Second language speakers' awareness of their own comprehensibility: Examining task repetition and self-assessment

Lauren Strachan, Sara Kennedy, and Pavel Trofimovich

Abstract

This study investigated whether second language (L2) speakers are aware of and can manipulate aspects of their speech contributing to comprehensibility. Forty Mandarin speakers of L2 English performed two versions of the same oral task. Before the second task, half of the speakers were asked to make their speech as easy for an interlocutor to understand as possible, while the other half received no additional prompt. Speakers self-assessed comprehensibility after each task and were interviewed about how they improved their comprehensibility. Native-speaking listeners evaluated speaker performances for five dimensions, rating speech similarly across groups and tasks. Overall, speakers did not become more comprehensible from the first to the second task, whether prompted or not. However, speakers who did demonstrate greater improvement in comprehensibility received higher ratings of flow. In addition, speakers' self-ratings of comprehensibility were aligned with listeners' assessments only in the second task. When discussing comprehensibility, speakers commented more on task content than linguistic dimensions. Results highlight the roles of task repetition and self-assessment in speakers' awareness of comprehensibility.

Keywords: comprehensibility; second language; awareness; pronunciation; task repetition; self-assessment

Second Language Speakers' Awareness of their own Comprehensibility:
Examining Task Repetition and Self-Assessment

The general function of speech, including speech by second language (L2) speakers, is to convey speakers' intended meanings to their interlocutors. Opinions on how to do so effectively vary from speaker to speaker. Many L2 speakers report that they wish to sound nativelike (Derwing, 2003). To this end, speakers can adopt an explicit target for learning—a specific language model or models (e.g., General American). Other speakers instead believe that L2 speech should be primarily understandable to listeners, not necessarily nativelike (Ranta, 2010). However, without an explicit model, these speakers may struggle to identify elements of L2 speech associated with listener understanding. For instance, for L2 English, the particular dimensions of speech associated with speakers' comprehensibility (ease of understanding as rated by listeners) appear to depend on speakers' first language (L1) and the complexity of speaking tasks (Crowther, Trofimovich, Isaacs, & Saito, 2015a; Crowther, Trofimovich, Saito, & Isaacs, 2015b). This study's objective was therefore to examine whether L2 speakers are generally aware of how comprehensible they sound to their interlocutor and how they can improve their comprehensibility, in order to explore speakers' awareness of the different dimensions of L2 comprehensibility.

1. Linguistic Dimensions of Comprehensibility

Research on L2 English comprehensibility shows that comprehensibility can be generally linked to two linguistic dimensions (e.g., Isaacs & Trofimovich, 2012; Jułkowska & Cebrian, 2015): pronunciation (accurate production of individual sounds and prosody) and lexicogrammar (varied/appropriate use of words and accurate/complex grammar). However, contextual factors such as speakers' and listeners' L1 background and type of speaking task can also determine which linguistic

dimensions are important for comprehensibility (e.g., Isaacs & Trofimovich, 2012; Saito & Shintani, 2016). In languages other than English, speakers' comprehensibility has also been linked to multiple dimensions, including lexical stress, morphology, fluency, and accuracy in the use of lexis, and pronunciation of individual sounds for L2 German (O'Brien, 2014); fluency, appropriateness and variation of lexis, and placement of pitch for L2 Japanese (Saito & Akiyama, 2017); and pronunciation accuracy, fluency, richness of lexis, and complexity of grammar for L2 French (Bergeron & Trofimovich, 2017). Put simply, L2 comprehensibility is tied to many linguistic dimensions which vary across speakers and tasks, suggesting that speakers might have difficulty judging when and to what extent their speech is comprehensible to listeners.

2. What Do Speakers Know about Comprehensibility?

Early research into L2 comprehensibility was conducted from the interactionist perspective, which posits that interaction offers L2 speakers opportunities for language development through negotiation for meaning (e.g., Swain & Lapkin, 1995). Generally, speakers can and do modify their output to make it more comprehensible, both when interlocutors express difficulty in understanding and when speakers themselves initiate modifications, often targeting L2 semantics and morphosyntax (e.g., Pica, Holliday, Lewis, & Morgenthaler, 1989; van den Branden, 1997). Yet no interactionist study has focused on what speakers think about how or why their speech is more or less comprehensible to interlocutors.

Working from the perspective of language awareness, defined as speakers' "explicit knowledge about language and conscious perception and sensitivity in language learning… and language use" (Association for Language Awareness, 2017), several researchers have examined relationships between speakers' awareness of L2 speech and their comprehensibility. Venkatagiri and Levis (2007) found that 19% of the

variance in L2 English speakers' comprehensibility ratings could be attributed to speakers' awareness of phonological processes (e.g., syllabification of words) and their short-term memory capacity. In a classroom study, Kennedy & Trofimovich (2010) showed that L2 speakers with more comprehensible speech were more likely to discuss their language experiences in qualitative terms (describing language as a way to express meaning and language learning as a means of extracting meaning from a context). In contrast, L2 speakers with less comprehensible speech were more likely to write about their language experiences quantitatively (describing language as a discrete set of items and language learning as a process of assimilating those items).

In a survey of 100 English speakers from different L1 backgrounds (Derwing & Rossiter, 2002), over 50% of respondents identified pronunciation or a combination of pronunciation and language problems as the source of communication breakdowns in their English interactions. For the 61% who identified specific problems, the majority (84%) pertained to individual sounds, with an additional 10% linked to rhythm or intonation. More than half of L2 speakers reported being able to control their speech through explicit focus, changing its pace or clarity, self-repetition, or other strategies. In an interview study (Zoss, 2015), advanced L2 English speakers also cited pronunciation of individual sounds as a source of their interlocutors' understanding difficulty and reported challenges in retrieving vocabulary and grammar forms as a factor influencing their comprehensibility.

Finally, through interviews and observations of interactions in different contexts and with different interlocutors, Rajadurai (2006) explored how three proficient Malaysian speakers of L2 English changed their pronunciation in order to be understood. When speakers believed they needed to be clear, they used intonation and phrasing to chunk information, highlighting prominent or important detail through

phrasal stress and slowing their speech rate. Speakers also changed their pronunciation depending on the listener, with aspiration of voiceless stops being more frequent with non-Malaysian interlocutors, and nonstandard use of glottal stops (e.g., preceding a final consonant) being more prevalent with Malaysian interlocutors. To sum up, L2 speakers appear to demonstrate some awareness of their communication difficulty, linking their comprehensibility to various dimensions of speech, most notably, pronunciation of individual sounds. However, it remains unclear how these beliefs and their associated behaviours are related to interlocutor understanding.

3. Comprehensibility and Speakers' Self-Assessments

Research into L2 speakers' self-assessment of their performance, including comprehensibility, provides further insight into speaker awareness of comprehensible speech. Speakers' self-assessment in oral tasks is often weakly or not at all related to instructors' evaluations of the same tasks (Lee & Chang, 2005; Patri, 2002). However, others have reported medium or strong associations between speakers' self-assessment and instructors' assessments, particularly after training in self-reflection and self-assessment (Babaii, Taghaddomi, & Pashmforoosh, 2016; Chen, 2008). Qualitative analyses of speakers' unstructured self-assessments revealed multiple criteria that were used: content, fluency, comprehensibility, organization, use of language forms, task and time management (Babaii et al., 2016; Dlaska & Krekeler, 2008; Kissling & O'Donnell, 2015; Lee & Chang, 2008). And speakers who self-assessed their speech over time noticed improvement in fluency, pronunciation, connected speech processes, organization, and vocabulary use (de Saint Leger, 2009; Kissling & O'Donnell, 2015). L2 speakers' self-assessments appear to depend on their skill level, in line with research in math or science (e.g., Dunning, Johnson, Ehrlinger, & Kruger, 2003). Speakers at lower levels of L2 comprehensibility overestimate how easily listeners understand them,

while speakers at higher levels underestimate listeners' ease of understanding (Trofimovich, Isaacs, Kennedy, Saito, & Crowther, 2016). In addition, the discrepancy between L2 speakers' self-assessment and listeners' evaluations seems to vary for different aspects of speech. In L2 French (Lappin-Fortin & Rye, 2014), instructors' assessments of vowel production and prosody were unrelated to speakers' self-assessments, but instructors' evaluations converged with speakers' self-assessments when the use of connected speech processes was targeted. And the relative difference between speaker and listener assessments of comprehensibility also seems to be linked to linguistic dimensions that are specific to L1 groups, such as individual sounds and intonation for L1 Chinese speakers and word stress and rhythm for L1 Hindi speakers (Crowther et al., 2015b). In sum, L2 speakers' self-assessment of their speech can be consistent with the assessment of other listeners, particularly after receiving self-assessment training. However, many speakers misjudge how aspects of their speech (e.g., pronunciation) are viewed by listeners, suggesting that they cannot identify which linguistic dimensions affect how easily listeners understand them.

4. Comprehensibility and Task Repetition

The benefits of task repetition for L2 speakers' performance in oral tasks are well established (e.g., Ahmadian & Tavakoli, 2011; Lynch & Maclean, 2000). However, it is unknown whether speakers can become more comprehensible to their interlocutor simply through the practice of performing the same task more than once. As Jung, Kim, and Murphy (2017) outline, task repetition research comprises three types of repetition: exact repetition (same task procedure and content repeated), procedural repetition (same task procedure repeated with new content), and content repetition (same task content repeated with a new procedure). To date, most research on oral task repetition has examined speakers' performance through measures of speech complexity,

accuracy, and fluency (e.g., Ahmadian & Tavakoli, 2011; Kim & Tracy-Ventura, 2013). However, these measures describe only the linguistic dimensions of speech; they do not reveal listeners' or speakers' reactions to task performance nor show how speakers modify their strategies or discourse organization in completing the task.

The few task repetition studies with listener-based or task performance measures were conducted in classroom settings. In a study on L2 English speakers' repeated poster presentations to a rotating audience, Lynch and Maclean (2000) found that one speaker adopted more targetlike lexis and pronunciation that she had heard from interlocutors, though the speaker was unaware of making any changes to her speech. Another speaker modified her explanation of an unfamiliar statistical technique, rephrasing and reorganizing her successive descriptions. In a study combining task repetition and self-assessment, Castañeda and Rodríguez-González (2011) had L2 Spanish speakers record four prompt-elicited monologues over a semester and self-assess their performance after each recording. The instructor, who rated the recordings, did not notice substantial changes in task performance. But the researchers, who holistically analysed the recordings, noted that repeated tasks "contained new vocabulary, new language chunks, elaboration and linking of ideas, and improved communicative strategies" (p. 490). In their self-assessments, speakers commented most frequently on grammar, pauses/hesitancy, and pronunciation, with comments on early tasks emphasizing linguistic form (e.g., verb tenses), speaking rate, and pronunciation, while comments on later tasks focused on fluency and strategies for communicative effectiveness. In a final survey, speakers reported that task repetition led to improved performance, in terms of content, organization, and grammar.

5. The Current Study

Through decades of sustained research attention, L2 speakers' comprehensibility has emerged as a complex, multidimensional construct, involving not simply pronunciation but other aspects of language and speech (e.g., Isaacs & Trofimovich, 2012). Speakers' awareness of their own comprehensibility has sometimes reflected this multidimensional nature, yet when speakers assess their own comprehensibility, their assessments often diverge from those provided by listeners or do not reflect the many dimensions of language and speech which matter for comprehensibility (e.g., Lappin-Fortin & Rye, 2014). It appears therefore that many L2 speakers may not be fully aware of, or proficient in, ways to make their speech more comprehensible, often despite years of language instruction and exposure. In this study, we evaluated this claim by providing L2 speakers the opportunity to reflect on their comprehensibility through self-assessment and also to repeat the same speaking task with new content, which can allow speakers to shift their focus to aspects of language, speech, or the task which continue to be challenging for comprehensibility (e.g., Castañeda & Rodríguez-González, 2011). A subset of speakers was also prompted to consciously consider how to make their speech comprehensible before repeating the task, to provide additional insight into speakers' awareness of comprehensible speech. Speakers' performances from both tasks were rated by English-speaking listeners for comprehensibility and four other linguistic dimensions (segmental errors, word stress errors, intonation, flow), so that speakers' performances and self-assessments could be evaluated in reference to listener judgments. Finally, all speakers were interviewed to probe their understanding of comprehensible L2 speech. Motivated by the overall goal of providing insight into L2 speakers' awareness of comprehensible speech, this study was guided by the following three sets of questions:

(1) Does repeating the same task, with or without the prompt to sound as

comprehensible as possible, result in perceptible differences in listener ratings of

L2 speakers' comprehensibility? Which linguistic dimensions in speakers'

speech are associated with such differences?

(2) Do L2 speakers' self-ratings of comprehensibility become more aligned with

listener-based assessments as a function of repeated task, with or without the

prompt to sound as comprehensible as possible?

(3) How do L2 speakers understand the concept of making L2 speech more

comprehensible for the listener?

5.1 Method

    5.1.1. L2 Speakers. L2 speakers were 40 international students enrolled in

degree programs at the BA (17), MA (20), and PhD (3) levels in an English-

medium university in [location blinded]. The speakers, who were born and raised

in Mainland China (39) or Taiwan (1), grew up in unilingual homes speaking

Mandarin Chinese (38) or Cantonese (2). Because the two Cantonese-speaking

participants did not perform differently from the remaining speakers, their data

were included in the total sample. As, all participants had been admitted to degree

programs, they thus met the university's proficiency requirement of International

English Language Testing System (IELTS) band score of 6.5 or higher.

    The 40 speakers were randomly assigned to two equal groups

(unprompted, prompted), which differed only in the type of instructions received

before the second task. Speakers in the prompted group were instructed to make

their speech as easy for the interlocutor (researcher) to understand as possible,

while speakers in the unprompted group received no additional instructions

targeting their comprehensibility. As shown in Table 1, the two groups did not

differ in any background or language variables, $t(38) < 1.11$, $p > .27$.

Table 1.

5.1.2. Tasks and Procedure. L2 speakers performed two versions of the

same graph-based interpretation task, adapted from practice materials (Papajohn,

2005) for the retired Test of Spoken English, which was used to examine

international students' speaking skills before the introduction of TOEFL iBT

(ETS, 2011). The task requires speakers to study a data graph on a particular topic

and then give a 1-minute summary of the information portrayed in it. Two

versions of the task were used (see Appendix A): one focusing on student

admissions at a university across four decades of enrolment and the other

targeting expenditures for road construction and maintenance, again across four

decades. Compared to non-academic tasks, such as picture descriptions, a graph-

based interpretation task was considered to be suitable for eliciting university

students' speech in an academic setting.

To control for order effects, half the speakers performed two versions of the task

in one order (enrolment followed by road construction), while the other half completed

them in the opposite order (road construction followed by enrolment). The first task

(Task 1), was introduced in the same manner to all speakers, who were asked to

examine a printed graph for 15 seconds in order to formulate their response and then

had 1 minute to produce their summary. After completing Task 1, speakers self-rated

their performance for accentedness (not reported here) and comprehensibility.

Comprehensibility was defined as amount of effort it takes to understand what someone

is saying, and speakers were reminded that comprehensibility is separate from

accentedness, in that it is possible for a speaker to have an accent and be very easy to understand, or to have little accent and still be difficult to understand. Each speaker's 1-minute recording was played back to him or her so that the speaker could use a 1,000-point sliding scale, programmed in custom-designed software, to self-rate his or her performance. The scale endpoints were labeled as "hard to understand" (equivalent to the rating of 0) and "easy to understand" (equivalent to the rating of 1,000), but the scale contained no interval or numerical markings. After completing their self-ratings, speakers used paper-based 9-point scales to estimate how difficult the speaking task was for them and how difficult it was for them to rate their speech (1 = "very easy," 9 = "very difficult"), and finally how well they understood the rated category of comprehensibility (1 = "did not understand it at all," 9 = "understood very well").

After a brief filler task, speakers performed the second version of the task (Task 2). The filler task was a 604-word reading passage on dinosaur extinction followed by seven multiple-choice comprehension questions, which took approximately 15 minutes to complete. The passage was adapted from an online TOEFL practice test (https://www.ets.org/Media/Tests/TOEFL/pdf/SampleQuestions.pdf) to keep with the academic theme of the study procedure and to allow for a short break between two repetitions of the target tasks. The only procedural difference for Task 2 was the instructions given to the two speaker groups. Speakers in the unprompted group received no instructions in addition to those provided for Task 1, while speakers in the prompted group heard the following instructions from the researcher as she introduced Task 2: "This time, I want you to make your speech as easy to understand as you can. I want to be able to easily understand your speech while you are doing the task." The remainder of Task 2 proceeded in the same manner as in Task 1, with speakers recording their 1-minute summary, then self-rating their performance, and finally

estimating task and rating difficulty and their understanding of the comprehensibility

construct. Speakers' speech in both tasks was audio recorded directly onto a computer

using a high-quality microphone and, as discussed above, was immediately available for

self-rating. At the end of the session, each speaker was interviewed individually, using

eight guiding questions (see Appendix B) that probed speakers' performance in each

task, their conceptions of comprehensible speech, and the changes they might have

introduced to their speech in Task 2.

     5.1.3. Listeners and Listening Procedure. L2 speakers' performance in Tasks

1 and 2 was evaluated by 10 listeners (8 female, 2 male), all native English speakers

(M_{age} = 33.6 years, SD = 5.4). All listeners were born and raised in monolingual English

families in Canada or the United States and had been schooled in English, with two

reporting bilingual English-French education at the elementary and high school levels.

All but one listeners were enrolled in, or had recently completed, MA or PhD programs

in applied linguistics or education and were residents of [location blinded], where

speakers pursued university studies. Of the 10 listeners, nine had L2 teaching

experience (M = 5.4 years, SD = 4.1), and all reported high rates of daily English

speaking (M = 80.0%, SD = 15.6) and familiarity with L2-accented English (M = 8.1,

SD = 1.5) on a 9-point scale (1 = "not familiar," 9 "very familiar"), including by

speakers of Mandarin and Cantonese.

     Listeners evaluated excerpts from each speaker's recordings of Tasks 1 and 2.

The 80 recordings (40 per task) were first matched in amplitude; then, initial

dysfluencies were removed and the full-length files were truncated to the initial 30

seconds of speech produced, which is consistent with previous research using 20–60

second samples for listener judgments (e.g., Derwing, Rossiter, Munro, & Thomson,

2004). Listeners evaluated audio files individually, providing two sets of ratings across

two rating blocks: (a) for accentedness (used here only for description of speaker

groups, as shown in Table 1) and comprehensibility, and (b) for four specific

pronunciation variables (segmental errors, word stress errors, intonation, flow), all

summarized in Table 2. Listeners used similar 1,000-point sliding scales to those used

by speakers to self-rate their performance. The endpoints in each scale were only

labelled with verbal descriptions, and the scale contained no interval markings.

Accentedness and comprehensibility scales appeared on screen together, and listeners

were required to listen to each audio once before providing their ratings, on the

assumption that these variables reflect initial, intuitive perceptual judgments. The scales

for the remaining four variables (segmental errors, word stress errors, intonation, flow)

also appeared together, but listeners could replay each audio multiple times before

providing their ratings. For both sets of ratings, each listener experienced all 80 audio

files (from both tasks and both speaker groups) in a unique randomized order per

listener.

Table 2.

Listeners were first shown the materials for each task, and then instructed about

each rating category using definitions and examples. They evaluated four extra practice

samples before proceeding to rate the target audio files (using headphones), with a 10–

15 minute break between the two blocks. After the rating was completed, listeners filled

out the final debrief questionnaire, where they used 9-point scales to rate how well they

understood each rating category (1 = "did not understand at all," 9 = "understood this

concept well") and how easily they could evaluate each rated category (1 = "very

difficult," 9 = "very easy/comfortable").

5.1.4. Data Analysis. Initial checks of the internal consistency of all listener-

based ratings revealed that listeners were generally consistent in their evaluations. In all

cases, Cronbach's *a* values surpassed the benchmark values of .70–.80 (Hall, 2009): *a* = .79 (comprehensibility), *a* = .76 (segmental errors), *a* = .71 (word stress errors), *a* = .71 (intonation), and *a* = .88 (flow). Listeners also rated their understanding of each rated category as being on average at or above 8.0 on a 9-point scale: 8.0 (comprehensibility), 8.8 (segmental errors), 8.5 (word stress errors), 8.5 (intonation), and 8.7 (flow). They also reported feeling generally at ease when using these categories, with a mean rating of 7.5 or above on a 9-point scale: 7.5 (comprehensibility), 8.4 (segmental errors), 7.6 (word stress errors), 7.5 (intonation), and 8.5 (flow). Therefore, individual speaker-based scores were computed by averaging across all 10 listeners' ratings for each measure, separately for Tasks 1 and 2.

The first research question had two parts. To answer the first part—whether self- and listener comprehensibility ratings were affected by task repetition or prompts to be more comprehensible—comprehensibility scores were submitted to a two-way mixed ANOVA, with group (prompted, unprompted) as between-subjects factors and task (task 1, 2) as within-subjects factors. To answer the second part—whether linguistic dimensions (segmental errors, word stress errors, intonation, flow) were associated with changes in comprehensibility ratings from task 1 to task 2—difference scores were obtained (task 2 – task 1) for listener comprehensibility ratings and a Pearson *r* correlational analysis with the rated linguistic dimensions was performed. The second research question asked whether self- and listener comprehensibility ratings became more aligned as a function of task repetition and/or prompts to be more comprehensible. To answer this question, difference scores were obtained by subtracting listener ratings from self-ratings. Low difference scores indicated high alignment between listener and self- comprehensibility ratings. Difference scores were submitted to a two-way mixed ANOVA with group (prompted, unprompted) as between-subjects factor and task (1, 2)

as within-subjects factor. Next, in order to better understand individual differences in alignment, self- and listener comprehensibility ratings were submitted to correlational analyses (Pearson *r*).

The debrief interviews (Appendix B), which were audio-recorded, were reviewed by the first author, who conducted the interviews and has an MA degree in applied linguistics and has completed graduate-level courses in research methodology. For the current report, only the responses to the three most relevant sets of questions were analysed: (a) Was your speech easier to understand in Speaking Task 1 or Speaking Task 2? (b) Did you change your speech from the first to the second recording? Why? How? and (c) Do you make any changes to your speech when you want others to understand you easily? What changes do you make? Why/why not? The first author reviewed and transcribed participants' responses question by question. A review of 20% of the transcripts by the second author showed 90% accuracy in transcribing, which was deemed acceptable. Thematic coding of responses for each question was done through empirical coding drawn from the content of the transcripts (Gibson & Brown, 2009). In an iterative process, the first author derived codes for themes and subthemes from the transcribed responses for each question. The second author reviewed the coding, suggesting some modifications to the coding of certain themes and subthemes, which was then commented on by the first author. This process continued until there was full consensus on the themes and subthemes and on coding decisions.

6. Results

6.1. Comprehensibility Across Repeated Tasks

The first analysis targeted the first research question, which asked whether repeating the same task, with or without the prompt to sound as comprehensible as

possible, resulted in perceptible differences in listener ratings of speakers'

comprehensibility. A two-way mixed ANOVA, with group (prompted, unprompted) as

a between-subjects factor and task (Task 1, Task 2) as a within-subjects factor revealed

no statistically significant main or interaction effects, $F(1, 38) < 0.99$, $p > .32$, $\eta_\rho^2 < .03$,

suggesting that speakers were evaluated similarly in comprehensibility across the two

tasks, regardless of the prompt received. The ANOVAs targeting the remaining rated

dimensions (segmental errors, word stress errors, intonation, and flow) similarly

revealed no significant main or interaction effects, $F(1, 38) < 1.59$, $p > .21$, $\eta_\rho^2 < .04$,

indicating that speakers were also evaluated similarly across tasks in these linguistic

variables (see Table 3).

Table 3.

6.2. Comprehensibility and Linguistic Dimensions

Despite the absence of group-based differences in speakers'

comprehensibility, there was variability in individual speakers' performance.

Some speakers received lower comprehensibility ratings in Task 2 than in Task 1,

while others were evaluated as being more comprehensible in Task 2 than in Task

1. For example, in the unprompted group, Task 1–Task 2 comprehensibility

difference scores (Task 1 score subtracted from Task 2 score) ranged between –

307 and +202 on a 1,000-point scale, meaning that some speakers were evaluated

by as much as 307 points lower in Task 2 than in Task 1 but others received a

score by 202 points higher in Task 2 compared to Task 1. In the prompted group,

there were similar differences, between –194 and +237 points.

To examine whether individual variation in speakers' comprehensibility across

tasks might have been associated with different aspects of their linguistic performance,

the next analysis focused on associations (Pearson r) between the extent of speakers'

Task 2–Task 1 comprehensibility differences and their performance in the remaining

four rated linguistic dimensions (segmental errors, word stress errors, intonation, and

flow) in Task 1 and specifically in Task 2, where a change in speech performance was

likely to occur due to task repetition. These analyses, summarized in Table 4,

demonstrated that the extent of Task 2–Task 1 comprehensibility differences was only

associated with speakers' performance in the dimension of flow, and only in Task 2 ($r =$

.41, $p = .008$). Put simply, the speakers who showed greater improvement in

comprehensibly were also those who received higher flow ratings in Task 2.

Table 4.

The relationship between Task 2–Task 1 comprehensibility differences and flow

ratings, which was medium in strength according to Plonsky and Oswald's (2014)

guidelines, is depicted in Figure 1. However, if the data for one outlier whose Task 2–

Task 1 difference score was –307 (shown by the lowest datapoint in Figure 1) is

removed, the association increases to $r = .54$, which approaches the benchmark for a

strong correlation (.60). One interpretation of these findings is that the speakers who

showed greater improvements in their comprehensibility from Task 1 to Task 2 might

have done so through production of more fluent speech in Task 2 (i.e., speech rated by

listeners as more natural and more fluid).

Figure 1.

6.3. Self- and Listener-Rated Comprehensibility

The next analysis focused on the second research question, which asked whether

L2 speakers' self-ratings of comprehensibility become more aligned with listener-based

assessments as a function of repeated task, with or without the prompt to sound as

comprehensible as possible. For this analysis, rating difference scores for

comprehensibility were computed first (listener-based ratings subtracted from self-ratings), separately for each task and group. Positive difference scores represented speakers who overestimated their comprehensibility, while negative scores designated speakers underestimating their comprehensibility, relative to listener-based assessment. In general, compared to listeners' evaluations, speakers' self-ratings tended to be lower in all cases—in the unprompted group in Task 1 ($M = -217$, $SD = 194$) and in Task 2 ($M = -189$, $SD = 161$) and in the prompted group in Task 1 ($M = -194$, $SD = 306$) and in Task 2 ($M = -191$, $SD = 254$)—meaning that speakers underestimated their comprehensibility by an average of 200 points on a 1,000-point scale. To compare rating difference scores, these scores were submitted to a mixed ANOVA, with group (prompted, unprompted) as a between-subjects factor and task (Task 1, Task 2) as a within-subjects factor. This analysis yielded no significant main or interaction effects, $F(1, 38) < 0.28$, $p > .60$, $\eta_p^2 < .01$, suggesting that the extent to which speakers' self-ratings differed from the assessment of listeners did not change from Task 1 to Task 2 and did not differ between prompted and unprompted groups.

However, as shown by large standard deviation scores, speakers again varied widely in the degree to which they diverged in their self-assessments from listeners' ratings of comprehensibility. Individual speakers' performance was examined through correlation analyses (Pearson r) computed between speakers' self-ratings and listener-based assessments of comprehensibility ($n = 40$). In Task 1, there was no relationship between the two sets of ratings ($r = .10$), which suggested that speakers' perceptions of their comprehensibility were independent from listeners' assessments of this construct. However, in Task 2, the association between self- and listener-based assessments was medium in strength ($r = .40$, $p = .009$), meaning that those speakers who tended to self-assess comprehensibility higher were also evaluated as being easier to understand by

listeners. As shown in Figure 2, which illustrates this association, the link between self-
and listener-assessed comprehensibility was generally weak, but it seemed more
pronounced at the lower end of the comprehensibility continuum. In fact, non-linear
(cubic) regression function fitted to the data (dotted line in Figure 2) appeared to fit the
data better, resulting in improvement in overall fit (R^2 = .23 vs. R^2 = .16) and revealing
a stronger association for speakers of lower comprehensibility (steeper regression line)
and a weaker, non-existent association for speakers of higher comprehensibility
(plateaued, levelled-off line). Put simply, the speakers at a lower level of
comprehensibility appeared to be more in sync in Task 2 with external listeners in terms
of comprehensibility assessment, compared to speakers whose comprehensibility was
greater. This is noteworthy given that no such relationship was found in Task 1.

Figure 2.

6.4. Speakers' Awareness of L2 Comprehensibility

The final analysis targeted the third research question, which focused on
speakers' understanding of the concept of making L2 speech more comprehensible for
the listener. In the prompted group, 12 out of 20 speakers believed their speech was
easier to understand in Task 2 than in Task 1. The pattern was the opposite in the
unprompted group, where 11 speakers thought they were more comprehensible in Task
1 than in Task 2. When asked what they had changed about their speech in Task 2 (see
Table 5), prompted speakers reported changing how they organized information within
utterances or the entire presentation (4 speakers), how they described particular content,
such as graph titles or numerical values (2 speakers), or mentioned using simpler words
(4 speakers) and sentences (2 speakers) and sounding generally more fluent (1 speaker).
Unprompted speakers also reported changing the organization of elements of the
presentation or making explicit connections between elements (4 speakers), adding new

content (2 speakers), and using different or more technical vocabulary (2 speakers) or

speaking in simpler utterances (1 speaker). Thus, at least half of the speakers in each

group who reported making changes mentioned the organization of individual sentences

or the presentation structure and reported changing the type or complexity of words and

sentences.

Table 5.

Speakers were also asked whether they ordinarily made any changes to their

speech in order to be easily understood. In the prompted group, 17 out of 20 speakers

reported that they made changes, such as using more frequent vocabulary (7 speakers),

simplifying sentence structure (4 speakers), changing their volume and rate of speech,

such as speaking more loudly and more slowly (2 speakers), changing the lexical stress

of words (1 speaker) and attempting to speak with a native speaker accent (1 speaker).

Two speakers mentioned that they provide more information, and one speaker stated

that he organizes information into discrete parts. In terms of general strategies, three

speakers reported rephrasing or reformulating utterances, and two mentioned external

aids, such as drawing pictures, using dictionaries, or asking someone else to explain.

In the unprompted group, 15 out of 20 speakers reported making changes to

their speech to increase understanding, by using task-relevant or frequent vocabulary (6

speakers), producing simpler or more appropriate sentence structure (2 speakers),

attempting to pronounce individual sounds clearly (1 speaker), or increasing overall

clarity of expression (2 speakers). Two speakers described providing extra details or

many examples, and one described using discourse markers to show organization. For

general strategies, two speakers reported reformulating or rephrasing unclear speech,

one reported preparing and rehearsing speech, and three reported use of external aids,

such as notes, pictures, or dictionaries. In sum, across both groups, speakers' most

frequent changes to improve comprehensibility included changes to the lexical content and organization of their speech, rephrasing of grammar structures, and various strategies to clarify the content of their speech.

7. Discussion

This study investigated the role of task repetition in comprehensibility and examined speaker awareness of their own comprehensibility. Results showed that flow may have contributed to increased comprehensibility in Task 2, that speakers with lower comprehensibility show better alignment with listeners' comprehensibility ratings after task repetition, and that (when asked) speakers identify a range of factors which may affect their own comprehensibility.

7.1. Task Repetition and Comprehensibility

The first research question asked whether repeating the same task, with or without the prompt to sound as comprehensible as possible, results in perceptible differences in listener ratings of L2 speakers' comprehensibility and whether these differences are associated with specific linguistic dimensions in speakers' speech. Results showed that there was no difference overall in listener comprehensibility ratings between tasks (i.e., whether speakers were evaluated in Task 1 or Task 2) or between groups (i.e., whether speakers were prompted to be more comprehensible or not). Thus, an increase in L2 comprehensibility (with or without a prompt for speakers to sound as comprehensible as possible) was generally unlikely to occur after procedural repetition involving a brief oral task. There was also no difference in listener ratings of segmental errors, word stress errors, intonation, and flow between the groups or tasks. Therefore, listener-identified improvement in L2 speakers' oral performance (e.g., in lexis, content and organization, and communication strategy use) likely requires more intensive and

extensive task repetition—such as multiple oral presentations (Lynch & Maclean, 2000) or four prompt-elicited monologues delivered over an academic term (Castañeda & Rodriguez-González, 2011)—compared to a single task repeated here.

The absence of group effects for task repetition could also be attributed to a variety of factors, including individual differences in speakers' working memory capacity (Robinson, 2011) or their language analytic ability (Skehan, 2015). For example, through task repetition, speakers with greater processing capacity and those with greater sensitivity to linguistic structure in speech might be more likely to improve their comprehensibility, compared to those with less processing capacity or less sensitivity to linguistic structure. Additionally, different from this study, most previous research showing effects of task repetition also included some form of instruction (Chen, 2008), whose effects might amplify speakers' awareness of various language forms. Moreover, the absence of group differences might also be task related. Although the effect of task type on L2 speakers' comprehensibility has been attested for both L2 English and L2 French (Bergeron & Trofimovich, 2017; Crowther et al., 2015a), the particular academic task used here (graph-based interpretation) may have been more familiar to some speakers than to others. Without task-focused instruction, some participants may have gained little benefit from repeating an otherwise unfamiliar task. While the task remained the same (i.e., its procedure was repeated), the content varied slightly (*road construction* vs. *university enrolment*). This may have increased the cognitive load for the second (repeated) task, in that participants would have to conceive of new vocabulary, despite similarities in task format. However, this was unlikely as participants were asked to describe the graph both times, rather than to elaborate on the topic, and the trends depicted in each graph were similar. In other words, new vocabulary was indeed necessary for each task (i.e., *construction* vs.

enrolment), but the vocabulary most necessary for task completion (i.e., *gradual increase*, *slight decrease*) remained largely similar. Finally, as was pointed out by one reviewer, the finding that participants did not improve from Task 1 to Task 2 may have implications for research, in that perhaps researchers need not be extremely concerned with practice effects. This reviewer also pointed out that it is possible that speakers are generally performing their best, whether prompted or not.

As was shown descriptively, individual speakers' comprehensibility ratings varied greatly between tasks, with some speakers perceived as being more and others as less comprehensible in Task 2 than in Task 1. However, the speakers who were rated more comprehensible in Task 2 compared to Task 1 also received high ratings for flow in Task 2, implying a possible link between listener-rated fluency on the second task and speakers' improvement in comprehensibility from Task 1 to Task 2. Links between L2 speakers' flow or fluency and their comprehensibility have been well established in prior work (e.g., Derwing et al., 2004; Saito, Trofimovich, & Isaacs, 2016), yet the current results provide, to our knowledge, one of the first demonstrations of a relationship between L2 speakers' fluency and *improvement* in their comprehensibility.

The linguistic dimension associated with improvement in comprehensibility— for some speakers in Task 2—was flow (see Figure 1). An increase in flow for these speakers might be due to faster online planning through freeing up of cognitive resources used for planning during the first task performance (e.g., Bygate, 2001). Given the brief gap between the two versions of the same task, flow or fluency would seem to be the most straightforward dimension to change, as speakers did not have the opportunity to reflect on or analyze their first task performance before repeating the task. It is also possible that speakers focused on dimensions of speech which were not part of listeners' rating scales or judgments of comprehensibility. For example, in

debriefing interviews, speakers often mentioned idea organization or accurate graph descriptions rather than speech-based constructs. If speakers made changes in organization or content accuracy, those changes were not noticed or taken into account by listeners.

7.2. Self-Assessment and Comprehensibility

The second research question asked whether speakers' self-ratings of comprehensibility become more aligned with listener-based assessments as a function of repeated task, with or without a prompt for comprehensibility. Overall, speakers from both groups across both tasks underestimated their comprehensibility relative to listener ratings. This is similar to Lee and Chang's (2005) findings, where participants underrated their presentation abilities compared to teacher- and peer-ratings, but counter to other pronunciation self-assessment studies, where speakers' self-ratings corresponded to (Dlaska & Krekeler, 2008), or were more lenient than (Lappin-Fortin & Rye, 2014), listener ratings. Low achievers have also been shown to overestimate, while high achievers underestimate, their oral presentation skills (Patri, 2002; Trofimovich et al., 2016), consistent with broader education and psychology literature (e.g., Dunning et al., 2003). Possible reasons for the differences between the current findings and results of prior research could be due to the way self-assessment was conducted. In this study, speakers were asked to self-rate using a global pronunciation measure, such as comprehensibility, rather than being provided a rubric or training for self-assessment or having the chance compare their performance to peers (Babaii et al., 2016; Chen, 2008; Lappin-Fortin & Rye, 2014). Without training, a specific rubric, or a baseline comparison, speakers were likely self-assessing based only on their own perceptions. Perhaps most importantly, speakers were not self-assessing their L2 speech but were estimating the judgment of their speech by unknown listeners. Given that

comprehensibility can be affected by listener characteristics (Crowther, Trofimovich, & Isaacs, 2016), speakers may have assumed that the (imagined) listeners—unlike the actual listeners—would have little familiarity with L2 English.

Although no overall alignment was found between self- and other-assessment, there was a medium-strength correlation between speaker and listener assessments of comprehensibility, and this association was stronger for speakers who had received low comprehensibility ratings. Put differently, those whose self-ratings were closest to listener ratings were also those with the lowest comprehensibility. This relationship was true only for Task 2, which implies that speakers might become attuned to their level of comprehensibility to listeners through some form of repeated practice. One source of repeated practice would be speakers' experience of repeating the same task procedure. In this case, the positive effect of task repetition would be akin to the positive role of peer-assessment practice or training in self-reflection in helping speakers become more aware of how their self-assessments align with the assessments of others (Babaii et al., 2016; Chen, 2008; Lappin-Fortin & Rye, 2014; Patri, 2002). Another source of repeated practice would be speakers' experience of engaging in self-assessment more than once (Kissling & O'Donnell, 2015), which could make speakers more sensitive to how easily their speech is understood by others. Whatever the source of the association between self- and other-assessment, the current findings imply that procedural task repetition, coupled with repeated self-assessment, has the potential to assist speakers in aligning their perceptions of comprehensibility with those of other listeners.

7.3. Speakers' Awareness of Comprehensibility

The third research question asked how L2 speakers understand the concept of making L2 speech more comprehensible for the listener. Responses to interview questions indicated that speakers associated many factors with their understanding of

comprehensibility. With respect to the specific factors (see Table 5), speakers

mentioned many pronunciation and lexis/grammar features identified in prior research

as feeding into listener judgments of comprehensibility (e.g., Isaacs & Trofimovich,

2012). Nevertheless, awareness was quite varied across speakers and did not generally

reflect the quantitative results which showed a link between improved

comprehensibility and flow (fluency) in Task 2.

      Overall, the Task 2 changes reported by speakers from both groups focused on

their use of vocabulary and organization of sentence and presentation structure. These

aspects of language use were not explicitly rated by listeners, so any relationship

between them and speakers' comprehensibility could not be determined. Rossiter (2009)

found that both temporal (e.g., speech rate) and non-temporal aspects of speech (e.g.,

grammar, vocabulary) contributed to listeners' fluency ratings for L2 English speakers.

So changes in speakers' vocabulary use as well as organization and structure of their

output could have contributed to their flow or fluency, which was a rated aspect linked

to speakers' comprehensibility.

      Generally, speakers' reports of how they *typically* change their speech to be

more comprehensible to listeners were not particularly different between the two

groups, comprising a range of different aspects. The prompted group more frequently

cited syntax and pronunciation as aspects they changed, but speakers in both groups

mentioned multiple aspects. Overall, speakers showed awareness that different

dimensions of speech and language use can contribute to comprehensibility. Speakers

seldom mentioned flow or fluency in particular, but many of the other dimensions they

identified were not rated by listeners, so it is unclear how these dimensions would be

related to listeners' comprehensibility judgments. One unambiguous finding is that,

overall, speakers were unable to significantly increase their comprehensibility for

listeners simply by repeating a task or by trying to speak very comprehensibly, which has implications for instruction.

7.4. Pedagogical Implications

The current findings support the use of task repetition and self-assessment for language learning and teaching in several ways. First, for some speakers, flow/fluency was associated with better comprehensibility as task procedure was repeated. Thus, repetition of fluency-based activities focusing on meaning rather than form (Ellis, 2005) may help L2 speakers increase both their fluency and comprehensibility. Second, through repeated task and/or repeated self-assessment, some speakers (especially those with lower comprehensibility scores from listeners) seemed to become aligned with listeners in self-assessing their comprehensibility. This finding supports the use of task repetition and self-assessment in classes which include a focus on L2 speech, even for low-proficiency speakers. As positive effects have been reported in self-assessment studies where speakers are explicitly instructed on self-assessment or language (Babaii et al., 2016) or provided an evaluation rubric (Casteñeda & Rodriguez-González, 2011), adding these elements to self-assessment activities, which may or may not include task repetition, could boost their effectiveness. Overall, it is heartening that speakers showed awareness of many domains contributing to their comprehensibility. However, neither group was able to significantly increase their comprehensibility through task repetition, which implies that L2 speakers would benefit from consistent and close analysis, by themselves and by others, of their comprehensibility and of how they produce various aspects of L2 speech.

7.5. Limitations and Future Directions

This study investigated the role of task repetition and self-assessment in comprehensibility and speaker awareness of comprehensibility. The lack of overall

change in comprehensibility might be partly explained by the absence of detailed evaluation rubrics or speaker training in self-assessment, which should be included in future research. In addition, the absence of strong relationships between ratings of comprehensibility and measures of linguistic dimensions of speech could reflect the particular linguistic group targeted (Mandarin/Cantonese speakers of L2 English) and the restricted range of speech measures used, which excluded measures of vocabulary, grammar, and discourse structure and organization. Future studies might explore how these and other dimensions of oral output contribute to comprehensibility through task repetition and self-assessment for other groups of L2 speakers. Last but not least, results of this study may reflect the particular group of listeners, who (being MA and PhD students in education and applied linguistics) were familiar with accented speech; in fact, expert listeners (those with training in applied linguistics and familiarity with the target accent) have been shown to provide more lenient speech ratings compared to novice listeners (e.g., Saito, Trofimovich, Isaacs, & Webb, 2017). L2 speakers' judgments of comprehensibility must be compared to evaluations by other populations of listeners, most importantly including fellow L2 speakers, to increase the validity of findings in global contexts of L2 use.

8. Conclusion

The findings help to demonstrate the complexity of L2 comprehensibility as a construct. Although L2 speakers were aware of the range of contributors to comprehensibility, the use of task repetition and self-assessment did not typically result in robust changes in comprehensibility or to a strong alignment of speakers' self-ratings with other listeners' assessments. Taken together, results point to the importance of repeated fluency-oriented tasks and the need for a sustained focus

on raising speakers' awareness of L2 comprehensibility and of practical strategies to help speakers become more comprehensible to listeners.

References

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, *15*, 35–59.

Association for Language Awareness (2017). Retrieved 15 September 2017 from http://www.lexically.net/ala/la_defined.htm

Babaii, E., Taghaddomi, S., & Pashmforoosh, R. (2016). Speaking self-assessment: Mismatches between learners' and teachers' criteria. *Language Testing*, *33*, 1–27.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Oxford, UK: Routledge.

Castañeda, M., & Rodríguez-González, E. (2011). L2 speaking self-ability perceptions through multiple video speech drafts. *Hispania*, *94*, 483–501.

Chen, Y. M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, *12*, 235–262.

Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility. *Journal of Second Language Pronunciation*, *2*, 160–182.

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

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2015b). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly*, *49*, 814–837.

de Saint Léger, D. (2009). Self-assessment of speaking skills and participation in a foreign language class. *Foreign Language Annals*, *42*, 158–178.

Derwing, T. (2003). What do ESL students say about their accents? *Canadian Modern Language Review*, *59*, 547–567.

Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their
    pronunciation needs and strategies. *System*, *30*, 155–166.

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

Dlaska, A., & Krekeler, C. (2008). Self-assessment of pronunciation. *System*, *36*, 506–
    516.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to
    recognize their own incompetence. *Current Directions in Psychological Science*,
    12, 83–87.

Ellis, R. (2005). Principles of instructed language learning. *System*, *33*, 209–224.

ETS (2011). *TOEFL iBT® research insight: TOEFL® program history*. Princeton, NJ:
    Educational Testing Service.

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

Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility. *Studies in
    Second Language Acquisition*, *34*, 475–505.

Jułkowska, I. A., & Cebrian, J. (2015). Effects of listener factors and stimulus
    properties on the intelligibility, comprehensibility and accentedness of L2
    speech. *Journal of Second Language Pronunciation*, *1*, 211–237.

Jung, Y., Kim, Y., & Murphy, J. (2017). The role of task repetition in learning word-
    stress patterns through auditory priming tasks. *Studies in Second Language
    Acquisition*, *39*, 319–346.

Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language
    pronunciation: A classroom study. *Language Awareness*, *19*, 171–185.

Kim, Y., & Tracy-Ventura, N. (2013). The role of task repetition in L2 performance
    development: What needs to be repeated during task-based interaction? *System*,
    41, 829–840.

Kissling, E. M., & O'Donnell, M. E. (2015). Increasing language awareness and self-
    efficacy of FL students using self-assessment and the ACTFL proficiency
    guidelines. *Language Awareness*, *24*, 283–302.

Larson-Hall, J. (2009). *A Guide to doing statistics in second language research using
    SPSS*. New York: Routledge.

Lappin-Fortin, K., & Rye, B. J. (2014). The use of pre-/posttest and self-assessment
    tools in a French pronunciation course. *Foreign Language Annals*, *47*, 300–320.

Lee, S., & Chang, S. (2005). Learner involvement in self-and peer-assessment of task-based oral performance. *Language Research*, *41*, 711–735.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, *4*, 221–250.

O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, *64*, 715–748.

Papajohn, D. (2005). *Toward Speaking Excellence: The Michigan Guide to maximizing your performance in the TSE ® test and other speaking tests*. Ann Arbour, MI: University of Michigan Press.

Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing*, *19*, 109–131.

Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, *11*, 63–90.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878-912.

Rajadurai, J. (2006). Pronunciation issues in non-native contexts: A Malaysian case study. *Malaysian Journal of ELT Research*, *2*, 43–59.

Ranta, E. (2010). English in the real world vs. English at school: Finnish English teachers' and students' views. *International Journal of Applied Linguistics*, *20*, 156–177.

Robinson, P. (2011). Task-based language learning: A review of issues. *Language Learning*, *61*, 1–36.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, 65, 395–412.

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

Saito, K., & Shintani, N. (2016). Do native speakers of North American and Singapore English differentially perceive comprehensibility in second language speech? *TESOL Quarterly*, *50*, 421–446.

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

Saito, K., Trofimovich, P., Isaacs, T., & Webb, S. (2017). Re-examining phonological and lexical correlates of second language comprehensibility: The role of rater experience. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment: Interdisciplinary perspectives* (pp. 141–156). Bristol, UK: Multilingual Matters.

Skehan, P. (2015). Foreign language aptitude and its relationship with grammar: A critical review. *Applied Linguistics*, *36*, 367–384.

Swain, M. & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics*, *16*, 371–391.

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self-and other-perception of second language speech. *Bilingualism: Language and Cognition*, *19*, 122–140.

van den Branden, K. (1997). Effects of negotiation on language learners' output. *Language Learning*, *47*, 589–636.

Venkatagiri, H. S., & Levis, J. M. (2007). Phonological awareness and speech comprehensibility: An exploratory study. *Language Awareness*, *16*, 263–277.

Zoss, J. M. (2015). *Adult English learners' perceptions of their pronunciation and linguistic self-confidence*. (Unpublished Master's thesis). Hamline University, Saint Paul, Minnesota.