Concordia Working Papers in Applied Linguistics, 6, 2015

Using What You Know: Can Cross-Linguistic Instruction Improve L2 Pronunciation?

Dustin Crowther
Michigan State University

Abstract

Aiming to take advantage of known differences between second language learners' first language (L1) and their target language, an approach towards the teaching of pronunciation was explored, as L1 Japanese speakers (n = 15) received four sessions of explicit instruction on word and syllable structure/timing differences between Japanese and English in the form of perceptual training and production practice. Analysis of three spontaneous speech recordings (pre, post, and delayed) indicated that participants did not improve in either accentedness or comprehensibility ratings, but approached statistically significant improvement in word stress error rate from pretest to delayed posttest (p = .055). As the overall power of the study was low (.64), as well as initial word stress error rates (.13), further research is needed with an appropriately sized sample of low proficiency learners to determine if cross-linguistic instruction may lead to actual improvement in these areas of speech. Issues regarding methodological difficulties are also addressed.

Approaches toward teaching pronunciation in the second language (L2) classroom have seemingly been underrepresented in L2 research, yet studies that have been performed indicate that instruction does indeed yield positive results (Lee, Jang, & Plonsky, 2015; Saito, 2012). It is

surprising that such an underrepresentation exists, given the importance of pronunciation for L2 learners. This importance can be seen in a variety of consequences that an L2 user speaking with an accent may experience (Flege, 1988), including negative social evaluation and loss of intelligibility, with the latter likely being of most importance.

Socially, possessing an accent can lead to situations of discrimination and discomfort. In looking at accent discrimination in Canada, Munro (2003) identifies stereotyping, harassment, and even occupational loss as negative ramifications of possessing a foreign accent in English. While, all three examples demonstrate a negative perception of accented L2 speech, issues related to accent are not only limited to negative social perception. In simple conversations, where social judgment is not an issue between the native and L2 speaker, accent still leads to frustration and embarrassment for both speaker and listener. On the speaker's behalf, this frustration stems from the fact that despite successful control over grammar and vocabulary, their message is still not understood (Derwing & Munro, 2015). While it is beyond both language learners' and teachers' ability to affect the negative social actions of those around them, it is not beyond their capability to improve L2 speakers' ability to deliver their message in an understandable way.

Several views exist on how to acquire the ability to deliver a clear message when speaking in an L2. For many L2 speakers, there is a feeling that in order to achieve clear, oral communication with a native-speaker, it is necessary to speak with a native-like accent (Derwing, 2003; Tokumoto & Shibata, 2011). Yet, there are other L2 speakers who see their accented speech as a symbol of their cultural identity (Baker, 1983; Gatbonton, Trofimovich & Magid, 2005; Tokumoto & Shibata, 2011). Looking beyond learner opinion, research shows that clarity of meaning is found in *comprehensibility* and *intelligibility* more so than in *accentedness*. Following Derwing & Munro (2015) these constructs are defined as follows:

a) Accentedness: how different an utterance by an L2 speaker sounds from that of a native speaker
b) Comprehensibility: the ease in which a listener can understand an L2 speaker's utterance
c) Intelligibility: the listener's actual comprehension of an L2 speaker's utterance

In terms of the relationship between the three, while some speakers found to be heavily accented have been shown to also be highly intelligible and comprehensible, those who possess low intelligibility and

comprehensibility are almost always highly accented (Derwing & Munro, 2015).

Whether focusing on accent or intelligibility and comprehensibility, there is a need to identify which linguistic measures of speech affect the perception of L2 speech. A number of studies have identified individual measures in isolation of each other, such as segmental accuracy (Derwing, Munro, & Wiebe, 1998), word stress, pitch range, pause or syllable length (Field, 2005; Kang, 2010; Kang, Rubin, & Pickering, 2010), and speech rate (Munro & Derwing, 2001; Trofimovich & Baker, 2006), as well as the importance of appropriate grammar and lexical usage (Fayer & Krasinski, 1987). More recent studies have considered the relative weight of these different measures. Trofimovich and Isaacs (2012) looked in depth at which specific linguistic aspects of L2 speech were related to comprehensibility and accentedness. Comprehensibility was chosen as the target, as its method of measurement, the use of 9-point Likert scales, is more reflective of those used on high stakes assessment contexts, such as TOEFL and IELTS (Isaacs & Trofimovich, 2012). Over four distinct categories (phonology, fluency, lexis/grammar, discourse), 19 different speech measures were identified and measured in the speech of 40 native French speakers learning English. While accent was linked primarily to pronunciation measures, comprehensibility was associated with both pronunciation and lexicogrammar measures (Trofimovich & Isaacs, 2012). The one linguistic aspect that proved common to both accentedness and comprehensibility was word stress error rate (e.g. *UN-ha-ppy* instead of *un-HA-ppy*), which was identified by the authors as being a result of the difference in syllable timing between participants' first language (L1), French, and the target language, English. English is stress-timed, where a stressed syllable receives greater energy than an unstressed syllable, causing a difference in length between the stressed and unstressed (Landefoged, 2001). French does not have such a contrast.

As the authors point out, such a contrast between the learner's L1 and L2 would present a major learning challenge, and would have a great effect on a listener's judgment of both their accentedness and comprehensibility. As well, it is suggested that such a contrast in word stress likely plays a far greater global role in the learning of an L2. Indeed, Saito, Trofimovich, and Isaacs (2015) found similar results for L1 Japanese speakers of English. This similarity is not surprising, as Japanese is also a non-contrastive language in terms of word stress, with each syllable (referred to as mora) given equal length (Ladefoged, 2001). Having identified the importance of correct usage of word stress in the perception

of Japanese-accented English speech, and knowing that these speakers potential difficulty stems from a difference in how word stress is handled between their native and target languages, placing an explicit focus on this difference would seem an appropriate approach towards instruction.

Explicit instruction focused on improving L2 speakers' English pronunciation, although limited, has shown a definitive improvement in the ability of participants (Saito, 2012). Working with high-intermediate adult learners, Couper (2006) found that an explicit focus on syllables and connected speech improved his learners' ability to avoid the use of epenthesis. In another study comparing three sets of learners, Derwing, Munro, and Wiebe (1998) found that a treatment group receiving instruction on stress, intonation, and rhythm showed greater improvement in pronunciation than a treatment group receiving segmental instruction, and both showed greater improvement over a control group. Additionally, Abe (2011) found that after only one month, Japanese high school students receiving form-focused instruction showed greater improvement learning English weak forms than those who did not. These studies make clear the effectiveness that explicit instruction can have on pronunciation (see Saito, 2012, for a more in depth overview of the role of explicit instruction in pronunciation), although none of the above studies placed a specified focus on cross-linguistic differences during instruction.

Cross-linguistic influence (CLI) has always been a major consideration when looking at pronunciation. With pronunciation being one of the more difficult aspects to acquire in an L2, it is understandable that L2 pronunciation will be affected by L1 phonetics and phonology (Odlin, 1989). Previous research on L2 phonology acquisition echoes this view on the importance of L1 influence (Eckman, 2004), and several books exist highlighting where L1 based difficulties are likely to occur for learners from different language backgrounds, such as Avery and Ehrlich (1992) and Swan and Smith (2001). In addition to the differences between word stress already highlighted, potential difficulties may involve learners categorizing separate phonemes in the L2 as a single L1 category (Best, 1995). An example of this is in how the /l/ and /r/ distinction found in English does not exist in Japanese, a language where the two sounds are mapped onto the same phoneme. This difference makes it difficult for L1 Japanese speakers to perceive and produce this distinction when speaking English. As well, to allow for easier pronunciation of L2 consonant clusters or word-end codas that do not exist in their L1, learners may insert (ellipses) or remove (ellision) vowel sounds (Ross, 1994; Dupoux,

Parlato, Frota, Hirose, & Peperkamp, 2011). Knowing that the L1 can play a heavy role in the acquisition of L2 pronunciation, a focus on how two languages are phonologically different would seem to offer great benefits to L2 speakers.

The value of raising learners' awareness of how their L1 differs from their target L2 has been demonstrated for both grammar and vocabulary. Young Francophone learners of English in Quebec, Canada showed a greater ability to correctly judge and produce question forms after L1/L2 differences were highlighted (Ammar, Lightbown, & Spada, 2010). Similar results were found among 15-16 year old L1 Hebrew learners of English in both the acquisition of vocabulary and the usage of compound nouns and reduced restrictive relative clauses (Laufer & Girsai, 2008; Kupferberg & Olshtain, 1996). The term *cross-linguistic instruction* was used by Kupferberg and Olshtain to describe this approach to instruction, where the differences between learners' L1 and the target L2 are highlighted to aid acquisition. Such an approach is less frequent in pronunciation training, although Saito (2011) did find that a treatment approach that included information on segmental differences between his Japanese learners' L1 and their target English led to an improvement in their comprehensibility ratings. However, no study has investigated whether such an approach focusing on word stress differences between the two languages would have a similar effect, on either correct production of word stress or perceived accentedness and comprehensibility. As such, this study intends to test the effectiveness of cross-linguistic instruction on the acquisition of English word stress by L1 Japanese learners. This leads to the first two research questions of this study:

RQ#1: Will instruction highlighting the difference between Japanese learners' L1 and their target L2 improve their word stress error rate?

RQ#2: Will instruction highlighting the difference between Japanese learners' L1 and their target L2 improve their ratings in accentedness and comprehensibility?

Following previous literature on the success of both explicit instruction on pronunciation, as well as the success of cross-linguistic instruction in other areas of language acquisition, it is expected that participants will show improvement in both word stress error rate, as well as in accentedness and comprehensibility.

The final research question addresses the idea that the relationship between word stress error rate and accentedness and comprehensibility found among French learners of English may be a global issue (Trofimovich & Isaacs, 2012):

RQ#3: For Japanese learners, what relationship exists between their word stress error rate and their accent and comprehensibility ratings?

As the L1 to be investigated, Japanese, features similar word stress to that of French, it is expected that a similar relationship will be found.

METHOD

Participants

All participants (N = 15; male = 5, female = 10) were L1 Japanese speakers living in Montreal, Quebec, Canada, ranging in age from 19 to 50 (*mean* = 30, *SD* = 9), recruited through an advertisement placed on the Japanese classified website from-Montreal.com. The range of time spent living abroad was vast, from two months to 20 years, although only four participants had been in Montreal for more than four years. Various reasons were reported for living in Montreal, including work (N = 3), study (N = 11), and living (N = 2). Of those studying, one reported having a job, one was completing an MA at a local university, and another was an exchange student at a different local university. The remaining nine were attending English language schools in the city. Two participants reported knowing both Chinese and French as an L2, four reported knowing only French, and one reported knowing only Chinese. An a priori power analysis conducted through the program *R* revealed that for *power* = .8 and *f* = .4, 21 participants would be needed. With only 15 participants in this study, *power* = .64 and *f* = .4.

Target of Cross-Linguistic Focus

The structure of Japanese syllables differs greatly from that of English. Whereas syllable structure in English can range in complexity (i.e., V, CV, CVC, CCV, VCC), Japanese is either V or CV (i.e., a, i, u, e, o, ka, ki, ku, ke,

ko). Each Japanese syllable within a word is referred to as a mora. Accompanying this structural difference, when morae are combined to form words (i.e., mi + do + ri = *midori*, or English *green*) their resulting prosodic patterns are more similar to French than they are to English. As with French syllables, there is no stress contrast from one mora to the next, allowing for equal length of each articulated mora (Landefoged, 2001). This pattern differs from English, where a stressed syllable receives greater energy, causing a difference in length between the stressed and unstressed (Landefoged, 2001), and this difference may potentially lead to difficulties in assigning correct stress for L1 Japanese speakers of English. As discussed previously, Saito et al. (2015) have shown the importance of correct word stress placement in how Japanese-accented speech is perceived.

Materials

Test materials. Three tests were created to measure participants' use of English word stress, each consisting of a spontaneous speech task. Three speech topics were used, based on previous topics used by Eckman (1991), each requesting two pieces of information. Test 1 asked whom had influenced the participant's life and in which way; test 2 asked what the participant was doing in Canada, and why they had chosen to do this there; test 3 asked about the difficulty of obtaining a visa for Canada, and what the procedure had been. The test materials were piloted with an L1 Chinese speaker and an L1 Japanese speaker, which led to the addition of prompts to aid participants in brainstorming for each topic. All test recordings were made on either a Sony or Olympus digital voice recorder, supplied to each participant by the researcher.

Instructional materials. Two sets of materials were developed for instructional purposes during the study. The first set was a four-page worksheet focused on the structural differences between Japanese and English, highlighting *word* and *syllable* structure and timing. It was important to raise learners' awareness of how structural differences could negatively affect production (such as ellision and epenthesis), as their ability to produce correct word stress depended on their ability to produce words with the correct number of syllables ("sports" rather than "suportsu"). Two handouts listing common consonant clusters in *word-initial* and *word-final* position (taken from Eckman (1991)) accompanied the worksheets. The second set of materials was a two-page worksheet

identifying the differences between English *content* and *function* words, as well as how these types of words affect stress within a sentence.

Perceptual practice worksheets. Two sets of worksheets were used. Set one, targeting consonant clusters, measured participants' ability to identify whether an epenthetic vowel was present or not within a word. In total, 20 nonce words were used, 10 taken from Masuda and Arai (2010), which featured consonant clusters mid-word (five with a vowel separating consonants, five without), and 10 taken from Cardoso (2011), featuring word end codas (five with a vowel after the coda, five without). Worksheet one required participants to chose which of two words they heard (with the epenthetic vowel or without). Worksheet two required participants to identify whether the second word of a three-word list matched the word preceding or following it. The second set of worksheets highlighted correct word stress. Worksheet one required participants to indicate which of 15 polysyllabic words heard featured contrastive word stress. Worksheet two required participants to identify, using the same 15 polysyllabic words, which syllable featured word stress and whether it was used correctly or not. All recordings used were of a native speaker of English, made using the computer-based program Audacity.

Production exercises. Three sets of exercises were developed for Sessions 2 to 4. For set one, targeting consonant clusters, two half-page paragraphs were created, with epenthetic vowels included within and at the end of words. Students were required to eliminate the epenthetic vowels in their paragraph and read it to their partner. The listening student was then required to answer four comprehension questions based on their partner's reading. For set two, which focused on word stress, two handouts were created. Handout one featured 10 examples each of two, three, and four syllable words, and three examples of five syllable words, all of which students were required to select the correct syllable to stress. Students then proceeded to check their answers and practice saying the words out loud with a partner. All words were taken from an academic word list developed by Coxhead (2000). Handout two featured four questions for students to ask a partner, and space for them to record all polysyllabic words used in the response. Set three was designed to allow students to practice the correct stress of content and function words. This involved students working in partners to first identify correct stress within a sentence, and second to produce these sentences allowed. Ten original sentences were created for this purpose.

Design

Using a within-subject design, this study measured the effectiveness of the independent variable *instruction* by measuring students' scores on three dependent variables over three test times (pre, post, and delayed). The dependent variables measured included word stress error rate, accentedness ratings, and comprehensibility ratings.

Procedure

In total, participants attended six sessions over seven weeks, with a two-week break between the fifth and sixth sessions. All sessions were 60 minutes in length, except Session 1, which was 120 minutes to allow for a pretest before the beginning of treatment.

The first hour of Session 1, along with Sessions 5 and 6 were used to complete each of the three tests. At the beginning of each session, participants received instructions on how to operate the digital voice recorders provided. In two separate groups, participants were given two minutes to brainstorm on the spontaneous speech topic and then spoke for between 90 seconds and two minutes on the topic.

Following the pretest in Session 1, participants received explicit instruction from the researcher using the first set of instructional worksheets. Sessions 2 and 3 were composed of a warm-up activity, two perceptual activities, and a practice activity. Session 2 used the *Consonant Cluster-based* worksheets, while Session 3 used the *Word Stress-based* worksheets. Session 4 consisted of a warm-up, followed by the second set of instructional worksheets, and finished with the *Sentence Stress* worksheets.

Analysis

For each recording, a 35-45 second sample was chosen for analysis. Whenever possible, samples lacked extended pauses and finished with the completion of a thought. All samples were then transcribed, with 25% being verified by a second transcriber. The agreement rate between transcribers was 98% (total words transcribed = 876, total changes = 16). As the agreement rate was high, all transcripts were accepted in their original form.

Coding

Each transcript was coded in three different ways.

Word stress error rate. Following Trofimovich and Isaacs (2012), word stress error rate was operationalized as "the total number of instances of misplaced or missing primary stress in polysyllabic words divided by the total number of polysyllabic words produced" (p. 908). Samples of such errors include *HA-ppy* spoken as *ha-PPY* or *sur-VIVE* spoken as *SUR-vive*. Initial coding was performed by the researcher, and 25% of the recordings were subsequently recoded to ensure reliability of coding. A test of inter-rater reliability showed low agreement (*kappa* = .38). Disagreements between raters were then discussed and negotiated until agreement was reached. All disagreements were then recoded by a third rater, which again revealed low agreement (*kappa* = .22). Due to the consistently low agreement, each potential polysyllabic word that was disagreed on was discussed between the researcher and two raters to determine what led to these disagreements. These included 1) segmental errors, such as /r/ pronounced as /l/, 2) nonnative-like choices, where the coder felt the learner's production of *city* as /'sIti/ was nonnativelike, with /'sIdi/ being the more accepted production, 3) regional accent, for example the word *laboratory* could be pronounced with either 3 or 4 syllables, and 4) word placement in a sentence, such as a change in pitch in words at the end of sentences to indicate a statement or question. Based on this discussion, agreement was reached for the disputed words in the 25% sample that was recoded, and the knowledge from these discussions was then used to recode the rest of the recordings.

Accentedness and comprehensibility. Using 9-point Likert scales, consistent with previous studies (i.e., Derwing & Munro, 2015; Trofimovich & Isaacs, 2012), three raters scored the participants on their accentedness and comprehensibility. Accentedness was defined as "how different the speaker sounds from a native speaker of North American English" and was rated from 1 ("heavily accented") to 9 ("no accent at all"). Comprehensibility was defined as "how easy the speaker is to understand, rated from 1 ("hard to understand") to 9 ("easy to understand"). Both term and scale definitions were adapted from Trofimovich and Isaacs (2012). All three raters were native speakers of English without previous experience with teaching English as a second/foreign language, and minimal contact with Japanese speakers of English. Using Cronbach's alpha inter-rater reliability was found to be at

an acceptable level ($\alpha_{acc.}$ = .77, $\alpha_{comp.}$ = .77) (Larson-Hall, 2009) and the three scores were thus averaged to determine a single mean score for each recording for all further analysis. For all tests, N = 15, and alpha was set at $p < .05$.

RESULTS

The first research question asked whether instruction highlighting the difference between learners' L1 and their target L2 would improve their word stress error rate. As the data were not normally distributed, violating a key assumption of repeated-measures ANOVA, a non-parametric Friedman test was used, revealing that participants' improvement in word stress error rate approached statistical significance (X^2 (2) = 5.8, p = .055). Further tests were then run to determine between which times this occurred. Figure 1 shows the learners' scores over the three test times.

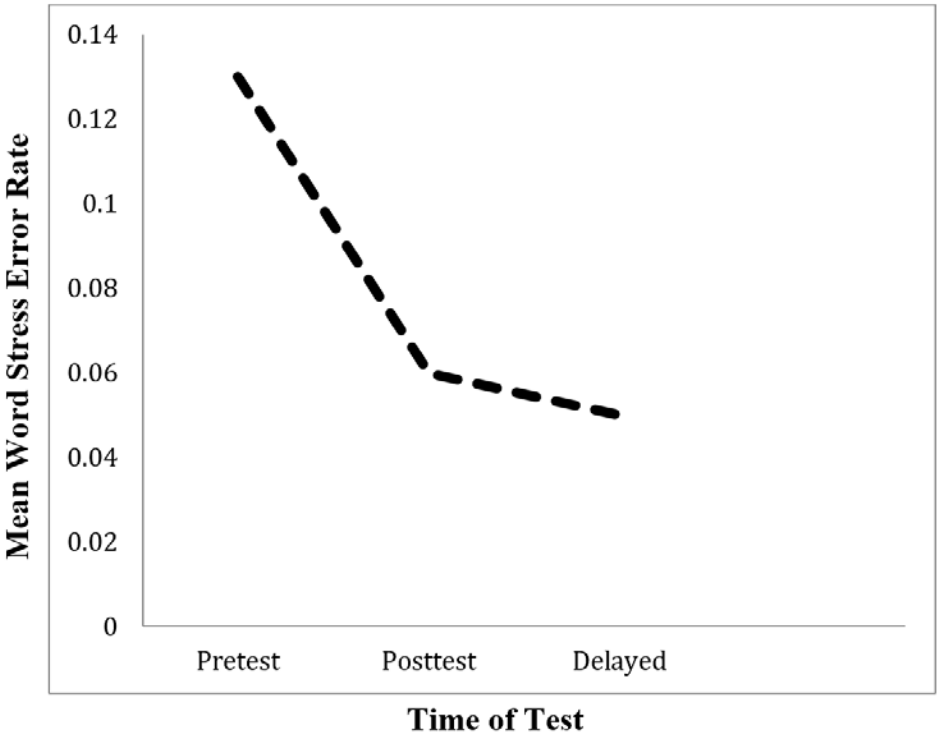

Figure 1. Mean word stress error ratings over time.

Post-hoc comparisons were conducted using a Wilcoxon signed-rank test, with a Bonferroni correction applied (a = .017). No significant improvement was found between pretest and posttest results (Z = 2.00, p = .05, r = .56) or posttest and delayed posttest results (Z = .71, p = .48, r = .22). The difference in results between pretest and delayed posttest approached significant improvement (Z = 2.35, p = .019), with a large effect size (r = .70).

The second research question asked whether instruction highlighting the difference between learners' L1 and their target L2 would improve their ratings in accentedness and comprehensibility. A comparison of ratings over time for accentedness and comprehensibility can be seen in *Figure 2.* A repeated-measures ANOVA was performed to test the effect of treatment between pretest, posttest, and delayed posttest recordings. For accentedness, no significant main effect was found (F(2,28) = .17, p = .84), indicating no statistically significant improvement over time. Effect sizes for all three relationships were small (Pretest to Posttest: d = .1; Pretest to Delayed Posttest: d = .1; Posttest to Delayed Posttest: d = 0).

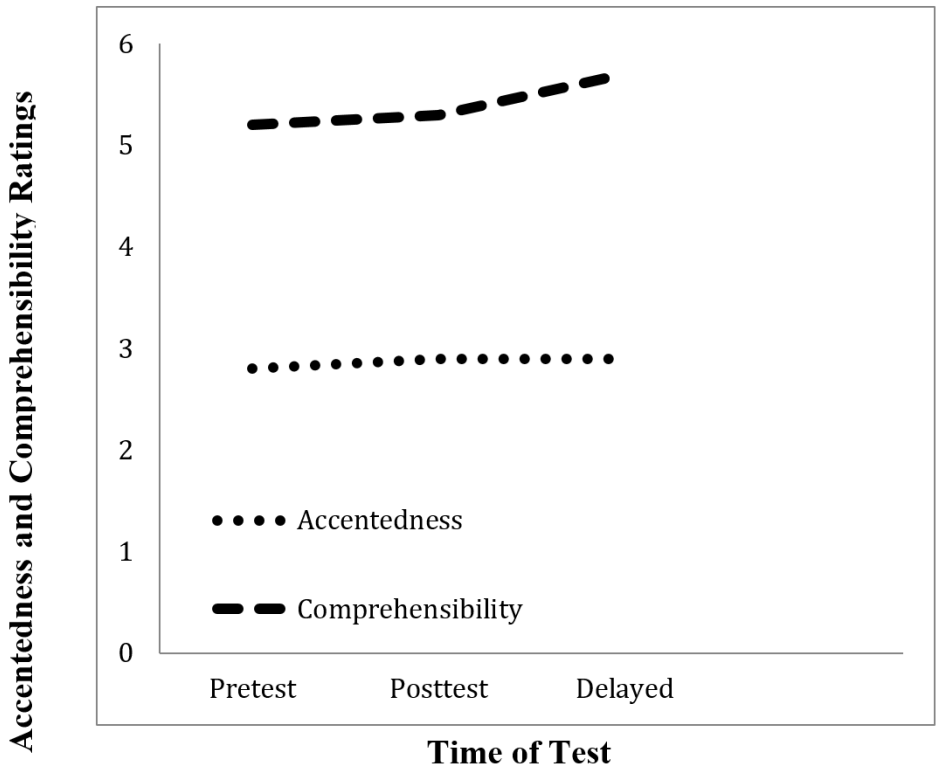

Figure 2. Mean accentedness and comprehensibility ratings over time.

For comprehensibility, no significant main effect was found ($F(2,28) = .17$, $p = .58$), indicating no statistically significant improvement over time. Effect size for the relationship between pretest and posttest ratings was small ($d = .05$), and for the other two potential relationships were either small or medium (pretest to delayed posttest: $d = .30$; posttest to delayed posttest: $d = .26$).

The third research question investigated the relationship between word stress and both accentedness and comprehensibility. To ensure the consistency of the relationship, a correlation was run for all three times of the treatment. As the error rates gathered were not normally distributed among participants, violating an assumption of a Pearson correlation, the non-parametric Spearman's *rho* test was used. A one-way correlation for word stress error rate with both accentedness and comprehensibility ratings on the pretest found no significant relationship with either accentedness (95% CI: -.63, .37; $r = .17$, N = 15, $p = .25$) or comprehensibility (95% CI: -.69, .28; $r = .27$, N = 15, $p = .17$), although following Larson-Hall (2009) the effect for both was moderate.

A one-way correlation on the posttest found a statistically significant relationship with a large effect size for word stress error rate when paired with comprehensibility (95% CI: -.83, -.05; $r = .55$, $N = 15$, $p = .02$). This indicates that as word stress error rates went down, comprehensibility ratings went up, and the strength of this relationship was large (Larson-Hall, 2009). However, no significant relationship was found between word stress error rate and accentedness (95% CI: -.66, .33; $r = .22$, N = 15, p = .22).

A one-way correlation on the delayed posttest found a statistically significant relationship with a medium effect size for word stress error rate when paired with accentedness (95% CI: -.80, .04; $r = .48$, $N = 15$, $p = .03$), but no significant relationship when paired with comprehensibility (95% CI: -.66, .33; $r = .22$, N = 15, p = .22). This indicates that as word stress error rates went down, accentedness scores went up. However, the strength of this claim is questionable, as the 95% confidence interval spans zero, which would indicate a non-significant finding.

Although a relationship was found between word stress error rate and comprehensibility during the posttest, and word stress error rate and accentedness during the delayed posttest, there was no consistent relationship found across all three test times, as would be expected if a true relationship existed.

DISCUSSION

A series of statistical tests revealed that no statistically significant improvement was found in either learners' word stress error rate or accentedness and comprehensibility ratings. As well, word stress error rate was not found to have any consistent statistically significant relationship with either accentedness or comprehensibility across the three test times. These findings all run contrary to previous studies, and there are various reasons that may explain why these contradictory results occurred.

In terms of the lack of improvement found between pretest and posttest, the actual size of the sample tested must be considered. With only 15 participants, the a priori power analysis was only .64, which means that there was a 36% chance that any effect that may have existed would not have been found. Despite the smaller sample size, statistical significance was still approached (p = .055). With a larger sample size, and thus a higher power level, improvement may actually be found. Previous studies on explicit instruction of pronunciation that have generated significant results support this theory as well, as both Couper (2006) and Abe (2011) had larger sample sizes of 21 and 30 participants respectively. Yet, another study conducted by Derwing, Munro, and Wiebe (1998) showed overall improvement in two different treatment groups, both consisting of only 16 participants. So even though it appears that a larger sample size may generate statistically significant findings, other considerations of why the current sample did not must be considered.

The actual instructional time given to the students during treatment was minimal, (four hours over four weeks). This may not be enough exposure for students to improve. For Derwing, Munro, and Wiebe (1998), instruction occurred in an ESL setting over 12 weeks, and for Couper (2006), students received 12 different instructional sessions over two weeks, averaging 30 minutes each. These two studies indicate that either prolonged exposure or intensive exposure to the instructional approach would be of benefit to the learner, neither of which was administered to the participants of this study.

A third consideration, and potentially the strongest, is the proficiency level of the participants. Due to the limited number of Japanese residents in Montreal, the sample consisted of volunteers of varying proficiency levels, which led to a low initial word stress error rate. Due to this treatment focusing specifically on only this one aspect, there may have been no room for improvement in word stress error rate, leading to no

improvement in accentedness or comprehensibility. A different sample of participants, all beginning with high error rates may produce a different result. This is supported by the fact that of the three participants who had initial error rates above .25, all showed improvement by the delayed postest, with participant one improving from .40 to .08, participant 5 from .40 to .07, and participant 7 from .25 to .07.

The last consideration is that cross-linguistic instruction may not be appropriate for pronunciation. Practically, such an approach will require the instructor to have a strong grasp on how the two phonological systems, native and target, differ. In the case of this study, the teacher's knowledge was based on three years experience teaching English to low-level high school learners in Japan. Such knowledge is not readily available to all teachers. Additionally, such an approach is seemingly appropriate only in foreign language learning contexts, where learners share a first language background (Celce-Murcia, Brinton, & Goodwin, 2010). Possibly more important however, is that there are learner components to speech production that are beyond teacher control. Previous examples of this cross-linguistic instruction have been in the areas of grammar (Ammar, Lightbown, & Spada, 2010; Kupferberg & Olshtain, 1996) and vocabulary (Laufer & Girsai, 2008), two areas in which teachers can offer concrete, rule-based examples that demonstrate how the L1 and L2 differ. Unlike grammar and vocabulary, however, pronunciation involves a physical act, and thus a focus on not only the cognitive skills, but also motor (Scovel, 2006). Even if teachers clearly explain how the phonological systems of the two languages differ, such as how syllables are structured and the use of word stress, as targeted in this study, they cannot control whether the learners can physically produce the correct target form. Additionally beyond their control is the perceptual aspect of speech production, with many studies claiming the importance of perceiving a difference before producing it (Flege, 1991; Cardoso, 2011). Again, even if teachers offer clear examples of how the two languages differ, learners' potential inability to hear this difference is likely to hinder their ability to produce it. While perceptual training was used during this study, no actual measurement of participants' perceptual ability was taken, as this was not a focus. Lastly, it may be that cross-linguistic instruction did not work because the difference between the two languages is not as severe as would seem. Japanese learners have been shown to be able to produce a native-like difference between stressed and unstressed vowel duration, although they still lack native-like production in terms of the actual quality of the reduced vowel in the unstressed

syllable (Lee, Guion, & Harada, 2006). It may be more pertinent, then, to test the effects of instruction on vowel reduction, more so than on actual word stress, and see if this leads to an improvement in accentedness or comprehensibility.

As discussed previously, agreement between coders on what constituted a word stress error proved difficult to attain. Subsequent discussions between coders identified several considerations not directly relevant to word stress, including segmental errors, nonnative-like choices, regional accent, and word placement within a sentence. This difficulty in coding may explain why no relationship was found between word stress and either accentedness or comprehensibility, which is in contrast with Trofimovich and Isaacs (2012). In Trofimovich and Isaacs (2012), there were 19 different linguistic aspects of speech being measured, possibly allowing coders an easier time differentiating word stress errors from other aspects of speech. In terms of the current study, with only one linguistic aspect being measured, the coders were possibly unable to place a singular focus on word stress errors, which led to the difficulty in attaining interrater reliability. It may simply be that no relationship was found between word stress error rate and accentedness or comprehensibility because it was not word stress error rate that was truly being measured in this study.

Based on this difficulty, two interesting questions in terms of how word stress affects listeners' perception of L2 speech are raised. The first is which other linguistic aspects of speech affect the perception of correct word stress and whether they can be filtered out when making a judgment on learners' actual correct usage. The second is the overall importance of word stress as a measurement of accentedness and comprehensibility. Is the reason these other linguistic aspects of speech are so prevalent to listeners because they hold greater importance, or could it be that we cannot actually consider any of these aspects individually from each other?

Several key limitations have already been mentioned, including word stress error rate coding, sample size, and lack of proficiency control at the outset of the study. Another limitation is that no native speaker speech samples were used during the accentedness and comprehensibility rating process. Initially, raters were informed that they would be rating the speech of Japanese learners of English, and as the L1 was so specific, it was felt that a native speaker of English would be too easily identified if inserted into the samples. Based on rater feedback, they did not believe this would have been the case, and any replication or follow-up study

should include native-speech samples to ensure the reliability of raters' ability to differentiate between native and nonnative speech.

CONCLUSION

Despite the cross-linguistic instruction provided, no statistically significant improvement was found in word stress error rate or accentedness and comprehensibility between pretest and either posttest. However, improvement in word stress error rate approach significance, and considering the low sample size, and subsequent power level, further research with an appropriate sample size may be warranted.

The coding of word stress error rate proved problematic, due to numerous other linguistic aspects of speech considered when judging whether a word had been stressed correctly. Further research looking into which aspects of speech affect listeners' perception of word stress, as well as the overall importance of each of these aspects in accentedness and comprehensibility, may allow for a more accurate measurement of actual word stress errors committed by L2 speakers of English in the future. More importantly, answers to these questions may allow for a better overall view of which aspects truly deserve focus when targeting pronunciation in the classroom.

REFERENCES

Abe, A. (2011). Effects of form-focused instruction on the acquisition of English weak forms by Japanese EFL learners. In W.-S. Lee & E. Zee (eds.), Proceedings of the 17th International Congress of Phonetic Sciences. (pp. 184-187), Hong Kong.

Ammar, A., Lightbown, P. M., & Spada, N. (2010). Awareness of L1/L2 differences: Does it matter? *Language Awareness, 19*, 129-146.

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation.* Oxford, UK: Oxford University Press.

Baker, H. D. R. (1983). Life in the cities: The emergence of Hong Kong man. *The China Quarterly, 95*, 469-79.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.

Cardoso, W. (2011). The development of coda perception in second language phonology: A variationist perspective. *Second Language Research, 27*(3), 1-33.

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

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect*, *21*(1), 46-66.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*, 213-238.

Derwing, T. M. (2003). What do ESL students say about their accents? *The Canadian Modern Language Review, 59*, 547-566.

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

Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*, 393-410.

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language, 64*, 199-210.

Eckman, F. R. (1991). The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition, 13*, 23-41.

Eckman, F. R. (2004). From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition, 26*, 513-549.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, *37*, 313-326.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39*, 399-423.

Flege, J. E. (1988). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), *Human communication and its disorders. A review* (pp. 224-401). Norwood, N.J.: Ablex.

Flege, J. E. (1991). Perception and production: The relevance of phonetic input to l2 phonological learning. In T. Hueber and C. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theories* (pp. 249-289). Amsterdam: John Benjamins.

Gatbonton, E., Trofimovich, P., & Magid, M. (2005). Learners' ethnic group affiliation and L2 pronunciation accuracy: A sociolinguistic investigation. *TESOL Quarterly, 3*, 489-511.

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

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

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

Kupferberg, I., & Olshtain, E. (1996). Explicit contrastive instruction facilitates the acquisition of difficult L2 forms. *Language Awareness, 5*, 149-165.

Ladefoged, P. (2001). *A Course in Phonetics (Fourth ed.)*. US: Heinle & Heinle.

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

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics, 29*, 694-716.

Lee, B., Guion, S. G., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition, 28*, 487-513.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics, 36*, 345-466.

Masuda, H. & Arai, T. (2010). Processing of consonant clusters by Japanese native speakers: Influence of English learning backgrounds. *Acoustical Science and Technology, 31*, 320-327.

Munro, M. J. (2003). A primer on accent discrimination in the Canadian context. *TESL Canada Journal/ Revue TESL du Canada, 20*(2), 38-51.

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition, 23*, 451-468.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. New York, NY: Cambridge University Press.

Ross, S. (1994). The ins and outs of paragoge and apocope in Japanese-English interphonology. *Second Language Research, 10*, 1-24.

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

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, 842-854.

Saito, K., Trofimovich, P., & Isaacs, T. (2015). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Linguistics*. Published online 2 February 2015.

Scovel, T. (2006). Age in L2 acquisition and teaching. In C. Abello-Contesse, R. Chacón-Beltrán, L.–J. M. D. & M. M. Torreblanca-López (Eds.). *Age in L2 acquisition and teaching* (pp. 31-48). Bern, Germany: Peter Lang.

Swan, M., & Smith, B. (Eds.) (2001). *Learner English: A teacher's guide to interference and other problems* (2nd ed.). Cambridge, UK: Cambridge University Press.

Tokumoto, M., & Shibata, M. (2011). Asian varieties of English: Attitudes towards pronunciation. *World Englishes, 30*, 392-408.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effects of L2 experience on 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*, 905-916.