Gordon, J. (2018). Second-language comprehensibility ratings: do ESL and EFL teachers rate in the same way? In J. Levis (Ed.), *Proceedings of the 9th Pronunciation in Second Language Learning and Teaching conference*, ISSN 2380-9566, University of Utah, September, 2017 (pp. 1-16). Ames, IA: Iowa State University.

SECOND LANGUAGE COMPREHENSIBILITY RATINGS: DO ESL AND EFL TEACHERS RATE IN THE SAME WAY?

Joshua Gordon, University of Northern Iowa

This mixed-method investigation analyzed the way native-ESL and nonnative-EFL teachers rated second language (L2) comprehensibility in ESL learners. Two groups of native-ESL and nonnative-EFL teachers rated spontaneous speech samples from ESL learners obtained before and after a stand-alone pronunciation course in an intensive ESL program. Speech samples from a group of L1-English speakers were also included for control purposes. The quantitative analyses indicated that the group of nonnative-EFL teachers rated the speech samples more severely than the native-ESL teachers. Additionally, stimulated-recall interviews carried out with each teacher after the rating task revealed similarities and differences between both groups of teachers. However, the most important differences seemed to be rooted in teachers' pedagogical knowledge, teacher training, and familiarity (or lack of familiarity) with different L2 accents. The results of this investigation are discussed in terms of their implications for L2 pronunciation instruction and teacher training.

INTRODUCTION

Studies that have explored how native (NSs) and nonnative speakers (NNSs) rate second language (L2) speech for comprehensibility (i.e., the perceived ease or difficulty of understanding a message, see Derwing & Munro, 2009) have found mixed results. Whereas NNSs have been found to rate L2 speech more severely than NSs (Fayer & Krasinski, 1987; Kang, 2012; Rossiter, 2009), other studies have found the opposite (Brown, 1995), while other investigations have not found any significant differences between both groups of listeners (Crowther, Trofimovich, & Isaacs, 2016; Derwing & Munro, 2013; Flege, 1988; MacKay, Flege, & Imani, 2006). However, the ways that L2 teachers rate L2 comprehensibility has remained mostly unexamined. Such a question is important to understand since the teacher is, in the majority of cases, the one who makes pedagogical decisions in class as to which students need to work more on their pronunciation. The mixed-method investigation which is discussed here analyzed the way native-ESL and nonnative-EFL teachers rated comprehensibility in ESL learners in order to explore the differences and similarities in how teachers rate, and to understand the types of criteria used by both groups of teachers to rate L2 speech. In the next pages, I will present a review of literature and previous studies that motivated my investigation, followed by the details of the methodology used to carry out this study. Finally, I will present quantitative and qualitative analyses, followed by a discussion of the significance of these findings for pronunciation instruction and teacher education.

Literature Review

Studies that have investigated what makes L2 speech comprehensible have demonstrated that in addition to phonological issues, lexical and syntactical factors can also affect the degree of comprehensibility in L2 speech (Crowther et al., 2016; Derwing & Munro, 2015; Trofimovich & Isaacs, 2012). In contrast, other studies that have investigated what constitutes a foreign accent in

L2 speech have determined that this dimension is mostly linked to phonological factors (see Derwing & Munro, 2009, 2015). Whereas these studies have analyzed what constitutes comprehensibility and foreign accent in L2 speech, other studies have investigated whether NSs or NNSs perceive L2 speech in the same way in terms of comprehensibility; the results of such studies have presented mixed results.

A few studies have determined that NNSs have a tendency to be stricter when it comes to rating L2 speech. For example, Kang (2012) investigated the effects of raters' backgrounds when rating the L2 speech of international teaching assistants (ITAs) from American universities. In this study, 70 undergraduate students (48 NSs and 22 NNSs of English) rated speech samples of 11 prospective ITAs from different language backgrounds. Kang found that NNSs were more severe than the NSs in their ratings, which confirmed previous results in which NNSs were stricter than NSs in rating L2 speech (e.g., Fayer & Krasinski, 1987; Rossiter, 2009). In other studies, however, the NSs have appeared to be more severe when rating L2 speech. For instance, Brown (1995) investigated the effects of raters' occupational and linguistic backgrounds at the moment of assessing L2 speakers of Japanese. In this case, 33 NSs and NNSs of Japanese rated speech samples of 51 NNSs of Japanese. Brown found that the NNSs were more lenient in their ratings when compared to the NSs.

The difference between NSs' and NNSs' ratings of L2 speech is even more complex in the sense that, contrary to the above-mentioned studies, other studies have not found significant differences between these groups of raters. In these studies, this lack of difference between NS and NNS raters may be rooted in the L2 proficiency of the NNSs. For example, in studies in which NNS raters had very high levels of L2 proficiency, those NNS raters typically tended to provide rating scores that were not significantly different from those of the NS raters (e.g., Derwing & Munro, 2013).[1] In a recent study, Crowther, et al. (2016) investigated the effects of listener status (NS vs. NNS) and language background when rating comprehensibility and accentedness of L2 speech. For this study, 26 NNSs of English (speakers of L2 English with French and Mandarin L1 backgrounds) rated spontaneous speech samples of 40 L2-English speakers (whose L1 was French). The ratings were compared with those of 60 NSs of English. There were no significant differences in the global ratings between the 60 NSs and the 26 NNSs, or between the two groups of NNSs. These results also agreed with other studies that did not find significant differences in the L2 speech ratings of NSs and NNSs (e.g., Derwing & Munro, 1997; 2013; Flege, 1988; MacKay, et al., 2006).

In addition to differences between NSs and NNSs, other studies have pointed out that different types of raters can provide different results. For example, Kennedy and Trofimovich (2008) stated that expert raters, like language teachers or linguistics students, can provide biased ratings because they regularly hear different accents, which could lead to more lenient ratings than those of an inexperienced rater (see also Thompson, 1991). Additionally, other studies have argued that familiarity with a specific accent (e.g., because of knowledge of another language) can also produce biased results. For example, Winke and Gass (2013) found that raters' familiarity with speakers' L1 affected how they came up with ratings of specific L2 speech samples, as opposed to other raters who were not familiar with specific languages or accents. In other words, this is an

[1] I would like to thank Tracey Derwing for bringing this issue to my attention.

area that has presented inconclusive results, as sharing the same L1 background between speakers and raters has been an advantage in some studies (e.g., Bent & Bradlow, 2003; Harding, 2012), but it has not been a determining factor in enhancing comprehensibility in listeners familiar with certain L1s in other studies (see Munro, Derwing, & Morton, 2006).

Other studies have analyzed how speaker-independent factors affect L2-speech perception. Levi, Winters, and Pisoni (2007) examined the effects of listening context and lexical frequency in the perception of foreign-accented speech. Listeners rated foreign accent in individual words produced by L1 and L2 speakers of English. The words were presented to the listeners in two contexts: auditory-only, and auditory+orthography. The results demonstrated that listeners perceived high-frequency words as less accented. Additionally, the use of orthographic cues caused L1 speakers of English to be perceived as less accented, whereas L2 speakers were rated as more accented. Levi et al. suggested that the high frequency words were probably rated as less accented because the listeners had more stored exemplars of them—that is, their representations were more robust in memory, so it was easier to perceive them in the rating task.

In a similar line of research, Saito and Shintani (2016) investigated the effects of bilingualism in the perception of comprehensibility of L2 speech. In this study, spontaneous speech samples from 50 L2-English speakers (L1-Japanese speakers) were presented to two groups of raters: 10 Canadian monolingual L1-English speakers, and 10 Singaporean L1-English speakers who were also proficient in other languages. Their results demonstrated two interesting findings. First, ratings were mainly influenced by phonological and temporal cues regardless of the raters' backgrounds. Second, the Singaporean raters were more lenient in rating comprehensibility. Saito and Shintani suggested that the Singaporean listeners, as speakers of different languages, were better accustomed to different phonological patterns, which made them more accommodating than the monolingual speakers when parsing L2 speech.

These conflicting results—ranging from the different aspects that listeners pay attention to in L2 speech to how raters' different language backgrounds affect how they rate—demonstrate that the factors underlying ratings of L2 comprehensibility are complex. Certainly, this type of research is clearly important for L2 pronunciation teaching and teacher education purposes; however, one group of raters that remains mostly unexamined is language teachers, and more specifically, NS and NNS teachers and how they may rate differently in different teaching contexts (i.e., second and foreign language contexts). This is especially important in the classroom, as language teachers decide which students need to work more on their pronunciation. However, as mentioned earlier, teachers' perceptions of their students' speech can be biased (Kennedy & Trofimovich, 2008) and their familiarity with different accents, or their NS/NNS status, can affect L2 speech perception (see Brown, 1995; Crowther et al., 2016; Kang, 2012; Winke & Gass, 2013). Therefore, it is important to compare the comprehensibility ratings of groups of teachers in different contexts.

The Current Study

In this study, I investigated how native-ESL and nonnative-EFL teachers rated L2 speech samples for comprehensibility to determine whether there were differences or similarities in the teachers' ratings, and to analyze on which aspects these two groups of teachers based their ratings. The study was guided by the following research questions:

1. Do ESL and EFL teachers rate L2 speech comprehensibility in the same way?

2. On which linguistic aspects (i.e., phonological, lexical, grammatical) do ESL and EFL teachers base their ratings?

METHODOLOGY

Three steps were followed to carry out this study. First, speech samples were collected from L2 learners in two stand-alone pronunciation classes in an intensive ESL program at a large American university in the Midwest. Speech samples from L1-English speakers were also collected as a control. Second, two groups of teachers (native-ESL & nonnative-EFL teachers) rated the speech samples for comprehensibility. Finally, stimulated-recall interviews were carried out with each teacher after the rating task to discuss their reasons for their ratings.

Stimuli

The speech samples used in the rating task were part of a corpus obtained from a previous study (see Gordon Zamora, 2015). Speech samples in the form of video-description narratives were collected from 10 ESL learners in two stand-alone pronunciation classes (Class A & Class B hereafter) in an intensive ESL program. There were 5 ESL learners in each class, who were enrolled in the highest two institutional levels of the program. They came from several countries, and spoke different L1s (see Table 1). These learners recorded descriptions of two different video cartoons found on the internet (Simon's Cat, 2009, 2010) at the beginning of their course (Time 1) and also at the end of their course (Time 2).[2] In addition to recordings from these L2 learners, 5 L1-English speakers also recorded descriptions of both video cartoons. These L1-English speakers were undergraduate students from the same school, and they recorded descriptions of both video cartoons only once. Because of the spontaneous nature of the samples, all the narratives were first transcribed to find similarities in the descriptions. There was a specific event in the plot of each cartoon that all the ESL learners and the L1-English speakers described across the board. Thus, passages between 15 or 17 seconds of those descriptions were presented to the two groups of ESL and EFL teachers in a rating task.

[2] The video cartoons Simon's Cat (2009) and Simon's Cat (2010) were presented to the ESL learners in Class A and B in inverted order to maintain a balance. For instance, the video that learners in Class A watched and described at Time 1 is the same video that learners in Class B watched and described at Time 2. In the same way, the video that learners in Class B watched at Time 1 is the same video that learners in Class A watched at Time 2.

Table 1.

ESL Learners from Class A and Class B.

CLASS A

ESL Learner	Country of Origin	Age	Native Language	Level in the Program
CAP01	Japan	20	Japanese	Level 7
CAP02	South Korea	25	Korean	Level 6
CAP03	Argentina	46	Spanish	Level 6
CAP04	Japan	20	Japanese	Level 7
CAP05	Kazakhstan	21	Kazakh	Level 6

CLASS B

ESL Learner	Country of Origin	Age	Native Language	Level in the Program
CBP01	Japan	21	Japanese	Level 7
CBP02	Saudi Arabia	35	Arabic	Level 7
CBP03	Japan	19	Japanese	Level 7
CBP04	Chile	40	Spanish	Level 6
CBP05	South Korea	20	Korean	Level 7

The Raters

All the speech samples were presented to a group of 12 native-ESL teachers and a group of 15 nonnative-EFL teachers. The ESL teachers were all born in the U.S., and all were L1 speakers of American English. All were instructors in two intensive ESL programs at two universities in the Midwest; they included both novice and experienced teachers, and all had advanced degrees in TESOL, Linguistics, or Applied Linguistics. All of them spoke at least one L2, and the majority even spoke or at least had knowledge of a third language. Additionally, each ESL instructor had taken a course on pedagogical phonology (i.e., a specific course where they studied how to teach L2 pronunciation), and the majority had lived in a country where one of their L2s was spoken (see Table 2). The EFL teachers were all from Costa Rica and were L1 speakers of Spanish. They all taught at the university level and had advanced degrees in teaching EFL. The majority worked as EFL teacher trainers preparing preservice teachers, and others worked as EFL

instructors for students from other majors.[3] The EFL teachers had taken only basic linguistics and phonetics/phonology courses as part of their training. They spoke only one L2 (English), and very few had knowledge of a third language at a basic reading level. In addition, most of the EFL teachers had not spent a considerable amount of time living in a country where their L2 was spoken (see Table 3). These two groups of teachers carried out an individual rating task, described below.

Table 2

Native-ESL Teachers' Background.

ESL Teacher	Age	Gender	Second Language	Self-Rated Proficiency	Other Languages	Self-Rated Proficiency	Highest Academic Degree	Years of Teaching Experience	Time Living Abroad
ESLN01	26	F	Spanish	Intermediate	Turkish	Beginner	M.A.	3	Spain, 6 months
ESLN02	25	F	Serbian	Intermediate	Russian	Intermediate	M.A.	2	Bosnia, 8 months
ESLN03	24	F	Spanish	Advanced	French	Beginner	M.A.	3	NA
ESLN04	23	F	German	Intermediate	Swedish	Beginner	M.A.	3	NA
ESLN05	24	M	Japanese	Advanced	Chinese	Beginner	M.A.	3	NA
ESLS01	32	F	Spanish	Advanced	NA	NA	M.A.	5	Mexico, 5 months
ESLS02	60	F	Dutch	Intermediate	NA	NA	M.A.	9	Holland, 20 years
ESLS03	45	M	Japanese	Beginner	NA	NA	M.A.	11	Japan, 1 year
ESLS04	30	F	Kyrgyz	Intermediate	Spanish	Intermediate	M.A.	7	Kyrgyzstan, 2 years
ESLS05	38	F	Spanish	Advanced	Quechua	Beginner	M.A.	12	Panama, 2 years
ESLS06	48	M	Japanese	Advanced	Mandarin	Intermediate	M.A.	15	Japan, 3 years
ESLS07	30	M	Spanish	Advanced	French	Intermediate	M.A.	7	Spain, 2 years

[3] Although no specific test was given to any of the teachers to assess their L2 proficiency (and the majority of EFL teachers in fact self-rated their L2 proficiency as "advanced" or "near-native"), the group of EFL teachers represented a homogeneous group of proficient speakers who had high levels of their L2 in order to perform some of their teaching duties (e.g., teacher training preservice teachers in different areas).

Table 3

Nonnative-EFL Teachers' Background.

EFL Teacher	Age	Gender	Second Language	Self-Rated Proficiency	Other Languages	Self-Rated Proficiency	Highest Academic Degree	Years of Teaching Experience	Time Living Abroad
EFLN01	24	M	English	Advanced	NA	NA	Lic.	2	NA
EFLN02	26	F	English	Advanced	NA	NA	Lic.	2	NA
EFLN03	25	F	English	Advanced	NA	NA	Lic.	3	NA
EFLN04	25	F	English	Advanced	NA	NA	Lic.	4	NA
EFLN05	23	M	English	Advanced	French	Reading Knowledge	Lic.	1	NA
EFLS01	36	F	English	Advanced	NA	NA	M.A.	13	NA
EFLS02	36	M	English	Advanced	NA	NA	M.A.	12	U.S. 1 month
EFLS04	33	F	English	Advanced	NA	NA	M.A.	10	U.S. 4 months
EFLS05	35	F	English	Advanced	French	Reading Knowledge	Ms.C.	13	NA
EFLS06	45	F	English	Near-Native	French	Reading Knowledge	M.A.	17	NA
EFLS07	33	F	English	Advanced	NA	NA	M.A.	11	NA
EFLS08	36	F	English	Advanced	NA	NA	M.A.	14	NA
EFLS09	34	M	English	Advanced	NA	NA	M.A.	12	NA
EFLS10	33	M	English	Advanced	NA	NA	M.A.	12	NA
EFLS11	34	F	English	Advanced	NA	NA	M.A.	13	NA

Rating Task & Stimulated-Recall Interviews

The rating task was carried out individually on a personal computer in a quiet office or library. First, the teachers watched the two video cartoons described by the speakers to avoid biased ratings with the initial samples (see Derwing, Rossiter, Munro, & Thomson, 2004). They also carried out a short warm-up to get familiar with the task.[4] The teachers then completed the rating task, which was programmed using PRAAT (Boersma & Weenink, 2015). The teachers heard the randomized speech samples through high-quality headphones, and rated each one by clicking

[4] For the warm-up task, the raters listened to 10 different sentences spoken by individuals who did not participate in the actual task. The sentences did not have any relationship with the videos being described by the speakers in the actual task.

on a 9-point Likert scale on the computer screen (1=*extremely easy to understand*, 9=*impossible to understand*; see Munro & Derwing, 1995).

Once the rating task was over, ten samples were randomly selected for a stimulated-recall interview with each teacher (Gass & Mackey, 2000). The samples were played one by one, and the teachers were reminded of the ratings given to each specific sample. They were asked about possible problems that made the speech samples difficult to understand (see Appendix 1). Further questions were asked for clarification, and if the teachers requested to hear a sample again, it was played as many times as necessary. Finally, all the stimulated-recall interviews were transcribed and comments were classified according to common themes (see Richards, 2003).

RESULTS

Quantitative

The inter-rater reliability coefficients (Cronbach's alpha) computed across all ratings given by the ESL and EFL teachers were very high at both times 1 and 2 (ESL = T1 .95, T2 .95; EFL = T1 .97, T2 .97). This indicated a strong agreement (Larson-Hall, 2009). As expected, both groups of raters found the NS control group the most comprehensible, as seen in the mean scores reported in Table 4.

Table 4.

Mean Comprehensibility Ratings.

Rater Group	Group	Time 1	Std. Error	Time 2	Std. Error
ESL Teachers	Class A (*n*=5)	3.96	0.34	4.51	0.35
	Class B (*n*=5)	3.57	0.34	3.57	0.36
	NS Group (*n*=5)	1.10	0.34	1.12	0.36

Rater Group	Group	Time 1	Std. Error	Time 2	Std. Error
EFL Teachers	Class A (*n*=5)	4.99	0.33	5.54	0.34
	Class B (*n*=5)	5.12	0.33	4.64	0.34
	NS Group (*n*=5)	1.52	0.33	1.42	0.34

A mixed-method analysis was carried out to determine if there were significant differences in the ratings of both groups of teachers based on class or time. The Type 3 Test of Fixed Effects found a significant effect of *Time*, $F(1, 791) = 1.20$, $p = 0.2735$, and no interaction of *Rater Group* over *Time*, $F(1, 791) = 1.59$, $p = 0.2073$. These results further strengthen the consistency of both groups of teachers in their ratings at times 1 and 2. However, the results also yielded a significant effect of *Rater Group*, $F(1, 791) = 41.49$, $p < 0.0001$, according to which the EFL teachers consistently rated the three groups of speakers (Class A, Class B, and NSs) as less

comprehensible, compared to the ESL teachers. Figure 1 below shows the different ratings given by the two groups of teachers over time (in general), whereas Figure 2 shows the same ratings given by both groups of teachers to the three different speaker groups (Class A, Class B, and NS).

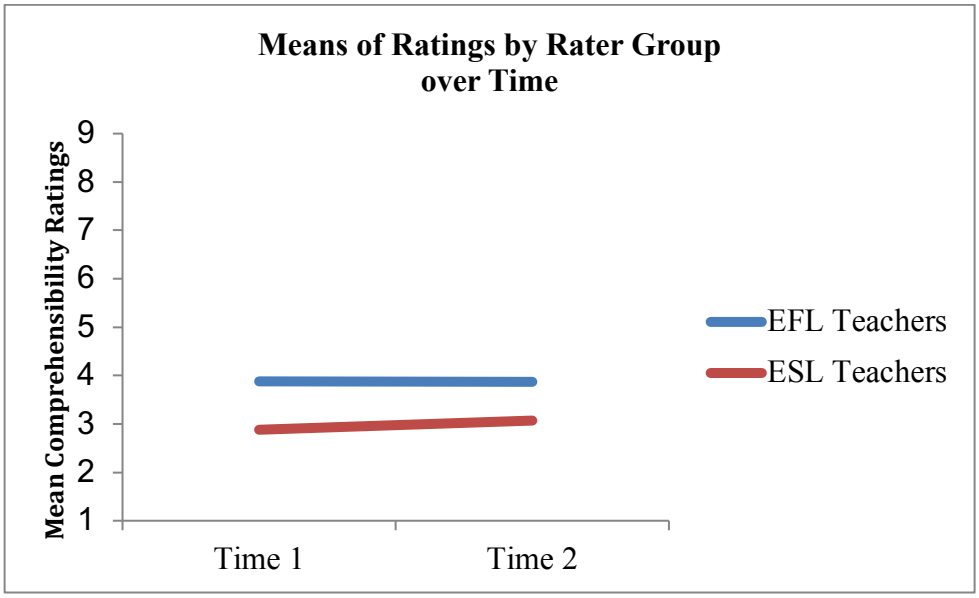

Figure 1. *Means of Ratings by Rater Group over Time.*

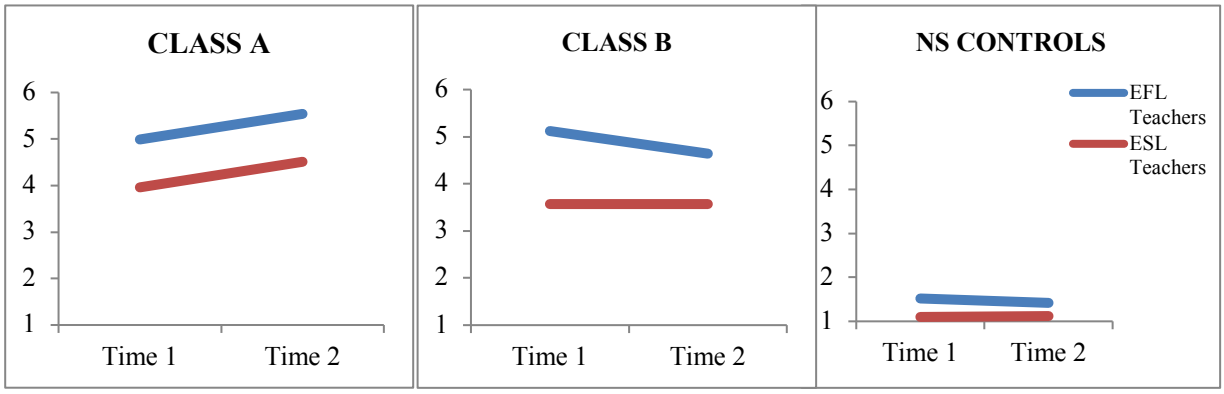

Figure 2. Means of Ratings by Rater Group over Time in Three Speaker Groups.

Qualitative

The qualitative results demonstrated commonalities between both groups of teachers' ratings, but they also showed important differences. One issue that both groups agreed that affected comprehensibility was problems with grammar and vocabulary. Mixing up tenses (e.g., not conjugating verbs in past tense) or using the wrong name for nouns (e.g., *bee* instead of *fly*) were pointed out as issues that affected comprehensibility.

The two groups of raters also mentioned that fluency and suprasegmentals affected the overall comprehensibility of the message. For example, very long pauses, or taking too much time to retrieve lexical items made L2 speech sound unnatural. Additionally, problems with word and sentence stress also made the speech more difficult to understand for both groups (see sample

comments *a* and *b* in Appendix 2). Whereas problems with fluency and suprasegmentals affected the overall comprehension of messages, problems at the segmental level hindered the comprehension of specific words. For example, substitution of one sound for another was the main issue affecting the comprehension of words (e.g., *mouth* sounded like *mouse*, or *playing* sounded like *praying*), but difficulties with the pronunciation of vowels, epenthesis, and clusters were also mentioned as factors that affected comprehensibility (see Appendix 2 *c* and *d*).

Although both groups had differences in familiarity with L1 backgrounds, many teachers recognized that such familiarity (or lack of it) may have affected their ratings. For example, some ESL teachers mentioned being stricter when rating speech that had an accent with which they were more familiar. In contrast, the EFL teachers were mostly familiar with Spanish-accented English, and they expressed that they had more difficulty understanding accents other than Spanish (see comments *e* and *f* in Appendix 2). Finally, and closely related to this last point, is the fact that the EFL teachers reported that accentedness resulted in lowered comprehensibility ratings for them; that is, they felt that they had struggled to overlook it. Although the EFL teachers claimed that they understood the messages, accentedness seemed to have forced them to concentrate harder on understanding the said messages—particularly when spoken with unfamiliar L1 accents. This did not seem to be the case with the ESL teachers, who claimed that the samples were comprehensible in spite of having different degrees of accentedness (see comments *g* and *h* in Appendix 2).

DISCUSSION AND CONCLUSIONS

Going back to the original research questions, the results demonstrated important differences and similarities between both groups of teachers: (1) *Do ESL and EFL teachers rate L2 speech comprehensibility in the same way?* According to the quantitative results, it appears that both groups followed similar tendencies in ratings, as shown by the very high levels of consistency between the two groups and by the absence of interaction between *Rater Group* over *Time*. However, the EFL teachers (all NNSs) were more severe when rating L2 speech, in contrast to the ESL teachers, who found the speech samples more comprehensible. Thus, the results here are consistent with previous research that has found NNSs to be more severe when rating L2 speech (Fayer & Krasinski, 1987; Kang, 2012; Rossiter, 2009). One possible reason for this is that NNSs may process L2 speech differently from NSs, and that they may also have slightly less robust representations for L2 words, which can affect how judgments about accentedness or comprehensibility are made (e.g. Levi, Winters & Pisoni, 2007).

The results also suggest that knowledge of different L2s and familiarity with accented speech may give certain listeners an advantage in terms of comprehensibility. Whereas the EFL teachers did not speak more than one L2 and had not spent a considerable amount of time in an L2 context, the ESL teachers spoke two or three L2s, had lived in L2 contexts, and were familiar with different accents because their classes routinely included a variety of learners with different L1 backgrounds. Therefore, it is possible that such familiarity gave the ESL teachers an advantage that helped them understand L2 speech more easily. As pointed out by Saito and Shintani (2016), more experience with L2 speech may give listeners a cognitive advantage to parse L2 speech more easily, which is reflected in the fact that it is easier for L2-experienced listeners to understand L2 speech as opposed to listeners who lack such familiarity with L2s (e.g., monolingual speakers). In the context of this study, the ESL teachers were NSs with a

higher level of experience listening to a variety of L2 accents, which may have made it easier for them to parse L2 speech. In contrast, the EFL teachers were NNSs and did not have much experience with different L2 accents. This differences in L2 background probably resulted in more difficulties for the latter group of teachers to parse the L2 speech, which they then perceived as less comprehensible.

As for the second research question (2) *On which linguistic aspects (i.e., phonological, lexical, syntactical) do ESL and EFL teachers base their ratings?*, the results also demonstrated important differences and similarities. For example, both groups indicated that syntactic and lexical problems affected the comprehension of L2 speech, a point which is consistent with previous studies demonstrating that comprehensibility is affected not only by phonological factors but also by lexical and syntactical issues (Crowther et al., 2016; Trofimovich & Isaacs, 2012). Additionally, the two groups of teachers reported that problems with suprasegmentals affected comprehension of the message as a whole, while problems at the segmental level affected comprehension of specific words. These insights expressed by both ESL and EFL teachers were also consistent with previous studies that attributed fluency and prosody as having major roles in comprehensibility (Crowther et al., 2016; Kang, Rubin, & Pickering, 2010; Saito, Trofimovich, & Isaacs, 2016) and that underscore the importance of segmentals in retrieving the meaning of specific words (see Derwing & Munro, 2015).

One of the key findings of this research is that even though the two groups of raters were ESL and EFL teachers, their ratings were in fact different when rating the same L2 speech samples. First, the ESL teachers had more experience with different accents and languages; this may have facilitated their parsing of L2 speech. This means that teacher education programs preparing teachers for pronunciation instruction should make an attempt not only to provide training in different theoretical and practical aspects of pedagogical phonology, but also to familiarize preservice teachers with different types of English accents—in both ESL and EFL contexts—to become aware of what constitutes intelligible and comprehensible L2 speech. This is particularly important in the current state of English in the world where NNSs of the language outnumber NSs (see Crystal, 1997; Graddol, 2006), and in which problems with comprehensibility and intelligibility may arise not only between speakers of the inner and outer circles, but more so among speakers of the outer and expanding circles (Jenkins, 2002; Levis, 2005).

Finally, and closely related to this last issue, is the fact that many teacher education programs do not necessarily provide a solid theoretical foundation for instructors to teach pronunciation (see Murphy, 2014; Thomson, 2013). Thus, the fact that there were differences in training in both groups of ESL and EFL teachers may raise the question of whether some of these teachers were treating both comprehensibility and foreign accent as an equal dimension—a question further implied by the EFL teachers' reports that accentedness was difficult to separate from their comprehensibility ratings. Such results may also question whether they are a reflection of classroom practices or not; that is, whether teachers try to unrealistically focus on eliminating their students' foreign accent instead of focusing on enhancing comprehensibility and intelligibility in their speech. Therefore, teacher education programs could help future preservice teachers not only by providing training on important aspects of phonetics and phonology, but also by exposing teachers to different theories of L2 phonology and L2 acquisition in general and to the appropriate practical application in class of different aspects of those theories in the teaching of L2 pronunciation.

ABOUT THE AUTHOR

Joshua Gordon is Assistant Professor in the Department of Languages and Literatures at the University of Northern Iowa. He has taught ESL, EFL, and Spanish as a foreign language. He has trained pre-service and in-service ESL and EFL teachers in the United States and in Costa Rica. He obtained a PhD in Second Language Studies from Indiana University, Bloomington. His research interests include L2 phonology and L2 pedagogy, particularly the interface between both areas in the teaching and learning of oral skills.

Contact Information

Department of Languages and Literatures
University of Northern Iowa
Bartlett Hall 2039
Cedar Falls, IA 50614
Telephone: (319) 273-2821

E-Mail: joshua.gordon@uni.edu

REFERENCES

Bent, T. & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America, 114,* 1600-1610.

Boersma, P. & Weenink, D. (2015). Praat: Doing phonetics by computer [Computer program] Version 5.3.23. Retrieved 6 May 2015 from http://www.praat.org/.

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

Crowther, D., Trofimovich, P., & Isaacs, T. (2016). Linguistic dimensions of second language accent and comprehensibility: Nonnative listeners' perspectives. *Journal of Second Language Pronunciation, 2,* 160-182.

Crystal, D. (1997). *English as a global language*. Cambridge, UK: Cambridge University Press.

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

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

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

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

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

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

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

Gass, S. M. & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Earlbaum Associates.

Gordon Zamora, J. (2015). *Teaching and Learning L2 pronunciation: A closer look at classroom and extra-classroom factors in the development of comprehensibility in ESL learners*. Unpublished doctoral dissertation, Indiana University.

Graddol, D. (2006). *English next: Why global English may mean the end of 'English as a Foreign Language.'* London, UK: The British Council.

Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective. *Language Testing, 29,* 163-180.

Hayes-Harb, R. & Hacking, J. F. (2015). Beyond rating data: What do listeners believe underlies their accentedness judgments? *Journal of Second Language Pronunciation,* 1, 1, 43-64.

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

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

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

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

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

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

MacKay, I. R., Flege, J. E., & Imai, S. (2006). Evaluating the effects of chronological age and

sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics, 27,* 157-183.

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

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

Murphy, J. M. (2014). ESL teachers are ready to teach pronunciation. In L. Grant (Ed.), *Pronunciation myths*. Ann Arbor: University of Michigan Press.

Richards, K. (2003). *Qualitative inquiry in TESOL*. New York: Palgrave Mcmillan.

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. & 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*(2), 217-240.

Simon's Cat. (2009, July 24). *Fly guy* [Video file]. Retrieved from https://www.youtube.com/watch?v=I1qHVVbYG8Y

Simon's Cat. (2010, February 10). *Snow business* [Video file]. Retrieved from https://www.youtube.com/watch?v=Tuf61OjvoPQ

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

Thomson, R. I. (2013). ESL teachers' beliefs and practices in pronunciation teaching: Confidently right or confidently wrong? In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 224-233). Ames: Iowa State University.

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

Winke, P. & Gass, S. (2013). The influence of second language experience and accent familiarity on oral proficiency rating: A qualitative investigation. *TESOL Quarterly, 47,* 762-789.

APPENDIX 1

Sample Guide for Stimulated-Recall Interview (based on Hayes-Harb & Hacking, 2015).

1. *Are there particular sounds that influenced your comprehensibility rating?*
2. *Are there particular words that influenced your comprehensibility rating?*
3. *Did the rhythm and melody of the speech influence your comprehensibility rating? If so, how?*
4. *Did the intonation influence your comprehensibility rating? If so, how?*
5. *Did the fluency and speed of speech influence your comprehensibility rating? If so, how?*

APPENDIX 2

Qualitative Findings from the Stimulated-Recall Interviews.

a. The message is comprehensible in the sense that if I put all the words together I get it, but then it should be more fluent and he isn't. […] I don't think the pronunciation of specific sounds is a problem. I guess that even though he is not a native speaker, his message is clear somehow. But in my opinion pronunciation is not the problem but his pauses, his lack of fluency. –EFLS06.

b. He was very slow, not fluent at all, he keeps correcting himself to the point I barely know which word he's trying to get, the [di] (*the*) for *that*. If his fluency was ok, I don't think that would be a big deal, but he's "*so*:::" and his vowels are a little off, but yeah very unnatural pauses. –ESLS07.

c. She says that there was a "fry frying" [instead of "fly flying")], so there were problems with /r-l/ that changed the meaning, and also "mouse" [instead of "mouth"] […] some words were weird or difficult for me because we are accustomed to saying "very" [vɛɹi] and not [bɛɹi]." –EFLN05.

d. I couldn't tell what she was saying there "no eight"? and then [bʌtsə˞] [instead of "but her…"] I also had a hard time with her "thetas" [θ]. "thought" sounds like [tɔt], maybe that's what she's trying to say?? I mean her rhythm sounds natural, but her accuracy in terms of sounds just wasn't there. So sometimes instead of "cat" she's extending it to [katə], making it almost like two syllables, but I could understand it but that's just more just accent. –ESLS05.

e. However, let's say that I got the message, but again, I guess that she is not a native speaker. I'm not sure what her L1 is but she sounds like she speaks an Asian language but I don't know which one because I don't know their differences. –EFLS02.

f. I probably have some biases when I hear more of the East Asian speakers I'm probably a little harsher than I am on others who are Middle Eastern or wherever they are from. I guess his native language is Kazakh or maybe Russian. He didn't sound so Kazakh there, he sounded more Russian. –ESLS06.

g. What I feel about his accent is like he's not a native speaker, so I don't know, it is an advantage for me because I can understand, I'm accustomed to listening to this with my students every single day, but as for the message then I feel like I have to pay attention to understand. […] but it [accent] gives me the idea that I have to pay extra attention so that I don't miss anything. Because he's not going to speak like a native speaker, so probably I would have to… I don't know, like really focus on what he's saying to understand. –EFLS03.

h. He sounds fluent but he doesn't sound native-like […] there were some native speakers in the samples and their speech was like, how can I say? more fluent, yeah. Fluency is another tricky aspect to define, but yeah he doesn't sound very fluent to me.