Linguistic Dimensions of Accentedness and Comprehensibility: Exploring Task and Listener Effects in Second Language French

Annie Bergeron
Concordia University

Pavel Trofimovich
Concordia University

Abstract: *This study explored the effects of task and listeners' linguistic background on judgments of accentedness and comprehensibility in second language (L2) French. Forty Spanish speakers of L2 French recorded a picture narrative and answered an interview question. These audio samples were assessed by native French listeners with (n = 10) and without (n = 10) knowledge of L2 Spanish using 1,000-point sliding scales to evaluate accentedness and comprehensibility, as well as nine linguistic variables targeting pronunciation, fluency, lexis, grammar, and discourse. In the picture narrative task (cognitively less complex), accentedness could be distinguished from comprehensibility at the level of the linguistic dimensions associated with each construct. While accentedness was related to pronunciation and fluency (consonant and vowel errors, intonation, speech rate), comprehensibility was additionally linked to lexis (accuracy, richness), grammar (accuracy, complexity), and discourse richness. However, in the interview task (cognitively more complex), both accentedness and comprehensibility had similar linguistic profiles, associated with all nine linguistic variables. Listeners' experience with L2 Spanish had little impact on listener ratings of accentedness or comprehensibility in both tasks. Findings are discussed in relation to L2 instruction and the role of task complexity in targeting various dimensions of L2 accentedness and comprehensibility.*

Key words: *French, accentedness, comprehensibility, listener experience, task effect*

Introduction

The ability to successfully communicate in a second language (L2) has become increasingly important as interactions across cultural and linguistic boundaries become increasingly frequent. Within this larger context, L2 communication is often equated with speakers' capacity to "pass" for a native speaker or their ability to sound nativelike (e.g., Jenkins, 2000). Levis (2005) described this belief in terms of the nativeness principle, referring to the reduction of a speaker's accent as being one of the goals of language teaching. Nativeness (defined in this manner) is typically measured through listeners' accentedness ratings, targeting how strongly L2 speech is influenced by the speaker's native language (L1) or is colored by other nonnative features (Derwing & Munro, 2015).

According to Levis (2005), intelligibility and comprehensibility offer another frame for understanding pronunciation. Intelligibility refers to listeners' actual understanding of L2 speech and is assessed through listeners' transcriptions of an utterance (e.g., Munro & Derwing, 1999) or their performance on cloze tests (e.g., Hayes-Harb, Smith, Bent, & Bradlow, 2008). Comprehensibility, which denotes listeners' perceived ease or difficulty in understanding a speaker's L2 speech, is traditionally measured through listener-based scalar ratings. Because comprehensibility ratings are highly correlated with intelligibility measures (Derwing & Munro, 2015), comprehensibility is both a user-friendly and a popular metric of understanding in a broad sense (Levis, 2005) included in various high-stakes assessment instruments (e.g., the Test of English as a Foreign Language [TOEFL], the International English Language Test System [IELTS]). Focusing on the constructs of intelligibility and comprehensibility shifts the emphasis from nativelike speech to a focus on attaining understandable speech.

Regardless of L2 speakers' orientation toward nonaccented speech or toward intelligible or comprehensible performance,

a fundamental question for language researchers and instructors wishing to help their learners attain their goals is which linguistic dimensions of L2 speech are associated with accent and which are linked to comprehensibility. This study explored this question in L2 French. The specific goal was to determine to what extent the pronunciation, fluency, lexis, and grammar dimensions of speech might be associated with L2 French speakers' accentedness (global measure of nativelikeness) and which might be uniquely linked to their comprehensibility (global measure of understanding) across two tasks (a picture narrative and an interview) for listeners who had and did not have prior experience with speakers' L1.

Review of Literature

Linguistic Correlates of Accentedness and Comprehensibility

Research has shown that even in the presence of a strong accent, L2 utterances can still be fully understood (e.g., Derwing & Munro, 2015). However, what is less obvious is whether accentedness and comprehensibility can be distinguished at the level of the linguistic dimensions underlying each construct. For instance, various measures of pronunciation and fluency have been linked to the comprehensibility and intelligibility of L2 speech, including stress (Field, 2005), speech rate (Munro & Derwing, 2001), pitch range and pausing (Kang, Rubin, & Pickering, 2010), and poor grammar and inappropriate lexical choice (Fayer & Krasinski, 1987). Many linguistic dimensions have also been associated with L2 accentedness, including segmental accuracy (Derwing, Munro, & Wiebe, 1998), articulation rate (Baker & Trofimovich, 2006), pitch range, stress, and pausing (Kang, 2010). Trofimovich and Isaacs (2012) investigated 19 linguistic measures grouped under the categories of phonology, fluency, lexis, grammar, and discourse in an attempt to distinguish the linguistic correlates of comprehensibility

and accentedness. They showed that for French speakers of L2 English, comprehensibility was linked to measures of phonology, grammar, lexis, and discourse whereas accentedness was uniquely associated with measures of phonology. In a follow-up study targeting 120 Japanese speakers of L2 English, Saito, Trofimovich, and Isaacs (2016) reported that comprehensibility was linked to eight measures across multiple linguistic categories whereas accentedness was tied only to segmental accuracy. These findings, albeit specific to L2 English, led to the conclusion that accentedness and comprehensibility could be distinguished at the level of the linguistic dimensions underlying each construct (see also Saito, Trofimovich, & Isaacs, 2015).

However, because little research has been carried out in languages other than English, one question that remains unclear is whether the linguistic correlates of comprehensibility, as distinct from those underlying accentedness, also differ for L2s other than English. In a rare study targeting L2 German, O'Brien (2014) investigated the association between the linguistic categories of phonology, fluency, lexis, and grammar and the constructs of accentedness and comprehensibility for learners of L2 German as evaluated by L2 German listeners. In that study, comprehensibility and accentedness were predicted by a broad range of largely overlapping measures, suggesting that at least where L2 listeners are concerned, the linguistic independence of accentedness from comprehensibility may be less robust than previously thought. In a recent accentedness and comprehensibility study targeting L2 French, Trofimovich, Kennedy, and Blanchet (in press) showed that for classroom learners of L2 French assessed by native-speaking French listeners, both constructs were associated with intonation errors and pitch range, with comprehensibility additionally linked to fluency (longer speech runs, fewer hesitations). However, the analyses included only phonology and fluency measures of L2

speech, making it unclear how aspects of lexis and grammar feed into both constructs. Thus, the ostensibly robust findings from L2 English regarding the linguistic correlates of accentedness and comprehensibility need to be extended to other L2s to better understand whether the linguistic signatures of both constructs are specific to the target language.

Task Effects
The bulk of prior studies targeting the linguistic correlates of accentedness and comprehensibility has one methodological feature in common: the use of a single task (picture narrative) in which L2 speakers describe a sequence of ordered images. While this allows for direct comparisons of results (e.g., Munro & Derwing, 1999; O'Brien, 2014; Saito et al., 2015; Trofimovich & Isaacs, 2012), this clearly limits language educators' understanding of the extent to which L2 speakers' performance may differ across tasks, particularly because studies have shown that speakers pay attention to different linguistic dimensions depending on the cognitive demands of a speaking task (e.g., Long & Crookes, 1992; Robinson, 2005; Skehan, 2009). For instance, according to Skehan's (2009) limited capacity model, L2 speakers show greater performance in accuracy or complexity depending on the attentional resources that are required to process task content. Similarly, Robinson's (2005) cognition hypothesis holds that different tasks may elicit different types of linguistic output depending on such task characteristics as the availability of planning time or the requirement to discuss abstract topics.

Most commonly, researchers have examined task effects on L2 production by targeting grammar and vocabulary, generally reporting that task type influences the accuracy and complexity of language output (see Plonsky & Kim, 2016). To give a few examples, Michel, Kuiken, and Vedder (2007) observed that L2 speakers showed greater grammatical accuracy and lexical

complexity (fewer omissions, lexical and syntactic errors, and more lexically diverse speech) in a dialogic task than in a monologic task in L2 Dutch. Similarly, Révész (2011) found that L2 English speakers demonstrated enhanced grammatical accuracy and more varied lexis but also decreased syntactic complexity in a more complex task than in a simpler task.

To date, limited research has focused on how task complexity may impact various dimensions of L2 speech, with most prior work targeting fluency. For example, L2 speakers show greater fluency when engaged in dialogic tasks than in monologic tasks (Ejzenberg, 2000), with picture narratives eliciting speech rated as significantly less fluent than language output in monologic and conversational tasks (Derwing, Rossiter, Munro, & Thomson, 2004). For segmental accuracy, Rau, Chang, and Tarone (2009) reported task effects on the production of /θ/ (as in *think*) by speakers of L2 English from China and Taiwan. Word and passage reading were associated with increased accuracy compared to less formal tasks (interview, story retell). Crowther, Trofimovich, Isaacs, and Saito (2015) recently contrasted two speaking tasks (an IELTS long-turn task and a TOEFL iBT integrated task (see https://www.ets/org/toefl/ibt/about), investigating which of 10 rated linguistic dimensions of L2 output were associated with comprehensibility in each task. Whereas the phonology and fluency variables contributed to comprehensibility in both tasks, the lexis, grammar, and discourse variables showed stronger associations with comprehensibility in the more complex task (the TOEFL iBT) than in the simpler (the IELTS) task. However, to our knowledge, no research has examined task effects on linguistic variables associated with accentedness vs. comprehensibility for languages other than English.

Listener Effects

Because most human communication is dialogic in nature, both speakers and listeners play an important role in the success of a communicative exchange, particularly where L2 speakers are involved. For this reason, multiple studies have explored the effects of matched vs. mismatched listeners' and speakers' L1 backgrounds on various speech measures, including accentedness and comprehensibility. For instance, Munro, Derwing, and Morton (2006) did not find strong effects of listeners' L1 on the comprehensibility and intelligibility of L2 speech, reporting small shared language benefits for two of the four speaker and listener groups. These mixed results align with those of Major, Fitzmaurice, Bunta, and Balasubramanian (2002), who found that Spanish listeners benefited from receiving a lecture in Spanish-accented English while Chinese listeners did not enjoy a shared listener-speaker language benefit. However, Kang, Vo, and Moran (2016), who examined how listeners from multiple language backgrounds evaluated the speech of Vietnamese speakers of L2 English, reported strong effects of both listeners' L1 and the learning context (i.e., the pedagogical focus in L2 instruction) on multiple speech measures. In L2 German, O'Brien (2016) found no major differences in accentedness ratings between native and L2 German listeners; however, L2 listeners provided harsher comprehensibility ratings when compared to German listeners, likely due to L2 listeners' lower ability level.

While the effects of listeners' L1 background on assessments of L2 speech are complex and may depend on the specific L1-L2 combinations of listeners and speakers, the influence of listeners' experience—operationalized in terms of exposure to L2 speech, training in linguistics or phonetics, or experience with speakers' L1—has received consistent research support. With respect to listeners' exposure to L2 speech, Gass and Varonis (1984), for instance, found that L2 speakers were more intelligible to native English speakers who were familiar with particular accents, specific speakers, and L2 speech in general. As shown by Kennedy and Trofimovich

(2008), English as a second language teachers judged Mandarin speakers' L2 English speech as more intelligible, compared to that of native-speaking listeners who had little experience with L2 speech. Experienced listeners with training in applied linguistics and teaching experience also appear to be more consistent than novice listeners in their assessments of various dimensions of L2 speech, including comprehensibility and accentedness (Isaacs & Thomson, 2013), especially when evaluating less intuitive and more complex linguistic variables, such as lexical and grammatical accuracy (Saito et al., 2015).

Trained raters' familiarity with the speakers' L1 may also lead to a bias in their ratings, such that raters can show more leniency toward the L2 speakers if raters demonstrate some knowledge of the speakers' L1. For example, when comparing raters with self-assessed extended exposure vs. no exposure to speakers' L1, Carey, Mannell, and Dunn (2011) found that the former group gave speakers higher ratings on a pronunciation proficiency scale than did the latter group. In addition, listeners with a basic knowledge of L2 Spanish and Chinese were shown to be more lenient toward native speakers of Spanish and Chinese when judging their L2 English (Winke, Gass, & Myford, 2012). Xi and Mollaun (2009) also demonstrated that bilingual raters were more consistent in assessing L2 English speakers on the TOEFL iBT than were monolingual English raters. In sum, in an investigation of listener judgments of L2 speech, it would be crucial to consider listeners' familiarity with the speakers' L1 even if their familiarity and knowledge of the L2 is not extensive.

The Current Study

In light of previous research, whose focus has been nearly exclusively on speakers of L2 English performing picture narrative tasks, it is essential to look at the linguistic correlates of accentedness and comprehensibility to understand to what extent

they differ for L2s other than English and as a function of different speaking tasks and listeners' familiarity with the speakers' L1. To address this issue, the current study examined the linguistic dimensions of accentedness and comprehensibility for 40 native Spanish speakers of L2 French, targeting nine rated linguistic variables in their speech (spanning the domains of segmental and prosodic accuracy and fluency as well as lexis and grammar complexity and accuracy) across two tasks (a picture narrative and an interview) and for two sets of listeners (those familiar with and unfamiliar with Spanish). This study was guided by two research questions:

1. To what extent do the linguistic correlates of accentedness and comprehensibility in L2 French differ as a function of the speaking task performed?
2. To what extent do the linguistic correlates of accentedness and comprehensibility in L2 French differ as a function of raters' familiarity with the speakers' L1 (Spanish)?

Method

Participants

The audio materials included 40 recordings by native Latin American Spanish speakers of L2 French (21 males, 19 females) from Colombia (23), Venezuela (8), Ecuador (2), Cuba (2), Mexico (2), Peru (1), Panama (1), and Paraguay (1). The speakers were chosen from Latin American countries with shared aspects of L1 Spanish phonology, which include the neutralization of the distinction between /θ/ and /s/ (seseo), the opposition between /ʎ/ and /j/ (yeísmo), the aspiration of /x/, and the weakening and elision of voiced stop consonants between vowels (Canfield, 1981). The speakers ($M_{age} = 36.2$ years, range = 24–51) all started learning French after age 18 and had completed at least a beginner-level L2 French course in their home country. At the time of the study, they had lived in Montreal, where

French is the majority language, for a period of 2 months to 22 years ($M = 4.2$ years) and were students (18) in the *Certificat en français écrit pour non-francophones* (certificate in written L2 French), first-year graduate students (8) in a French-medium university, or full-time professionals (14) living and working in a French-speaking environment. Using a six-level proficiency flow chart (White & Turner, 2005), the speakers were rated by the first author as representing a range of L2 French ability, from intermediate to advanced ($M = 4.03$, range = 1–6).

The raters included 20 native speakers of Québec French (15 females, 5 males) who hailed from francophone families, completed all their education in French, and reported using French the majority of the time daily (91%), including to communicate with French speakers (82.5%). All raters ($M_{age} = 33.2$ years, range = 24–44) were experienced in the sense that they were pursuing an education or pedagogy degree at a French-medium university and were thus familiar with language learning and teaching issues (e.g., Isaacs & Thomson, 2013; Rossiter, 2009). Using a nine-point scale (1 = not at all familiar, 9 = very familiar), they self-reported high familiarity with

foreign-accented French ($M = 7.3$, range = 3–9) and specifically with Spanish-accented French ($M = 5.3$, range = 1–9). The 20 raters, who represented an otherwise relatively homogenous group (see Table 1 for background characteristics), were assigned to two equal groups based on their self-reported familiarity with L2 Spanish: those with no knowledge of Spanish and those who had taken a beginner-level course in L2 Spanish as part of their university studies. Thus the chief distinction between the two groups was their experience taking an L2 Spanish course.

Tasks

For this study, the two tasks that were recorded by the speakers (the picture narrative and the interview response) were selected to examine the effects of the speaking task on the linguistic correlates of accentedness and comprehensibility. The picture narrative task was based on two short stories from Jiang, Green, Henley, and Masten (2009), with images portraying the same two characters (a father and his young son baking a cake and going on a fishing trip). The speakers recorded the two narratives in sequence, in the same order,

TABLE 1

Raters' Background Characteristics (Means and Standard Deviations)

Background variables	No Spanish ($n = 10$)	L2 Spanish ($n = 10$)
L1	Québec French	Québec French
Self-assessed experience in L2 Spanish	—	Intermediate
Gender (m/f)	4/6	1/9
Age	33.2 (5.18)	29.9 (6.14)
Daily use of French (0–100%)	89.0 (14.49)	93.0 (6.75)
Media use in French (0–100%)	61.0 (35.10)	68.0 (20.44)
Interactions with French speakers (0–100%)	78.0 (28.21)	87.0 (8.23)
Familiarity with accented French	7.9 (1.10)	6.7 (1.94)
Familiarity with Spanish-accented French	6.0 (2.75)	5.4 (1.84)

and were given 30 seconds to look at the images before narrating each story. The recordings of the first story were used as practice files during the rating sessions (see "Speech Rating" below), while the recordings from the second story served as the target materials from this task. The speakers also participated in an informal interview with the first author, responding to the following prompt (originally in French): "Now that you are living in Montreal, what advice would you give a friend of yours living in your home country who would like to move to Montreal? What would you recommend that he or she do? What should he or she expect?" The goal was to elicit spontaneous speech, as speakers were unlikely to have experienced this topic through classroom instruction. The interview was the last task conducted, in order to reduce speakers' anxiety and increase familiarity with the interviewer.

According to Robinson (2001, 2007), tasks can be distinguished in terms of their complexity, associated with cognitive demands imposed by the task design. In this sense, the picture narrative appeared to be less cognitively complex than the interview response. The narrative task included a restricted character set, such as a man, a boy, and a policeman (+few elements) and a familiar theme (+prior knowledge); it required the use of the present tense (+here and now), with speakers given 30 seconds of preparation time (+planning time). As the raters were familiarized with the target images during the rating sessions, they were also likely oriented toward speakers' expected use of target vocabulary (e.g., Schmid & Yeni-Komshian, 1999), such as *canne à pêche* (fishing rod). Speakers (and raters) thus experienced little linguistic and thematic freedom, being restricted by the depicted content. In contrast, the interview response ostensibly imposed greater demands on the speaker and the listener, with responses generally not restricted to set topics or themes covered in classroom instruction (–few elements) and no preparation time available (–planning

time). This task also required speakers to deploy hypothetical and/or conditional structures as well as multiple temporal references (–here and now), leaving much freedom for them to decide how to formulate their response and, most important, what content to target, such as the importance of speaking languages, the lack of job opportunities, or respect for human rights. This enhanced freedom for the speaker to make various linguistic choices in the interview task also likely made the task of extracting meaning from speech more complex for the rater, at least compared to the picture narrative.

Speakers' output from both tasks (range = 32–660 seconds in duration) was recorded using a digital voice recorder (VN-8100PC), with all audio files subsequently matched for peak amplitude and then truncated to include the first 30 seconds of speech (excluding initial dysfluencies), in line with prior research targeting 30- to 60-second speech samples (e.g., Crowther et al., 2015; Derwing, Munro, & Thomson, 2008). All recordings were also transcribed by the first author, adhering to the following conventions to minimize pronunciation and fluency as possible confounds in judgments of lexis, grammar, and discourse (Crossley, Salsbury, & McNamara, 2014): removing hesitation markers, spelling clues signaling pronunciation-specific errors (e.g., *bonjour* [bɔ̃ʒuʁ] pronounced as *bonyou* [bɔnju] was transcribed as the targeted word *bonjour*), and punctuation to avoid transcriber influence. The resulting 80 audio files and 80 transcripts (40 speakers × two tasks) served as stimuli for the 20 raters' evaluations.

Speech Rating

The researcher (the first author) met each rater twice in a quiet office within 2 to 4 weeks, with a minimum of 3 days between the sessions. In the first session (2 hours), each rater assessed the audio files from both tasks (presented in counterbalanced orders

in separate blocks) for accentedness and comprehensibility. In the second session (3 hours), each rater evaluated the audio files (presented in counterbalanced blocks by task) for four phonology and fluency variables and then rated the transcripts (again in separate counterbalanced blocks by task) for five lexis and grammar measures (described in "Rated Categories" below). All 80 audio files and 80 transcripts were evaluated by the raters. The amount of time needed by each rater varied as they were not required to listen to the entire excerpt when rating for accentedness and comprehensibility, but they had the choice to repeat the audio files as many times as necessary and had the transcripts available to them for as long as needed in the second session. Before each set, raters were first trained on the relevant variables, with constructs presented in training booklets with examples. They then carried out five practice ratings to familiarize themselves with the L2 speech (Bradlow & Bent, 2008), the rating procedure, and the computer interface. For the picture narrative, the practice samples were drawn from the first (nontarget) narrative; for the interview response, the practice samples included speakers' responses to an unrelated question, originally asked in French (What are the advantages and disadvantages of the job you were practicing in your home country?).

For each set of audio- and transcript-based ratings, the raters used custom-based presentation software programmed in MATLAB that featured 1,000-point scales with a free-moving horizontal slider, initially appearing in the middle, with the left endpoint representing the rating of 0 and the right endpoint corresponding to the rating of 1,000. Each scale included a short description of the endpoints, with no numerical markings shown for the raters (see Table 2 for brief descriptions of each measure and scale endpoints). For all rating sets, all scales were visible simultaneously on the computer screen as simultaneous vs. consecutive scale presentation producing little difference in rating (O'Brien, 2016). In line with prior

research (e.g., Saito et al., 2015), for rating accentedness and comprehensibility, raters heard each file only once, since these represent relatively intuitive dimensions of speech. However, raters could listen to each recording multiple times to assess the four phonology and fluency variables; they were unable to record their rating until the file was played once in full. For the five lexis and grammar variables, each transcript remained on the screen until the rater felt satisfied with the assessments. Raters were encouraged to use the whole scale range and were told that even a slight movement of the slider could make an important change in the score. They were also informed that they could change their ratings multiple times until they proceeded to the next file or transcript. Each rater evaluated audio files or transcripts within each block using a unique randomization per rater.

Rated Categories

As shown in Table 2, the first set of rated measures included the global ratings of accentedness (defined in terms of how strongly speakers' speech was influenced by their L1 or was colored by other nonnative features) and comprehensibility (presented as the degree of ease or difficulty with which raters could understand speakers' speech). For the set of phonology and fluency ratings, raters evaluated speakers using each speaker's audio file for the following four variables (see Table 2):

1. Consonant errors, defined as errors at the level of individual consonants. These errors included phoneme substitutions (e.g., *essaminer* instead of *examiner* [to examine]), deletions (e.g., *macher* for *marcher* [to walk]), or insertions (e.g., *enfant* pronounced as *enfante* [child]).
2. Vowel errors, defined as errors at the level of individual vowels. These errors also included phoneme substitutions (e.g., *processous* instead of *processus* [process]), deletions (*agréable* perceived

TABLE 2

**Rated Variables, Endpoint Descriptors, Measure Summary
(Translated from French)**

Rated variable	Left endpoint	Right endpoint	Summary
Global			
Accentedness	Heavily accented	No accent at all	How different a speaker sounds from a native French speaker
Comprehensibility	Impossible to understand	Easy to understand	Ease or difficulty of raters' understanding of L2 speech
Phonology and fluency			
Vowel errors	Frequent	Infrequent or absent	Errors in production of individual vowels within a word
Consonant errors	Frequent	Infrequent or absent	Errors in production of individual consonants within a word
Intonation	Unnatural	Natural	Appropriateness of pitch moves within an utterance, such as rising tones in yes/no questions
Speech rate	Too slow or too fast	Optimal	Speakers' overall pacing and speed of utterance delivery
Lexis, grammar, and discourse			
Lexical appropriateness	Many inappropriate words used	Consistently uses appropriate vocabulary	Speakers' choice of words to accomplish a speaking task
Lexical richness	Few, simple words used	Rich and varied vocabulary	Sophistication of the vocabulary used by a speaker
Grammatical accuracy	Many grammar errors	Absence of grammar errors	Number of grammar errors made by a speaker
Grammatical complexity	Simple grammar	Elaborate grammar	Sophistication of the grammar used by a speaker
Discourse richness	Simple structure, few details	Detailed and sophisticated	Richness and sophistication of the utterance content

as *agrable* [pleasant]), or insertions (*eskier* for *skier* [to ski]).

3. Intonation, defined as the melody of speech, with reference to natural fluctuations of pitch across utterances. Natural intonation patterns were exemplified with respect to the (expected) rising patterns in yes/no questions (e.g., *Souperas-tu à la maison demain?* ↑ [Will you be home for dinner tomorrow?]), falling patterns at the end of affirmative statements (e.g., *Oui, je vais être à la maison* ↓ [Yes, I will be home]), or natural pitch moves in longer utterances (e.g., *Oui, je vais être à la maison* ↓ . . . ↑ *mais jusqu'à 7 heures* ↓ [Yes, I will be home . . . but until 7 o'clock]).

4. Speech rate, defined as the overall speed of the speech delivery. A low rating for this measure would correspond to slow and laborious speech delivery, full of pauses and hesitations, or to speech that is too fast to be comfortable to listen to. A high rating would indicate that speech flow is optimal for listening, with speech delivered naturally.

For the set of lexis and grammar ratings, raters evaluated speakers using each speaker's transcript for the following five variables (see Table 2):

1. Lexical accuracy, defined as the appropriateness of words used by the speaker to accomplish the task. Low lexical accuracy includes incorrect or inappropriate word usage as well as lexical intrusions from other languages (e.g., *Ils prennent l'arme pour shooter* [They take the gun to shoot], where *shooter* was used in place of *tirer* [to shoot]). Inappropriate word usage corresponds to the use of more generic, extended word meanings instead of more precise lexis (e.g., *Le petit garçon et son grand-père sont allés pêcher dans l'eau* [The little boy and his grandfather went fishing on the water], where *étang* [pond] would be more appropriate than *l'eau* [water]).

2. Lexical richness, defined as a range and variety of words used by the speaker.

Poor lexical richness would correspond to multiple repeated tokens of the same word and the general absence of sophisticated vocabulary (e.g., *Dans la première **image**, on voit le **garçon** et dans la deuxième **image**, c'est encore le même **garçon*** [In the first **picture**, we see a **boy** and in the second **picture**, it's the same **boy**], compared to *L'enfant dans la première image semble malheureux, alors que le bambin suivant démontre de la joie* [The kid in the first picture looks sad, while in the following one, the child shows some happiness]).

3. Grammatical accuracy, defined as the number of grammar errors concerning individual words or longer units (such as sentences) made by the speaker. Grammar errors include errors of word order (*C'est une **école petite*** instead of *C'est une **petite école*** [It's a school small] instead of [It's a small school]), conjugation (*Il faut que je **dis** la bonne réponse* for *Il faut que je **dise** la bonne réponse* [I have to say the right answer]), and agreement (*Ce sont des **petits enfants*** produced as *Ce sont des **petit enfants*** [These are small children], that is, without the liaison between the adjective and the noun).

4. Grammatical complexity, defined as the sophistication of the speaker's grammar. Raters were asked to attend to the use of complex (subordinated) multiclause utterances in place of multiple (coordinated) structures. For example, *Le petit garçon allait à la pêche alors qu'il voyait son grand-père au loin qui s'approchait* [The little boy was on his way to go fishing while he saw his grandfather approaching in the distance] includes more elaborate structures, compared to the same idea expressed in short, simple utterances, as in *Le garçon va à la pêche . . . le grand-père est proche . . . il va aussi pêcher* [The boy goes fishing . . . the grandfather is close . . . he is also going fishing]).

5. Discourse richness, defined as the richness and coherence of speech content.

TABLE 3

Rater Consistency (Cronbach's α) for Rated Categories by Task and Rater Group

Rated variable	Narrative		Interview response	
	No Spanish	With Spanish	No Spanish	With Spanish
Accentedness	0.94	0.95	0.95	0.95
Comprehensibility	0.91	0.91	0.94	0.93
Consonant errors	0.89	0.90	0.85	0.91
Vowel errors	0.89	0.91	0.90	0.91
Intonation	0.84	0.80	0.81	0.75
Speech rate	0.91	0.93	0.89	0.90
Lexical accuracy	0.86	0.87	0.89	0.85
Lexical richness	0.88	0.85	0.91	0.89
Grammatical accuracy	0.76	0.86	0.85	0.89
Grammatical complexity	0.84	0.83	0.90	0.89

Low discourse richness corresponds to a simple narrative or a response that lacks details and features an undeveloped story structure. High discourse richness corresponds to narratives or responses with a high degree of detail and developed, sophisticated content.

Understanding of Rated Categories and Rating Consistency

After completing each rating set, raters filled out a posttask questionnaire (9-point scales) to elicit data on their understanding of the rated categories (1 = *je n'ai pas compris du tout* [I did not understand at all], 5 = *neutre* [neutral], 9 = *j'ai très bien compris* [I understood this concept well]) and the ease with which these categories could be applied (1 = *très difficile* [very difficult], 5 = *neutre* [neutral], and 9 = *très facile* [very easy]). Raters from both groups demonstrated a clear understanding of all rated categories (M = 8.32, range = 7.90–8.80) and considered them easy to use (M = 7.26, range = 6.20–8.40). As shown in Table 3, high interrater consistency was

observed for the raters' evaluations of accentedness (Cronbach's α = 0.94–0.95) and comprehensibility (α = 0.91–0.94), as well as for all rated linguistic categories, with α values ranging between 0.75 and 0.93, in all cases exceeding the recommended 0.70–0.80 benchmark (Larson-Hall, 2010). Reliability values were thus deemed sufficiently high for individual raters' assessments to be averaged for each rater group, separately for each category.

Results

Global Ratings

The first analysis targeted the global ratings of accentedness and comprehensibility (summarized in Table 4), with the two sets of ratings submitted to a three-way ANOVA with repeated measurements for dimension (accentedness, comprehensibility), rater group (with Spanish, no Spanish), and task (picture narrative, interview response). This analysis yielded no statistically significant main effects or interactions, $F < 3.01$, $p > 0.09$, $\eta_\rho^2 < 0.07$, suggesting that L2 speakers' output across two tasks

TABLE 4			
Means (Standard Deviations) for Accentedness and Comprehensibility (0–1,000 Scale)			
Task	Global measure	No Spanish	With Spanish
Narrative	Accentedness	686 (238)	669 (224)
	Comprehensibility	681 (171)	645 (191)
Interview response	Accentedness	646 (246)	667 (250)
	Comprehensibility	660 (211)	647 (215)

elicited similar ratings of accentedness and comprehensibility for raters with and without knowledge of Spanish.

Rating of Linguistic Categories

The next analysis focused on the nine rated linguistic dimensions of speech, as evaluated by the two rater groups in picture narratives and interview responses. Because the nine linguistic categories represented multiple and possibly overlapping dimensions, it was first important to uncover potential underlying relationships between them. Therefore, the rated dimensions for the 40 speakers were submitted to a principal component analysis (PCA) with Oblimin rotation, carried out separately per task for each rater group. The Kaiser–Meyer–Olkin values (0.83–0.88) all exceeded the minimum of 0.60 recommended by Hutcheson and Sofroniou (1999), who interpreted the factorability of such correlation matrixes as "meritorious" (p. 225). Bartlett's test of sphericity also showed that correlations were high enough for PCA: $\chi^2(36) = 413.50$, $p < 0.0001$ (picture narrative, raters with no knowledge of Spanish), $\chi^2(36) = 516.05$, $p < 0.0001$ (interview response, raters with no knowledge of Spanish), $\chi^2(36) = 359.28$, $p < 0.0001$ (picture narrative, raters with L2 knowledge of Spanish), and $\chi^2(36) = 460.75$, $p < 0.0001$ (interview response, raters with L2 knowledge of Spanish).

The underlying dimensions obtained from the PCA analyses in the picture narrative task are summarized in Table 5. In this task, there were two underlying linguistic factors for raters with no knowledge of Spanish (accounting for 82% of the variance). The variables that loaded on the first factor included consonant and vowel errors, intonation, and speech rate; therefore, this factor was labeled *phonology and fluency*. The remaining linguistic variables —all measures of lexis, grammar, and discourse richness and accuracy—patterned together; this factor was thus named *lexicogrammar richness and accuracy*. For raters with an L2 knowledge of Spanish, in the picture narrative task the PCA revealed three nonoverlapping factors (accounting for a total of 89% of variance). Once again, the measures of consonant and vowel errors, intonation, and speech rate loaded on the first factor, which was named *phonology and fluency*. Three measures dealing with lexicogrammar richness and complexity (lexical richness, discourse richness, and grammatical complexity) loaded on the second factor, which was described as *lexicogrammar richness and complexity*. Finally, both measures of lexis and grammar accuracy patterned together under the third factor, which was termed *lexicogrammar accuracy*.

The underlying dimensions obtained from the PCA analyses in the interview response are summarized in Table 6. Unlike the analyses targeting the picture narrative task, the PCA output for the interview task failed to distinguish separate underlying dimensions across the nine rated categories.

TABLE 5

Summary of PCA Solutions (Factor Loadings) for Linguistic Variables by Task and Rater Group in Picture Narrative

Rater group	PCA factor	Linguistic variable
No Spanish	Phonology and fluency	Consonant errors (0.99)
		Intonation (0.87)
		Vowel errors (0.85)
		Speech rate (0.60)
	Lexicogrammar richness and accuracy	Grammatical complexity (1.02)
		Lexical richness (0.97)
		Discourse richness (0.86)
		Lexical accuracy (0.73)
		Grammatical accuracy (0.69)
With Spanish	Phonology and fluency	Vowel errors (0.97)
		Consonant errors (0.82)
		Intonation (0.73)
		Speech rate (0.49)
	Lexicogrammar richness and complexity	Lexical richness (0.89)
		Discourse richness (0.88)
		Grammatical complexity (0.85)
	Lexicogrammar accuracy	Grammatical accuracy (0.84)
		Lexical accuracy (0.76)

Note: All eigenvalues > 1.

In fact, for both groups of raters, all nine linguistic categories formed a single factor, in each case accounting for 77% (raters with knowledge of Spanish) and 78% (raters with no knowledge of Spanish) of the total variance.

Linguistic Correlates of Accentedness and Comprehensibility

The final analysis sought to determine the possible linguistic correlates of accentedness and comprehensibility ratings in the picture narrative and interview response

TABLE 6

Summary of PCA Solutions for Linguistic Variables by Task and Rater Group in Interview Response

Rater group	PCA factor	Linguistic variable
No Spanish	Phonology, fluency, lexis, grammar, and discourse	Nine linguistic categories (total eigenvalue = 7.04)
With Spanish	Phonology, fluency, lexis, grammar, and discourse	Nine linguistic categories (total eigenvalue = 6.90)

TABLE 7

Pearson Correlations Between PCA Linguistic Factors and Accentedness and Comprehensibility by Task and Rater Group in Picture Narrative

Rater group	PCA factor	Accentedness	Comprehensibility
No Spanish	Phonology and fluency	0.77*	0.57*
	Lexicogrammar richness and accuracy	0.25	0.58*
With Spanish	Phonology and fluency	0.83*	0.83*
	Lexicogrammar richness and complexity	0.34	0.67*
	Lexicogrammar accuracy	0.32	0.50*

Notes: *$p < 0.001$, $\alpha = 0.004$.

>tasks for each rater group. For this analysis, Pearson correlations were computed between the ratings of accentedness and comprehensibility and the linguistic dimensions yielded through PCA analyses, separately per task and each rater group (Bonferroni-corrected $\alpha = 0.004$), with linguistic factor scores derived through the Anderson–Rubin method of obtaining noncorrelated factor values. Correlation analyses for the picture narrative task are summarized in Table 7. In this task, for both groups of raters, accentedness was uniquely associated with the *phonology and fluency* factor whereas comprehensibility was tied with all the factors, which together encompassed measures of phonology, fluency, lexis, grammar, and discourse. In all cases, the associations were medium to large in strength ($r \geq 0.50$), based on Plonsky and Oswald's (2014) recommendations.

Correlation analyses for the interview response task are summarized in Table 8. In the interview task, a single factor encompassing all nine linguistic variables was significantly correlated with both accentedness and comprehensibility (with strong associations in both cases), although the strength of the relationship was numerically larger for comprehensibility ($R^2 > 0.67$) than for accentedness ($R^2 = 0.46$).

Discussion

The aim of the current study was to investigate whether the linguistic correlates of accentedness and comprehensibility

TABLE 8

Pearson Correlations Between PCA Linguistic Factors and Accentedness and Comprehensibility by Task and Rater Group in Interview Response

Rater group	PCA factor	Accentedness	Comprehensibility
No Spanish	Phonology, fluency, lexis, grammar, and discourse	0.68*	0.84*
With Spanish	Phonology, fluency, lexis, grammar, and discourse	0.68*	0.82*

Notes: *$p < 0.001$, $\alpha = 0.004$.

depended on the demands of a speaking task and raters' familiarity with the speakers' L1. To explore task effects, L2 speakers' language production was compared across two tasks that varied in the cognitive demands that were imposed on speakers and in the predictability of the speech content for listeners, with the picture narrative task considered simpler than the spontaneous interview response, using Robinson's (2001, 2005, 2007) classification of task complexity. To explore rater familiarity effects, listener judgments were analyzed for two groups of raters, those who either did or did not have (basic) knowledge of the speakers' L1 (Spanish). The results overall revealed little influence of raters' knowledge of the speakers' L1 but demonstrated strong effects of task on the linguistic dimensions of accentedness and comprehensibility in that the distinction between accentedness and comprehensibility was clearest in the picture narrative but was blurred in the more complex interview response.

Task Effects on L2 Accentedness and Comprehensibility

The current study is among the first investigations conducted on L2 French focusing on linguistic variables underlying listeners' judgments of accentedness and comprehensibility. In the picture narrative task, regardless of raters' amount of familiarity with speakers' L1, accentedness was easily distinguishable from comprehensibility in that accentedness was uniquely associated with the phonology and fluency dimension of L2 speech while comprehensibility was broadly linked to multiple dimensions, spanning the categories of phonology, fluency, lexis, grammar, and discourse. This finding replicates prior research on L2 English using a similar picture description task in which accentedness was found to be more narrowly linked to phonology and fluency but comprehensibility was additionally associated with lexis, grammar, and discourse (Saito et al., 2015, 2016; Trofimovich & Isaacs, 2012; Varonis & Gass,

1982). However, the current study revealed a novel finding: The linguistic distinction between accentedness and comprehensibility disappears in an interview task, again regardless of listeners' knowledge of speakers' L1.

At least one reason for the obtained pattern of findings is that the relationship between accentedness and comprehensibility and the linguistic dimensions underlying each construct is task specific. That is, the demands of a given speaking task likely determine the degree to which various linguistic resources matter for listeners' perception of accentedness (nativelikeness) and comprehensibility (ease of understanding). In terms of comprehensibility, the current findings emphasize once again that the ease with which listeners extract meaning from L2 utterances is a matter of multiple linguistic resources, regardless of speaking task (Saito et al., 2015, 2016; Trofimovich & Isaacs, 2012; Varonis & Gass, 1982). This point about the ease of understanding being linked to more than just phonology (pronunciation) is worth emphasizing again to both L2 learners and their instructors, as it is often erroneously believed that only fluency and error-free pronunciation accuracy matter for comprehensible L2 speech.

However, what makes L2 speakers sound accented appears to depend on the demands of a speaking task. In a picture-description task where speakers' linguistic output was clearly predictable for listeners who also had access to the images prior to rating, both in terms of the expected lexical expression and to some degree grammatical structure, listeners' perception of accent was largely a matter of phonology and fluency. This task was less complex for speakers—who had access to planning time, followed a predetermined story structure, and made reference to depicted objects, relationships, and actions; it was also less complex for listeners—who had clear expectations of the linguistic content. Thus, what listeners chiefly considered as

nativelike was based on the form of speakers' utterances (phonology, fluency). In contrast, in a free-response situation such as an interview, where speakers did not have much support (e.g., in the form of images) to structure their response and needed to talk beyond the here-and-now, deploying multiple linguistic resources, thus making their speech less predictable for listeners, assessments of accentedness were associated with multiple linguistic considerations.

In essence, complex speaking tasks provide speakers with more linguistic freedom to make themselves understood using multiple linguistic resources but these tasks also make it more likely that issues of lexis and grammar, in addition to being relevant to comprehensibility, also contribute to listeners' assessments of L2 speakers' accent. Put differently, increased task demands require speakers—and to a certain degree also listeners—to use all the linguistic resources in their possession to express meaning and be understood, which makes it harder to distinguish between comprehensibility and accentedness because multiple linguistic categories of L2 speech become relevant to each construct. Based on O'Brien's (2014) research with nonnative listeners, task complexity clearly needs to be defined in reference to both speakers' and listeners' demands: A picture description task that was considered simple for native-speaking listeners here might have been more complex to evaluate for nonnative listeners in O'Brien's study, which would explain why L2 listeners in that study associated accentedness and comprehensibility with overlapping linguistic dimensions.

Listener Experience Effects on Linguistic Dimensions of Accent and Comprehensibility

Despite prior research demonstrating that listeners' experience, which is typically defined in terms of their exposure to L2 speech or training in language-related disciplines, matters for listener-based assessments of speech constructs (Gass & Varonis, 1984;

Isaacs & Thomson, 2013; Kennedy & Trofimovich, 2008; Saito et al., 2015; Winke et al., 2012), there was no clear evidence in this study of listener effects in terms of their familiarity with speakers' L1. In fact, listeners with knowledge of Spanish and those reporting no knowledge of Spanish were similar in their global L2 ratings and in the linguistic dimensions they associated with these ratings. Of course, one possibility is that the listeners with knowledge of Spanish were unclear or insufficiently familiar as to how various features of Spanish could make the L2 French speech of Hispanic speakers more or less accented or comprehensible, which made them behave like listeners with no prior knowledge of Spanish. A more plausible explanation, however, is that all listeners in this study were highly familiar with the speech of Hispanic speakers of L2 French. In fact, as shown in Table 1, both listener groups were highly aware of not only accented French but, most important, Spanish-accented French, with the group that reported no experience with Spanish estimating their familiarity with Spanish-accented French to be higher (at a mean of 6.0) than the group with knowledge of Spanish (5.4). Indeed, a familiarity with Spanish-accented French in Montreal may stem from a large proportion of nonnative speakers being Hispanic and may also be typical of raters studying and working in the area of L2 pedagogy and education.

This finding brings up an intriguing possibility that naturalistic exposure to accented L2 speech may compensate for listeners' lack of formal language learning experience in the speaker's L1. Put simply, what makes listeners experienced may in large part be related to their exposure to accented L2 speech (e.g., Gass & Varonis, 1984; Kennedy & Trofimovich, 2008), not necessarily to their experience having learned the L1 of the speaker (e.g., Carey et al., 2011; Winke et al., 2012). Such an explanation would be consistent with both prior literature showing that listeners rapidly adapt to and learn to cope with

unfamiliar L2 accents (e.g., Bradlow & Bent, 2008; Clarke, 2004) and that sharing an L1 background with speakers does not automatically confer advantages to listeners (e.g., Major et al., 2002; Munro et al., 2006). Put differently, the "experience" in experienced listeners could to a large degree be a matter of exposure to nonnative speech, rather than training in linguistics and phonetics or knowledge of specific languages. This interpretation of experience should be targeted in future research in reference to multiple variables as part of the complete portrait of the listener, including listener biases, stereotypes, social attitudes, and strength of ethnic and cultural beliefs (e.g., Brennan & Brennan, 1981; Gluszek & Dovidio, 2010).

Implications and Conclusions

This study offers evidence to support previous research targeting L2 accentedness and comprehensibility, extending the results of prior research to languages other than L2 English. This is an important step as it allows for a discussion of such broad constructs as L2 accentedness and comprehensibility with reference to a range of possible L2s, which is different from initiatives targeting "criterial features" of learner language—for example, in terms of grammatical, functional, and lexical features—in a single L2—notably, English (e.g., Hawkins & Filipović, 2012). From a theoretical viewpoint, the findings point to a conclusion that the relationship between accentedness and comprehensibility and the linguistic dimensions contributing to each construct is likely a dynamic one, shaped by the demands of a given speaking task and its impact on both the speaker and the listener.

In terms of practical considerations, the current findings offer good news for language learners. When it comes to using an L2 in complex speaking tasks, there is little distinction in terms of which dimensions of L2 speech feed into listeners' perceptions of accentedness and comprehensibility. In other words, whether their goal is accent-free L2 speech or comprehensible L2 output, learners will likely benefit equally from a focus on various aspects of L2 speech (targeting various aspects of phonology, fluency, lexis, grammar, and discourse) during complex speaking tasks. For language instructors, the current results imply that free-response tasks, as opposed to more structured activities, may encourage learners to draw on multiple linguistic resources and as a result may enable learners to practice various aspects of L2 pronunciation, lexicon, grammar, and discourse while also allowing them to simultaneously work on both their accent and comprehensibility. For instance, K–12, high school, and postsecondary students—depending on their proficiency level—may benefit from interactions performed in a range of tasks, from less to more complex ones. With a focus on keeping learners motivated (van Lier, 2014), simpler tasks could be structured around visual materials (digital or paper-based) including photos and comic strips, which learners could share with each other, narrating real or imagined stories, as part of individual or group activities. In more complex tasks, learners may work on multiple linguistic dimensions within the context of real-life task scenarios, such as debating the best movie of the year or participating in job interviews. Performing such tasks in dyads would also enable learners to see whether they are successful in their interaction (see Kaur, 2009, for various signs of communication breakdowns), which may provide a natural, conversation-driven stimulus to repair speech. Learners may also benefit from discussions of what makes L2 speech more or less comprehensible and accented, with the idea of raising learners' awareness of which linguistic dimensions of speech may matter more or less for successful performance in various speaking tasks.

With the view of future research, it would be important to investigate how nonlinguistic, affective, and social information that is available to interlocutors in face-to-face interaction—such as gestures, body

language, speech style, and attitude—may impact interlocutors' perceptions of accentedness and comprehensibility in L2 speech. Indeed, the monologic nature of L2 performances that were investigated in the current study (evaluated by listeners through audio-recordings of speech, not through video or authentic interaction) may not be fully representative of speakers' dialogic experiences in real-life communication. Thus, it remains crucial that L2 researchers pursue further investigations of accentedness and comprehensibility in dialogic performance, with the view of identifying the linguistic underpinnings of each construct. It is equally important that researchers extend their current work to L2s other than English so that current research findings can be both tested with and ultimately applied to speakers of multiple L2s.

References

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching*, *44*, 231–250.

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

Brennan, E. M., & Brennan, J. S. (1981). Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research*, *10*, 487–501.

Canfield, D. L. (1981). *Spanish pronunciation in the Americas*. Chicago: The University of Chicago Press.

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

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

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2014). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, *36*, 570–590.

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

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

Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, *29*, 359–380.

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

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.

Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 287–313). Ann Arbor: The University of Michigan Press.

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.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of non-native speech. *Language Learning*, *34*, 65–87.

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of non-native accents in communication. *Personality and Social Psychology Review*, *14*, 214–237.

Hawkins, J. A., & Filipović, L. (2012). *Criterial features in L2 English: Specifying the reference levels of the common European framework*. Cambridge, UK: Cambridge University Press.

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics*, *36*, 664–679.

Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. London: Sage.

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

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Jiang, M., Green, R. J., Henley, T. B., & Masten, W. G. (2009). Acculturation in relation to the acquisition of a second language. *Journal of Multilingual and Multicultural Development*, *30*, 481–492.

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). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal*, *94*, 554–566.

Kang, O., Vo, S. C. T., & Moran, M. K. (2016). Perceptual judgments of accented speech by listeners from different first language backgrounds. *The Electronic Journal for English as a Second Language*, *20*, 1–24.

Kaur, J. (2009). Pre-empting problems of understanding in English as a lingua franca. In A. Mauranen & E. Ranta (Eds.), *English as a lingua franca: Studies and findings* (pp. 107–125). Newcastle-upon-Tyne, UK: Cambridge Scholars.

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

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

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

Long, M. H., & Crookes, G. (1992). Three approaches to task-based syllabus design. *TESOL Quarterly*, *26*, 27–56.

Major, R. C., Fitzmaurice, S. F., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for ESL assessment. *TESOL Quarterly*, *36*, 173–190.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching*, *45*, 241–259.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *49*, 285–310.

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.

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

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

O'Brien, M. G. (2016). Methodological choices in rating speech samples. *Studies in Second Language Acquisition*, *38*, 587–605.

Plonsky, L. D., & Kim, Y. (2016). Task-based learner production. *Annual Review of Applied Linguistics*, *36*, 73–97.

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

Rau, D., Chang, H. H. A., & Tarone, E. E. (2009). Think or sink: Chinese learners' acquisition of the English voiceless interdental fricative. *Language Learning*, *59*, 581–621.

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *Modern Language Journal*, *95*, 162–181.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*, 27–57.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics in Language Teaching*, *43*, 1–32.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International*

Review of Applied Linguistics in Language Teaching, 45, 193–213.

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., Trofimovich, P., & Isaacs, T. (2015). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, 36, 1–25.

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.

Schmid, P. M., & Yeni-Komshian, G. H. (1999). The effects of speaker accent and target predictability on perception of mispronunciations. *Journal of Speech, Language, and Hearing Research*, 42, 56–64.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30, 510–532.

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

Trofimovich, P., Kennedy, S., & Blanchet, J. (in press). Development of L2 French oral skills in an instructed setting: A focus on speech ratings. *Canadian Journal of Applied Linguistics*.

Varonis, E. M., & Gass, S. (1982). The comprehensibility of nonnative speech. *Studies in Second Language Acquisition*, 4, 114–136.

van Lier, L. (2014). *Interaction in the language curriculum: Awareness, autonomy and authenticity*. New York: Routledge.

White, J., & Turner, C. (2005). Comparing children's oral ability in two ESL programs. *Canadian Modern Language Review*, 61, 491–517.

Winke, P., Gass, S., & Myford, C. (2012). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, 30(2), 231–252.

Xi, X., & Mollaun, P. (2009). *How do raters from India perform in scoring the TOEFL iBT speaking section and what kind of training helps*? TOEFL iBT Research Report RR-09–31. Princeton, NJ: Educational Testing Service.