher Education*, *40*(10), 66–87.

Burri, M., Baker, A., & Chen, H. (2017). "I feel like having a nervous breakdown": Pre-service and in-service teachers' developing beliefs and knowledge about pronunciation instruction.

Journal of Second Language Pronunciation, *3*(1), 109–135.

Buss, L. (2013). Pronunciation from the perspective of pre-service EFL teachers: An analysis of internship reports. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 255–264). Ames, IA: Iowa State University.

Buss, L. (2016). Beliefs and practices of Brazilian EFL teachers regarding pronunciation. *Language Teaching Research*, *20*(5), 619–637.

Buss, L. (2017). The role of training in shaping pre-service teacher cognition related to L2 pronunciation. *Ilha Do Desterro*, *70*(3), 201–226.

Buss, L., Cardoso, W., & Kennedy, S. (2015a). Discourse intonation in L2 academic presentations: A pilot study. In J. Levis, R. Mohammed, M. Qian, & Z. Zhou (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference*. Santa Barbara, CA (pp. 27-37). Ames, IA: Iowa State University.

Buss, L., Cardoso, W., & Kennedy, S. (2015b). Longitudinal changes in the use of paratones in L2 English speech by Mandarin speakers. *Organon*, *30*(58), 215–227.

Carey, M. D., Mannell, R. H., & Dunn, P. K. (2011). Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing*, *28*(2), 201–219.

Celce-Murcia, M., Brinton, D., & Goodwin, J. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). New York: Cambridge University Press.

Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago: University of Chicago Press.

Chaudron, C., & Richards, J. C. (1986). The effect of discourse markers on the comprehension of lectures. *Applied Linguistics*, *7*(2), 113–127.

Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam and Philadelphia: John Benjamins.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Couper, G. (2016). Teacher cognition of pronunciation teaching amongst English language teachers in Uruguay. *Journal of Second Language Pronunciation*, *2*(1), 29–55.

Crookes, G. (1989). Planning and Interlanguage Variation. *Studies in Second Language*

Acquisition, *11*(04), 367.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–533.

Derwing, T. M. (2008). Curriculum issues in teaching pronunciation to second language learners. In J. Edwards & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 347–369). Philadelphia, PA: John Benjamins.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *20*(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. (2008). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(04), 476.

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*(2), 163–185.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Philadelphia, PA: John Benjamins Publishing Company.

Derwing, T. M., Munro, M. J., Foote, J. a., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, *64*(3), 526–548.

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*(3), 359–380.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1997). Pronunciation instruction for fossilized learners. Can it help? *Applied Language Learning*, *8*(2), 217–235.

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.

Douglas, D., & Myers, R. K. (2000). Assessing the communication skills of veterinary students: Whose criteria? In A. Kunnan (Ed.), *Fairness in validation in language assessment* (pp. 60–81). Cambridge: Cambridge University Press.

Elder, C. (1993). How do subject specialists construe classroom language proficiency? *Language*

Testing, *10*(3), 235–254.

Elder, C., Mcnamara, T., Kim, H., Pill, J., & Sato, T. (2017). Interrogating the construct of communicative competence in language assessment contexts: What the non-language specialist can tell us. *Language Sciences*, *57*, 14–21.

Elder, C., Pill, J., Woodward-Kron, R., McNamara, T., Manias, E., Webb, G., & McColl, G. (2012). Health professionals' views of communication: Implications for assessing performance on a health-specific English language test. *TESOL Quarterly*, *46*(2), 409–419.

Elliott, A. R. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania*, *80*(1), 95.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, *30*(4), 474–509.

Ferris, D. (1998). Students' views of academic aural/oral skills: A comparative needs analysis. *TESOL Quarterly*, *32*(2), 289.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, *84*(1), 70–79.

Flege, J. E., Bohn, O., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.

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

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*, 527–552.

Flege, J. E., Munro, M. J., & Fox, R. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, *95*(6), 3623–3641.

Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995). Effects of age of second-language learning on the production. *Speech Communication*, *16*, 1–26.

Foote, J., Holtby, A., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal*, *29*(1), 1–22.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second laguage performance. *Studies in Second Language Acquisition*, *18*, 299–323.

Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In B. F.

Freed (Ed.), *Second language acquisition in a study-abroad context* (pp. 123–148). Philadelphia, PA: John Benjamins Publishing Company.

Gass, S. M., & Varonis, E. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, *34*(1), 65–89.

Ghaith, G., & Yaghi, H. (1997). Relationships among experience, teacher efficacy, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education*, *13*(4), 451–458.

Gibson, W., & Brown, A. (2009). *Working with qualitative data*. London: Sage.

Grosz, B. J., & Hirschberg, J. (1992). Some intonational characteristics of discourse structure. *Proceedings of ICSLP 92*, Banff, 429-432.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, *12*(3), 175–204.

Hahn, L. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, *38*(2), 201–223.

Henderson, A. (2013). The English Pronunciation Teaching in Europe Survey: Initial results and useful insights for collaborative work. In E. Waniek-Klimczak & L. R. Shockey (Eds.), *Teaching and researching English accents in native and non-native speakers* (pp. 123–136). Berlin, Heidelberg: Springer Berlin Heidelberg.

Hincks, R. (2005). Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism. *System*, *33*, 575–591.

Hismanoglu, M., & Hismanoglu, S. (2010). Language teachers' preferences of pronunciation teaching techniques: traditional or modern? *Procedia - Social and Behavioral Sciences*, *2*(2), 983–989.

Horwitz, E. (1985). Using student beliefs about language learning and teaching in the foreign language methods course. *Foreign Language Annals*, *18*(4), 333–340.

Ioup, G. (2008). Exploring the role of age in the acquisition of a second language phonology. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 41–62). Philadelphia, PA: John Benjamins Publishing Company.

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. *Studies in Second Language Acquisition*, *34*(3), 475–505.

Jacoby, S. (1998). *Science as performance: Socializing scientific discourse through the conference talk rehearsal*. (Unpublished doctoral disseration). University of California, Los Angeles.

Jacoby, S., & McNamara, T. (1999). Locating competence. *English for Specific Purposes*, *18*(3), 213–241.

Jenkins, J. (2005). Implementing an international approach to English pronunciation: The role of teacher attitudes and identity. *TESOL Quarterly*, *39*(3), 535–543.

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

Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly*, *9*(3), 249–269.

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

Kennedy, S., Foote, J. A., & Buss, L. K. S. (2015). Second language speakers at university: Longitudinal development and rater behaviour. *TESOL Quarterly*, *49*(1), 199–209.

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *The Canadian Modern Language Review*, *64*(3), 459–489.

Knoch, U. (2014). Using subject specialists to validate an ESP rating scale: The case of the International Civil Aviation Organization (ICAO) rating scale. *English for Specific Purposes*, *33*, 77–86.

Kraut, R., & Wulff, S. (2013). Foreign-accented speech perception ratings: a multifactorial case study. *Journal of Multilingual and Multicultural Development*, *34*(3), 249–263.

Labov, W., & Waletsky, J. (1967). Narrative analysis: Oral versions of personal experience. In *Essays on the verbal and visual arts* (pp. 12–44). Seattle, WA: University of Washington

Press.

Lapkin, S., Hart, D., & Swain, M. (1995). A Canadian interprovincial exchange: Evaluating the linguistic impact of a three-month stay in Quebec. In B. F. Freed (Ed.), *Second language acquisition in a study-abroad context* (pp. 67–94). Philadelphia, PA: John Benjamins Publishing Company.

Lehiste, I. (1975). The phonetic structure of paragraphs. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and Process in Speech Perception* (pp. 195–203). Berlin: Springer-Verlag.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387–417.

Levis, J., & Pickering, L. (2004). Teaching intonation in discourse using speech visualization technology. *System*, *32*(4), 505–524.

Lightbown, P., & Spada, N. (1995). *How languages are learned*. Oxford: Oxford University Press.

Llanes, A., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, *37*(3), 353–365.

Lucas, S. E. (2009). *The art of public speaking* (10th ed.). New York, NY: McGraw Hill.

Lumley, T. (1998). Perceptions of language-trained raters and occupational experts in a test of occupational English language profficiency. *English for Specific Purposes*, *17*(4), 347–367.

Lumley, T., Lynch, B. K., & McNamara, T. F. (1994). A new approach to standard-setting in language assessment. *Melbourne Papers in Language Testing*, *3*(2), 19–40.

MacDonald, M., Badger, R., & White, G. (2001). Changing values: what use are theories of language learning and teaching? *Teaching and Teacher Education*, *17*(8), 949–963.

Macdonald, S. (2002). Pronunciation: Views and practices of reluctant teachers. *Prospect*, *17*(3), 3–18.

Macintyre, P. D., Dörnyei, Z., Clément, R., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, *82*(4), 545–562.

Mackay, I. R. A., Flege, J. E., & Imai, S. (2006). Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics*, *27*(2), 157–183.

Mattheoudakis, M. (2007). Tracking changes in pre-service EFL teacher beliefs in Greece: A longitudinal study. *Teaching and Teacher Education*, *23*(8), 1272–1288.

Mennen, I. (1998). Can language learners ever acquire the intonation of a second language? In *ETRW on Speech Technology in Language Learning (STiLL)* (pp. 17–20). Marholmen, Sweden: International Speech Communication Association.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Thousand Oaks: Sage Publications.

Morita, N. (2000). Discourse socialization through oral classroom activities in a TESL graduate program. *TESOL Quarterly*, *34*(2), 279–310.

Muñoz, C., & LLanes, A. (2014). Study abroad and changes in degree of foreign accent in children and adults. *The Modern Language Journal*, *98*(1).

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, *36*(1), 39–66.

Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. Edwards & M. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193–218). Amsterdam: John Benjamins.

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

Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning*, *48*(2), 159–182.

Munro, M. J., & Derwing, T. M. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, *34*(4), 520–531.

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, *58*(3), 479–502.

Munro, M. J., & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century: A point of view. *Journal of Second Language Pronunciation*, *1*(1), 11–42.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*(01), 111–131.

Munro, M. J., Derwing, T. M., & Saito, K. (2013). English L2 vowel acquisition over seven years. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second*

Language Learning and Teaching Conference. Aug. 2012. (pp. 112-119). Ames, IA: Iowa State University.

Murphy, J. (2014). Teacher training programs provide adequate preparation in how to teach pronunciation. In L. Grant (Ed.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 188–219). Ann Arbor, MI: The University of Michigan Press.

Okamura, A. (1995). Teachers' and nonteachers' perception of elementary learners' spoken Japanese. *The Modern Language Journal*, *79*(1), 29–40.

Pickering, L. (2001). The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly*, *35*(2), 233–255.

Pickering, L. (2004). The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes*, *23*(1), 19–43.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in discourse. In *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.

Pill, J. (2016). Drawing on indigenous criteria for more authentic assessment in a specific-purpose language test: Health professionals interacting with patients. *Language Testing*, *33*(2), 175–193.

Piske, T., Mackay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191–215.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191–215.

Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, *100*, 190–208.

Read, J., & Wette, R. (2009). Achieving English proficiency for professional registration: The experience of overseas-qualified health professionals in the New Zealand context. In J. Osborne (Ed.), *IELTS research reports* (Vol. 10, p. Vol. 10, pp. 181–222). Canberra: IELTS Australia.

Reves, T., & Medgyes, P. (1994). The non-native English speaking EFL/ESL teacher's self-image: An international survey. *System*, *22*(3), 353–367.

Roberts, J., & Henson, R. (2000). *Self-efficacy Teaching and Knowledge Instrument for Science Teachers (SETAKIST): A proposal for a new efficacy instrument*. Paper presented at the

Annual Meeting of the Mid-South Educational Research Association (28th, Bowling Green, KY, November 17-19, 2000).

Rounds, P. (1987). Characterizing successful classroom discourse for NNS teaching assistant training. *TESOL Quarterly*, *21*(4), 643–671.

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.

Samimy, K., & Brutt-Griffler, J. (1999). Perceptions of NNS students in a graduate TESOL program. In G. Braine (Ed.), *Non-native educators in English language teaching* (pp. 129–146). Mahwah, NJ: Lawrence Erlbaum.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158.

Segalowitz, N., & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, *26*, 173–199.

Sifakis, N., & Sougari, A. (2005). Pronunciation issues and EIL pedagogy in the periphery: A survey of Greek state school teachers' beliefs. *TESOL Quarterly*, *39*(3), 467–488.

Smylie, M. A. (1988). The enhancement function of staff development: Organizational and psychological antecedents to individual teacher change. *American Educational Research Journal*, *25*(1), 1–30.

Szpyra-Kozłowska, J. (2013). On the irrelevance of sounds and prosody in foreign-accented English. In E. Waniek-Klimczak & L. R. Shockey (Eds.), *Teaching and researching English accents in native and non-native speakers* (pp. 15–29). Berlin: Springer.

Szpyra-Kozłowska, J., & Stasiak, S. (2010). From focus on sounds to focus on words in English pronunciation instruction. *Research in Language*, *8*, 1–12.

Tench, P. (1996). *The intonation systems of English*. London: Cassell.

Tergujeff, E. (2012a). English pronunciation teaching: Four case studies from Finland. *Journal of Language Teaching and Research*, *3*(4), 599–607.

Tergujeff, E. (2012b). The English Pronunciation Teaching in Europe Survey: Finland. *Apples*, *6*(1), 29–45.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, *27*(2), 237–246.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, *41*(2), 177–204.

Thompson, S. (1994). Aspects of cohesion in monologue. *Applied Linguistics*, *15*(1), 58–75.

Thompson, S. (2003). Text-structuring metadiscourse, intonation and the signalling of organisation in academic lectures. *Journal of English for Academic Purposes*, *2*(1), 5–20.

Timmis, I. (2002). Native-speaker norms and international English: a classroom view. *ELT Journal*, *56*(3), 240–249.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effects of L2 prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*, 1–30.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905–916.

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, *68*(2), 202–248.

Tyler, A. (1992). Discourse structure and the perception of incoherence in international teaching assistants' spoken discourse. *TESOL Quarterly*, *26*(4), 713–729.

Tyler, A., Jefferies, A., & Davies, C. (1988). The effect of discourse structuring devices on listener perceptions of coherence in non-native university teacher's spoken discourse. *World Englishes*, *7*(2), 101–110.

Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics*, *15*(4), 399–420.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse. *Studies in Second Language Acquisition*, *20*, 1–25.

Wennerstrom, A. (2001). *The music of everyday speech*. New York: Oxford University Press.

Williams, J. (1992). Planning, discourse marking, and the comprehensibility of international teaching assistants. *TESOL Quarterly*, *26*(4), 693–711.

Yao, Z., Saito, K., Trofimovich, P., & Isaacs, T. (2013). Z-Lab [Computer software]. Retrieved from https://github.com/ZeshanYao/Z-Lab.

Yule, G. (1980). Speakers' topics and major paratones. *Lingua*, *52*, 33–47.

Zappa-Hollman, S. (2007). Academic presentations across post-secondary contexts: The discourse socialization of non-native English speakers. *The Canadian Modern Language Review*, *63*(4), 455–485.

Zimmerer, F., Jügler, J., Adreeva, B., Möbius, B., & Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. In Campbell, Gibbon, & Hirst (Eds.), *Speech Prosody 7* (pp. 1037–1041). Dublin, Ireland.

Appendices

Appendix A: Written Instructions for Listeners (Study 1)

You will listen to 12 academic presentations, each about 5 minutes long. The speakers were asked to talk about an important concept from their field of research to an imagined audience of first-year undergraduate students. After listening to each presentation, you will rate it on 5 scales: *accentedness*, *comprehensibility*, *fluency*, *structure clarity*, and *overall quality*. Then, you will provide a brief comment explaining your rating of *overall quality*. Please rate each presentation to the best of your ability.

Scale	Explanation
Accentedness	The degree to which a speaker's speech sounds different from native speakers' speech. If a speaker's speech is very different from that of native speakers, that speaker has a very strong non-native accent. However, if a speaker's speech is very similar to that of native speakers, that speaker has minimal non-native accent or has a native accent.
Comprehensibility	How easy it is to understand the speaker's speech. If you find it difficult to identify the words that a speaker is saying, then that speaker has low comprehensibility. If you can easily identify the words that a speaker is saying, then that speaker is highly comprehensible.
Fluency	The degree to which speech is smooth and fluid, without inappropriate pauses or starts and stops. If a speaker has many inappropriate pauses or hesitations, then the speech has low fluency (i.e., it is very dysfluent). If a speaker speaks fluidly, pausing only in appropriate places, then the speech is highly fluent.

Structure clarity	The degree to which a speaker's presentation has a clear topic structure. If it is unclear or hard to identify when one topic ends and another one starts, then the presentation has low structure clarity (i.e., the structure is very unclear). If it is clear to the listener when one topic ends and another one starts, then it has high structure clarity (i.e., the structure is very clear).
Overall quality	Your overall impression of the presentation. If you think the quality was very poor, it is a very bad presentation. If you think the quality is very good, it is a very good presentation. Please comment briefly on the criteria you used to determine your rating.

BEFORE YOU BEGIN…

1) You will start by clicking LISTEN. You will hear each speech sample once and only once.

2) You must rate and provide comments for one sample before you can move on to the next one.

3) Ensure that you use the correct scale for each category:

The left side ☹ represents a low score while the right side ☺ a good score

4) Please use the entire scale (left to right: 1 to 1000) as much as you can.

5) You can modify your rating or comments at anytime during the recording and after.

6) When you are finished, click SUBMIT to send your comments, then NEXT.

Appendix B: Mean Listener Ratings and Mean Pitch Increase by Participant (Study 1)

	Accentedness		Comprehensibility		Fluency		Structure Clarity		Overall quality		Pitch increase	
	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2	T1	T2
Participant 1	20.9	33.5	61.6	74.5	54.7	76.6	77.5	83.7	71.0	84.5	52.0	88.4
Participant 2	24.7	32.7	69.9	68.8	65.0	66.0	75.8	81.0	77.2	81.1	67.9	59.4
Participant 3	25.2	29.0	69.9	75.3	55.7	70.8	78.3	79.8	77.8	74.8	42.0	40.0
Participant 4	30.3	31.1	68.1	69.3	67.2	44.1	81.8	72.4	80.0	72.4	16.3	29.8
Participant 5	39.9	49.0	77.3	83.3	69.9	68.5	86.5	70.0	85.6	73.1	54.1	39.6
Participant 6	57.1	62.7	88.3	89.8	91.9	95.0	85.5	62.4	85.2	64.4	53.4	36.8
Participant 7	22.0	22.4	73.1	79.3	61.0	74.7	85.4	73.2	87.0	74.4	54.4	49.9
Participant 8	14.4	18.7	24.5	32.3	30.5	47.7	47.9	55.8	42.8	53.9	31.9	14.2
Participant 9	19.6	20.4	28.8	44.4	23.2	27.8	44.9	52.1	36.3	49.0	57.1	56.3
Participant 10	9.0	24.7	38.0	52.9	57.2	74.4	74.7	69.4	70.3	66.9	42.3	46.5
Participant 11[a]	99.5	100.0	97.9	98.7	90.7	94.8	91.7	95.6	93.0	93.0	100.2	149.4
Participant 12[a]	99.0	99.8	98.6	99.2	95.8	98.9	94.1	85.6	95.4	90.1	169.3	184.4

Note. The listener ratings are all percentages from 0-100%. Pitch increase values are percentages of increase with no upper limit. T1 stands for Time 1 and T2 stands for Time 2.

[a] Native speaker control participant

Appendix C: Written Instructions for Listeners (Study 2)

You will listen to 44 speech samples, each about 1 minute long. After listening to each speech sample, you will rate the sample on 6 scales: accentedness, comprehensibility, content, organization, fluency, and speaking style. Then, you will provide comments about your reasons for your ratings.

These speech samples have been taken from brief presentations in which speakers were asked to present an important concept from their field of research. The speakers were asked to imagine an audience of first-year undergraduate students. Unlike last session, the speech samples have not been processed; you will hear the speech 'in the clear'.[7] Before hearing each presentation, you will hear a brief phrase giving the presentation topic. **Each sample you hear is from a unique presentation, with no repeated samples**. Please rate each sample as a unique presentation and to the best of your ability.

Word	Explanation
Accentedness	The degree to which a speaker's speech sounds different from native speakers' speech. If a speaker's speech is very different from that of native speakers, that speaker is highly accented. However, if a speaker's speech is very similar to that of native speakers, that speaker has minimal or no accent.

[7] The data used for this study are part of a larger project that also aimed to compare ratings given to two versions of the same speech samples: one that had been low-pass filtered so that it contained little or no segmental information and another that was "in the clear", with no filtering. The filtered samples were rated in a separate session, approximately two weeks before the ratings of the clear samples. Data from the filtered samples are not reported in this dissertation.

Comprehensibility	How easy it is to understand the speaker's speech. If you can easily identify the words that a speaker is saying, then that speaker is highly comprehensible. However, if you find it difficult to identify the words that a speaker is saying, then that speaker has low comprehensibility.
Content	The degree to which the content of the presentation is accurate and is suitable for the audience (first-year undergraduate students). If the content is highly accurate and highly appropriate for first-year undergrads, then the content is very accurate and suitable. If the content is quite inaccurate and inappropriate for first-year undergrads, then the content is very inaccurate and unsuitable.
Organization	The degree to which a speaker's presentation has a clear structure and connections between ideas. If you can clearly perceive that the presentation is structured, with clear connections between ideas, then the presentation is highly organized. If you cannot perceive any structure to the presentation, then it has low organization.
Fluency	The degree to which speech is smooth and fluid, without inappropriate pauses or starts and stops. If a speaker speaks fluidly, pausing only in appropriate places, then the speech is highly fluent. If a speaker has many inappropriate pauses or hesitations, then the speech has low fluency
Speaking style	The degree to which the speaker uses a dynamic speaking voice. If the speaker uses a range of intonation (changes in pitch) which makes his/her speech attractive, then the speech is highly dynamic. If the speaker changes his/her intonation very little, speaking in a monotone, then the speech is monotonous.

BEFORE YOU BEGIN…

1) You will start by clicking LISTEN. You will hear each speech sample once and only once.

2) You must rate and provide comments for one sample before you can move on to the next one.

3) Ensure that you use the correct scale for each category:

The left side  represents a low score while the right side  a good score

4) Please use the entire scale (left to right: 1 to 1000) as much as you can.

5) You can modify your rating or comments at anytime during the recording and after.

6) When you are finished, click SUBMIT to send your comments, then NEXT.

Appendix D: Treatment Group Native-Speaker Questionnaire – Time 1 (Study 3)

SECTION I – Background Information

1. Name: …….. 2. Gender: ☐ Male ☐ Female

3. Date of birth (DD/MM/YY): ……..… 4. Birthplace (city, country): ……

5. Email: ……..…… 6. Date you started the program (DD/MM/YY): …….….

7. What do you consider to be your native language(s)? ……...

8. Were you exposed to this (these) language(s) since birth? ……..…

9. If not, at what age did you start learning English? ……..…

10. Please rate your ability to speak, listen to, read and write **English** from 1 (extremely poor) to 9 (extremely fluent). Circle the appropriate number.

Speaking	Listening	Reading	Writing
1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9

11. **If you are bilingual**, please rate your ability to speak, listen to, read and write your **second native language** from 1 (extremely poor) to 9 (extremely fluent). Circle the appropriate number.

Speaking	Listening	Reading	Writing
1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9

12. Did you receive any instruction on how to teach ESL or English pronunciation prior to entering the TESL program? If yes, please indicate when, where, for how long, and the kind of instruction received.

…….…….

……..

13. What do you hope to accomplish with this TESL program? ……....

…….....……..…….

…….....……..…….

SECTION II – Beliefs about Pronunciation

1. How important is it for ESL instructors to teach the following knowledge and skills? Please circle a number.

No importance	Almost no importance	Little importance	Unsure or neutral	Moderate importance	Large importance	Extreme importance
1	**2**	**3**	**4**	**5**	**6**	**7**

a. Reading strategies	1	2	3	4	5	6	7	e. Fluency and pronunciation	1	2	3	4	5	6	7
b. Listening strategies	1	2	3	4	5	6	7	f. Grammar	1	2	3	4	5	6	7
c. Vocabulary	1	2	3	4	5	6	7	g. Writing skills	1	2	3	4	5	6	7
d. Conversational skills	1	2	3	4	5	6	7	h. Culture of English-speaking countries	1	2	3	4	5	6	7

2. Please indicate your level of agreement with the following statements by circling a number.

Strongly disagree	Disagree	Tend to disagree	Unsure or neutral	Tend to agree	Agree	Strongly agree
1	**2**	**3**	**4**	**5**	**6**	**7**

a. Teaching pronunciation is often unnecessary, as most learners are able to pick up on pronunciation when frequently exposed to good language input.	1 2 3 4 5 6 7
b. A heavy accent is a cause of discrimination against nonnative speakers.	1 2 3 4 5 6 7
c. Pronunciation is not really teachable – you are either naturally good at it or not.	1 2 3 4 5 6 7
d. For most people, the older you are when you start learning a language, the harder it is for you to acquire native-like pronunciation.	1 2 3 4 5 6 7
e. There is a relationship between learners' perception and production of English speech.	1 2 3 4 5 6 7
f. The goal of pronunciation teaching should be to eliminate, as much as possible, foreign accent.	1 2 3 4 5 6 7
g. It is possible to teach pronunciation communicatively (i.e., through meaningful language use).	1 2 3 4 5 6 7

h. When learners are aware of pronunciation rules, it can help them improve their pronunciation.	1 2 3 4 5 6 7
i. Native speakers should be the model for pronunciation teaching.	1 2 3 4 5 6 7
j. Some pronunciation errors have a greater impact on intelligibility (i.e., understanding a speaker) than others.	1 2 3 4 5 6 7
k. Some individuals resist changing their pronunciation in order to maintain their identity.	1 2 3 4 5 6 7
l. Pronunciation is learned best by trying to imitate good models.	1 2 3 4 5 6 7
m. It is important to know phonology (i.e., the rules of pronunciation) in order to teach English pronunciation.	1 2 3 4 5 6 7
n. Pronunciation teaching should help make students comfortably intelligible to listeners.	1 2 3 4 5 6 7
o. Teaching pronunciation is boring.	1 2 3 4 5 6 7
p. ESL teachers should avoid correcting or pointing out pronunciation errors on the spot.	1 2 3 4 5 6 7
q. English sounds (e.g., the "h" sound in the word "house") are easier to teach than global aspects like stress, rhythm and intonation.	1 2 3 4 5 6 7
r. Pronunciation instruction improves the accuracy of language production rather than communication on the whole.	1 2 3 4 5 6 7
s. Learners benefit from paying conscious attention to the input and becoming aware of how different features are produced.	1 2 3 4 5 6 7
t. An L2 speaker can have a strong foreign accent and still be perfectly understandable.	1 2 3 4 5 6 7
u. It is difficult to integrate pronunciation teaching into regular ESL classes.	1 2 3 4 5 6 7
v. The best person to teach pronunciation is a native speaker.	1 2 3 4 5 6 7

SECTION III – Yourself as a Current or Prospective ESL Teacher

1. Please indicate your level of agreement with the following statements by circling a number.

Strongly disagree	Disagree	Tend to disagree	Unsure or neutral	Tend to agree	Agree	Strongly agree
1	**2**	**3**	**4**	**5**	**6**	**7**

a. I don't have the necessary skills to teach pronunciation.	1 2 3 4 5 6 7
b. I can adequately produce English stress, rhythm and intonation patterns.	1 2 3 4 5 6 7
c. I would feel uncomfortable if the professor observed my pronunciation tutorials at the end of the term.	1 2 3 4 5 6 7
d. If I try really hard, I can make even the most difficult or unmotivated student learn English pronunciation.	1 2 3 4 5 6 7
e. I can adequately produce English sounds (e.g., the "th" sound in "mother").	1 2 3 4 5 6 7
f. I know enough about English pronunciation to teach it effectively.	1 2 3 4 5 6 7
g. I wish I had received more instruction on English pronunciation as a learner.	1 2 3 4 5 6 7
h. I feel anxious about having to teach English pronunciation.	1 2 3 4 5 6 7
i. I know what to do to teach pronunciation effectively.	1 2 3 4 5 6 7
j. As a current or prospective ESL teacher, I feel inferior compared to my peers who are nonnative speakers of the language.	1 2 3 4 5 6 7
k. Even if I try very hard, I can never be an effective pronunciation teacher.	1 2 3 4 5 6 7
l. I have to study more English phonology (i.e., the rules of pronunciation) to feel comfortable teaching pronunciation.	1 2 3 4 5 6 7
m. I am confident that as a teacher I can (or will be able to) help my students improve their pronunciation.	1 2 3 4 5 6 7
n. To be a good pronunciation teacher, I will need to work much harder than my peers who are nonnative English speakers.	1 2 3 4 5 6 7
o. I can (or will be able to) answer my students' questions about English pronunciation.	1 2 3 4 5 6 7
p. I find pronunciation a difficult topic to teach.	1 2 3 4 5 6 7
q. I need to improve my own pronunciation in English before I can teach it to others.	1 2 3 4 5 6 7
r. I know how to make students interested in pronunciation.	1 2 3 4 5 6 7
s. I am a good pronunciation model for ESL students.	1 2 3 4 5 6 7
t. I am (or will be) better at teaching grammar or vocabulary than pronunciation.	1 2 3 4 5 6 7
u. I am afraid that students might "catch" me making pronunciation mistakes when I teach.	1 2 3 4 5 6 7
v. Overall, I have a good grasp of English pronunciation rules.	1 2 3 4 5 6 7

SECTION IV – Further Opinions and Experiences

1. Have you ever studied a second language? ☐ Yes ☐ No (If not, please skip to #6)

2. What language(s) have you studied? ……..

3. As a second language learner, did you ever receive pronunciation instruction? ☐ Yes ☐ No

4. If yes, what types of activities and resources did your teachers use to teach pronunciation?

……

……

5. Do you think the pronunciation instruction you received was effective? Why? ……

……..

……

6. For the pronunciation tutorials at the end of the semester, you and a classmate will be asked to teach pronunciation lessons to an ESL learner together. Would you feel more comfortable if this classmate (i.e., your co-teacher) was a native English speaker or a nonnative English speaker? Why? …….

……..

……..

7. Do you have any experience teaching English? ☐ Yes (Please answer the questions below) ☐ No

8. If yes, for how long? ……..……..…...

9. Where have you taught? Please mention the city and the type of institution (e.g., public school).

……

10. When you teach (taught), on average, approximately what percentage of your weekly class time is (was) dedicated to explicit pronunciation instruction? ……...

11. What types of activities and resources do (did) you usually use to teach pronunciation? ……

……..

Appendix E: Questionnaire Results: Beliefs about Pronunciation Teaching and Learning (Study 3)

	Treatment Group								Comparison Group							
	Time 1				Time 2				Time 1				Time 2			
	D	U	A	M	D	U	A	M	D	U	A	M	D	U	A	M
a. Teaching pronunciation is often unnecessary, as most learners are able to pick up on pronunciation when frequently exposed to good language input.	14 82.4	0	3 17.6	2	14 77.8	2 11.1	2 11.1	1	11 73.3	0	4 26.7	2	5 33.3	5 33.3	5 33.3	3
b. A heavy accent is a cause of discrimination against nonnative speakers.	6 35.3	3 17.6	8 47.1	3	3 16.7	7 38.9	8 44.4	3	7 46.7	3 20	5 33.3	3	7 46.7	1 6.7	7 46.7	3
c. Pronunciation is not really teachable – you are either naturally good at it or not.	18 100	0	0	1	17 94.4	0	1 5.6	0.5	12 80	1 6.7	2 13.3	1	11 73.3	1 6.7	3 20	1
d. For most people, the older you are when you start learning a language, the harder it is for you to acquire native-like pronunciation.	3 16.7	1 5.6	14 77.8	4	1 5.6	1 5.6	16 88.9	5	0	1 6.7	14 93.3	5	0	2 13.3	13 86.7	5
e. There is a relationship between learners' perception and production of English speech.	2 11.1	4 22.2	12 66.7	4	1 5.6	2 11.1	15 83.3	4	0	5 33.3	10 66.7	4	1 6.7	1 6.7	13 86.7	5

f. The goal of pronunciation teaching should be to eliminate, as much as possible, foreign accent.	13 76.5	1 5.9	3 17.6	2	13 72.2	3 16.7	2 11.1	2	10 66.7	1 6.7	4 26.7	2	10 66.7	2 13.3	3 20	2
g. It is possible to teach pronunciation communicatively (i.e., through meaningful language use).	1 5.6	0	17 94.4	5	2 11.8	5 29.4	10 58.8	4	0	3 21.4	11 78.6	4.5	1 7.1	0	13 92.9	4.5
h. When learners are aware of pronunciation rules, it can help them improve their pronunciation.	0	2 11.8	15 88.2	5	0	2 11.1	16 88.9	5	1 6.7	3 20	11 73.3	5	1 6.7	1 6.7	13 86.7	5
i. Native speakers should be the model for pronunciation teaching.	4 22.2	3 16.7	11 61.1	4	4 22.2	5 27.8	9 50	3.5	1 6.7	3 20	11 73.3	4	3 20	7 46.7	5 33.3	3
j. Some pronunciation errors have a greater impact on intelligibility (i.e., understanding a speaker) than others.	1 5.6	2 11.1	15 83.3	5	0	1 5.6	17 94.4	5	2 13.3	1 6.7	12 80	5	1 6.7	2 13.3	12 80	5
k. Some individuals resist changing their pronunciation in order to maintain their identity.	3 16.7	2 11.1	13 72.2	5	1 5.6	4 22.2	13 72.2	4	3 20	6 40	6 40	3	3 20	4 26.7	8 53.3	4
l. Pronunciation is learned best by trying to imitate good models.	1 5.6	5 27.8	12 66.7	4	2 11.1	3 16.7	13 72.2	4	4 26.7	3 20	8 53.3	4	1 6.7	4 26.7	10 66.7	4
m. It is important to know phonology (i.e., the rules of pronunciation) in order to teach English pronunciation.	0	1 5.6	17 94.4	5	1 5.9	1 5.9	15 88.2	6	2 13.3	3 20	10 66.7	4	2 13.3	4 26.7	9 60	4
n. Pronunciation teaching should help make students comfortably intelligible to listeners.	0	0	18 100	5	0	2 11.8	15 88.2	5	2 13.3	0	13 86.7	4	0	2 13.3	13 86.7	5

o. Teaching pronunciation is boring.	17 94.4	0	1 5.6	1	16 88.9	1 5.6	1 5.6	1	10 66.7	3 20	2 13.3	2	7 46.7	3 20	5 33.3	3
p. ESL teachers should avoid correcting or pointing out pronunciation errors on the spot.	7 38.9	4 22.2	7 38.9	3	9 50	4 22.2	5 27.8	2.5	11 73.3	0	4 26.7	2	11 73.3	1 6.7	3 20	2
q. English sounds (e.g., the "h" sound in the word "house") are easier to teach than global aspects like stress, rhythm and intonation.	5 27.8	4 22.2	9 50	3.5	5 27.8	4 22.2	9 50	3.5	5 33.3	3 20	7 46.7	3	4 26.7	3 20	8 53.3	4
r. Pronunciation instruction improves the accuracy of language production rather than communication on the whole.	4 22.2	3 16.7	11 61.1	4	7 38.9	3 16.7	8 44.4	3	2 13.3	7 46.7	6 40	3	5 33.3	5 33.3	5 33.3	3
s. Learners benefit from paying conscious attention to the input and becoming aware of how different features are produced.	0	4 22.2	14 77.8	4	1 5.6	2 11.1	15 83.3	5	0	4 26.7	11 73.3	5	1 6.7	4 26.7	10 66.7	4
t. An L2 speaker can have a strong foreign accent and still be perfectly understandable.	1 5.9	1 5.9	15 88.2	5	3 17.6	0	14 82.4	5	1 6.7	1 6.7	13 86.7	5	2 13.3	1 6.7	12 80	4
u. It is difficult to integrate pronunciation teaching into regular ESL classes.	6 33.3	8 44.4	4 22.2	3	11 61.1	2 11.1	5 27.8	2	10 66.7	4 26.7	1 6.7	2	7 46.7	3 20	5 33.3	3
v. The best person to teach pronunciation is a native speaker.	11 61.1	0	7 38.9	2	9 50	2 11.1	7 38.9	2.5	5 33.3	2 13.3	8 53.3	4	5 33.3	7 46.7	3 20	3

Note. D = Disagree; U = Unsure or neutral; A = Agree; M = Median

Content of unshaded cells = Count and percentage