To be published in *Studies in Second Language Acquisition* (Cambridge University Press)

Linguistic Dimensions of Second Language Accentedness and Comprehensibility Vary across Speaking Tasks

Dustin Crowther
Pavel Trofimovich
Talia Isaacs
Kazuya Saito

Abstract

This study critically examined the previously reported partial independence between second language (L2) accentedness (degree to which L2 speech differs from the target variety) and comprehensibility (ease of understanding). In prior work, comprehensibility was linked to multiple linguistic dimensions of L2 speech (phonology, fluency, lexis, grammar) whereas accentedness was narrowly associated with L2 phonology. However, these findings stemmed from a single task (picture narrative), suggesting that task type could affect the particular linguistic measures distinguishing comprehensibility from accentedness. To address this limitation, speech ratings of 10 native listeners assessing 60 speakers of L2 English in three tasks (picture narrative, IELTS, TOEFL) were analyzed, targeting two global ratings (accentedness, comprehensibility) and 10 linguistic measures (segmental and word stress accuracy, intonation, rhythm, speech rate, grammatical accuracy and complexity, lexical richness and complexity, discourse richness). Linguistic distinctions between accentedness and comprehensibility were less pronounced in the cognitively complex task (TOEFL), with overlapping sets of phonology, lexis, and grammar variables contributing to listener ratings of accentedness and comprehensibility. This finding points to multifaceted, task-specific relationships between these two constructs.

Abstract

This study critically examined the previously reported partial independence between second language (L2) accentedness (degree to which L2 speech differs from the target variety) and comprehensibility (ease of understanding). In prior work, comprehensibility was linked to multiple linguistic dimensions of L2 speech (phonology, fluency, lexis, grammar) whereas accentedness was narrowly associated with L2 phonology. However, these findings stemmed from a single task (picture narrative), suggesting that task type could affect the particular linguistic measures distinguishing comprehensibility from accentedness. To address this limitation, speech ratings of 10 native listeners assessing 60 speakers of L2 English in three tasks (picture narrative, IELTS, TOEFL) were analyzed, targeting two global ratings (accentedness, comprehensibility) and 10 linguistic measures (segmental and word stress accuracy, intonation, rhythm, speech rate, grammatical accuracy and complexity, lexical richness and complexity, discourse richness). Linguistic distinctions between accentedness and comprehensibility were less pronounced in the cognitively complex task (TOEFL), with overlapping sets of phonology, lexis, and grammar variables contributing to listener ratings of accentedness and comprehensibility. This finding points to multifaceted, task-specific relationships between these two constructs.

Linguistic dimensions of L2 accentedness and comprehensibility vary across speaking tasks

Research-informed approaches towards second language (L2) pronunciation development in classroom and research contexts can be described according to two competing perspectives, the nativeness and intelligibility principles (Levis, 2005). The nativeness principle targets nonaccented L2 speech, or speech free from linguistic features that might mark the speaker as nonnative. By contrast, the intelligibility principle focuses on L2 speech that is understandable to an interlocutor, despite the presence of a detectable accent. Because accented speech in adults is viewed as normal and unavoidable and the attainment of nativelike speech is uncommon (e.g., Abrahamsson & Hyltenstam, 2009), there is a strong scholarly emphasis on the promotion of intelligible speech as the primary goal of L2 pronunciation development (e.g., Derwing & Munro, 2015; Levis, 2005).

In support of this goal, a number of studies have examined which linguistic dimensions of L2 speech are associated with accentedness versus comprehensibility—key constructs in L2 speech research (Derwing & Munro, 2015). Accentedness, aligned with the nativeness principle, captures listeners' perception of how strongly L2 speech is influenced by the speaker's native language or is colored by other nonnative features. Comprehensibility is aligned with the intelligibility principle, as this construct encompasses listeners' perception of ease or difficulty of understanding L2 speech. Although comprehensibility is not a measure of listeners' actual understanding of L2 speech (i.e., intelligibility), typically operationalized through listeners' orthographic transcriptions or retells of a speaker's utterance, comprehensibility offers an appropriate measure of understanding in a broad sense, particularly in real-life contexts (Isaacs & Trofimovich, 2012). For example, many rating scales for standardized proficiency tests use the term "intelligibility" (e.g., TOEFL, IELTS) when what is, in fact, being measured is

comprehensibility (Harding, 2017). Furthermore, for many speakers, it is their subjective perceptions of how easy or difficult it is to process linguistic input, more so than actual performance measures, that predict various cognitive and linguistic behaviors (Oppenheimer, 2008). Comprehensibility is therefore a construct common to many rating scales and is also reflective of people's general experience with speech.

Accentedness and comprehensibility appear to be partially independent in two key respects. First, L2 speakers can possess heavy accents and still be comprehensible (Derwing & Munro, 2015). Second, the linguistic measures of L2 speech that feed into comprehensibility are more diverse than those that underlie accentedness (Varonis & Gass, 1982), such that accentedness is primarily associated with segmental accuracy while comprehensibility is additionally linked to suprasegmental (e.g., word stress, intonation, rhythm), fluency, and lexis/grammar considerations (e.g., Crowther, Trofimovich, Isaacs, & Saito, 2015a; Isaacs & Trofimovich, 2012; Saito, Webb, Trofimovich, & Isaacs, 2016). However, most existing evidence for the linguistic independence of accentedness and comprehensibility, the focus of this study, comes from research utilizing a single task (picture narrative). Because L2 speakers draw on different linguistic resources (pronunciation, fluency, vocabulary, grammar) to complete different tasks (Robinson, 2005; Skehan, 2009), the linguistic underpinnings of accentedness and comprehensibility might crucially depend on the demands of the speaking task. The goal of this report was to investigate this possibility.

Linguistic Independence of Accentedness and Comprehensibility: A Possible Task Effect?

In early research on linguistic correlates of accentedness and comprehensibility, many measures of L2 speech were considered in isolation, often across separate publications (see Crowther, Trofimovich, Saito, & Isaacs, 2015b). To address this limitation, recent studies

examined the combined contribution of 19 linguistic measures to L2 accentedness and comprehensibility (Isaacs & Trofimovich, 2012; Trofimovich & Isaacs, 2012), and then expanded this work to investigate the roles of speakers' first language (L1) background (Crowther et al., 2015b), listeners' L1 status (Crowther, Trofimovich, & Isaacs, 2016), and rating scale type (Saito, Trofimovich, & Isaacs, 2015) in determining which linguistic dimensions of L2 speech pattern with accentedness and which with comprehensibility. Across these studies, the wider range of linguistic dimensions linked to comprehensibility—compared to accentedness, which is associated predominantly with segmental accuracy—was taken as support for the partial linguistic independence of the two constructs.

The most common task used to elicit L2 speech for analysis of accentedness and comprehensibility has been a picture narrative, where speakers describe a series of sequenced pictures. However, this methodological choice means that nearly all evidence for the linguistic independence of accentedness from comprehensibility comes from one task. In fact, to our knowledge, there is no research comparing linguistic dimensions of both accentedness and comprehensibility, within a single report, across different task types. Yet previous task-based research focusing on L2 oral production suggests that task effects should not be ignored. All theoretical frameworks that center on the effects of task on L2 learners' linguistic performance indicate that differences in task demands impact L2 spoken output in terms of segmental and prosodic content (Tarone, 1983) and lexical and grammatical features (Robinson, 2005; Skehan, 2009). With respect to pronunciation accuracy, for example, segmental accuracy is greater in read-aloud than in spontaneous tasks (Rau, Chang, & Tarone, 2009). There also appears to be a hierarchy of perceived fluency for listeners, with L2 output in dialogue-based tasks rated as more fluent than speech elicited through picture narratives (Derwing, Rossiter, Munro, & Thomson,

2004). In terms of lexis and grammar, differences across tasks in planning and completion time (Yuan & Ellis, 2003), completion objectives (Robinson, 2005), and topic familiarity (Foster & Skehan, 1996) elicit varying levels of lexical and grammatical accuracy and complexity. Clearly, a consideration of task demands is crucial if viable claims are to be made about which linguistic dimensions of L2 speech are relevant to accentedness and which pattern with comprehensibility.

The Current Study

While previous research has focused on which linguistic dimensions of speech promote comprehensible L2 speech over nativelike performance, the bulk of evidence for linguistic correlates of comprehensibility is based on a limited task repertoire. If increased demands of a speaking task place greater strain on speakers' production processes, encouraging them to resort to all available linguistic resources (Robinson, 2005; Skehan, 2009), it becomes necessary to understand if the linguistic measures that differentiate comprehensibility from accentedness also vary across task type. It could be that the linguistic distinction between comprehensibility and accentedness (in terms of the linguistic dimensions that pattern with each construct), which has been robust in previous work with picture narratives, might attenuate or disappear if L2 speakers are tested in other tasks, especially those that vary in difficulty. Put differently, complex speaking tasks, requiring speakers to deploy all available linguistic resources, might lead to multiple, overlapping linguistic dimensions feeding into listeners' perceptions of L2 accentedness and comprehensibility, reducing the extent to which the linguistic signatures of accentedness and comprehensibility are distinct. One reason for this might be that greater task demands would require speakers to draw upon a wider range of linguistic dimensions, which may reduce the extent to which these dimensions differentiate accentedness from comprehensibility. If this were to be the case, beyond an empirically based understanding of

what contributes to listeners' perceptions of speech, alignment between the two constructs during more demanding task performance may impact both pedagogical and assessment considerations.

Therefore, for this report, we revisited the large-scale dataset previously featured in two conceptually distinct publications (Crowther et al., 2015a, 2015b). In Crowther et al. (2015b), the focus was on the role of speakers' L1 in determining which linguistic dimensions are relevant to listeners' perception of L2 accentedness versus comprehensibility in a picture narrative. In Crowther et al. (2015a), the focus was on task differences (using two non-picture narrative tasks), but only in how they related to comprehensibility (and associated linguistic dimensions), which is in line with a scholarly emphasis on intelligibility over nativeness (e.g., Derwing & Munro, 2015; Levis, 2005). In essence, neither report examined how task impacted *both* accentedness and comprehensibility, specifically in regards to the linguistic dimensions associated with each. Subsequent analyses of the dataset in full revealed more nuanced findings in this regard, findings necessary to refine our understanding. Therefore, the goal of this report was to extend previous studies by (a) analyzing the data across three speaking tasks which were previously targeted in two separate publications and (b) comparing the data for accentedness and comprehensibility across three speaking tasks, which has not been done previously. As in prior work, the focus here was on 10 rated linguistic dimensions of L2 speech—namely, segmental accuracy, word stress accuracy, intonation, rhythm, speech rate, grammatical accuracy and complexity, lexical richness and complexity, and discourse richness—to determine whether the linguistic correlates of comprehensibility differ from those associated with accentednes across tasks. By consolidating all analyzed data for a single report, it was possible to address the following question, which previous studies were unable to examine: *Can L2 accentedness and comprehensibility be differentiated at the level of linguistic dimensions across speech elicitation*

tasks varying in task demands?

Method

Participants

Participants included 60 L2 speakers of English (22 female, 38 male) who provided audio recordings of speech and 10 native English-speaking listeners (8 female, 2 male) who evaluated the recordings. The 60 speakers were drawn from an unpublished corpus of 143 L2 university students representing 19 linguistic backgrounds. All speakers were in the first semester of undergraduate (n = 29) or graduate (n = 31) study at an English medium university in Montreal, Canada, and represented four distinct L1 backgrounds: Farsi, Hindi/Urdu, Mandarin, and Romance. Because the goal of this report was to determine relationships between linguistic dimensions of L2 speech and listeners' ratings of comprehensibility and accentedness for speakers from multiple L1s, the 60 speakers were considered as a single group (for L1 differences, see Crowther et al., 2015b). Speakers' biographical information is summarized in Table 1. The listeners included 10 native speakers of English (M_{age} = 32.7 years, *range* = 25–56), raised in English-speaking homes with at least one native English-speaking parent, all current or recent graduates in applied linguistics. The listeners reported using English 89% of the time daily (*range* = 80–100%). All listeners indicated high familiarity with accented English and had on average 6.6 years of language teaching experience. Experienced listeners were chosen over naïve listeners as they tend to demonstrate more consistency in their linguistic ratings (Saito et al., 2015).

TABLE 1

Speaking Tasks

Each speaker completed three tasks. The first was a picture narrative (hereafter, picture

task, available at http://www.iris-database.org), used in Derwing, Munro, Thomson, & Rossiter (2009) and other studies by these authors. Speakers described an eight-frame picture narrative, in which two strangers bumped into each other, accidentally exchanged their identical suitcases, and realized their mistake upon returning home. The second task was the IELTS long-turn task (hereafter, the IELTS task), in which speakers received a card with one of two assigned topics (*describe a sports event you enjoyed watching*, *describe a job you would like to do in the future*) and two suggested discussion points (IELTS, 2009). Following IELTS procedures, after 1 minute of preparation time, speakers were given 1–2 minutes to respond, with 1–2 minute optional follow-up questions by the interviewer. Though the interviewer did not have IELTS examiner qualifications, they followed the IELTS interviewing procedures consistently across all individual interviews. For the third task, speakers completed a TOEFL iBT integrated task (hereafter, the TOEFL task), with stimuli presented via a computer interface (Educational Testing Service, 2006). Speakers were allotted 45 seconds to read a 93–105 word passage and then listened to a 80–90 second audio lecture on a related topic. Drawing from both, speakers responded to a question related to the content of both the listening and reading stimuli. They had 30 seconds to prepare their response before speaking for up to 1 minute. Speakers responded to one of two topics (*audience effects in psychology*, *behavioral explanations in sociology*), with approximately half of the speakers assigned to each. Independent-samples *t* tests indicated no differences between the two IELTS and TOEFL task versions ($p > .05$), so data across task versions were combined.

To determine task type differences, the three task types were analyzed by the first two authors following Robinson's (2005) framework for task classification, with respect to resource-directing variables (see Table 2). The TOEFL task appeared to draw on different cognitive

resources, compared to the picture and IELTS tasks, specifically requiring reasoning and perspective taking. Neither was necessary to complete the picture and IELTS tasks. However, the picture and the IELTS tasks differed in the degree of topic familiarity (and associated linguistic freedom) they afforded the speaker. Whereas speakers needed to draw on their linguistic resources to address a familiar and personal topic in the IELTS task, the picture task constrained the range of lexical items required for completing the narrative in a given sequence without much choice in content and organization (see Foster & Skehan, 1996). For a more in-depth description of how task complexity was determined, see Crowther et al. (2015a).

TABLE 2

To further assess the hierarchy of task differences, participants were asked to rate the difficulty of each task on a 9-point Likert Scale (1 = *very easy*, 9 = *very hard*). While no significant difference was found between the picture and IELTS tasks (M_{diff} = .07, p = .82, 95% CI [–0.56, 0.71]), the TOEFL task was perceived as more difficult than both the picture (M_{diff} = .71, p = .009, 95% CI [–1.24, –0.19]) and IELTS (M_{diff} = .80, p = .005, 95% CI [–1.34, –0.26]) tasks. Speakers' perception of the TOEFL task as the most difficult aligned with our view that it would be the most cognitively challenging, similar to the findings in Révész, Michel, and Gilabert (2015), who studied learner versus expert judgments of task complexity. Thus, the three tasks likely represented different task demands, with the picture and IELTS tasks being less demanding than the TOEFL task.

Rating Procedure

All audio recordings were edited to include the initial 30 seconds of speech, minus initial fillers and disfluencies, and were normalized for peak amplitude. The recordings, transcribed and verified by two transcribers, served as the materials for listener-based assessments. The 10

listeners evaluated the 60 speakers in all tasks using 1,000-point scales in Z-Lab (Yao, Saito, Trofimovich, & Isaacs, 2013), a custom-designed MATLAB program. Listeners first familiarized themselves with the task materials (e.g., images from the picture task), then perceptually evaluated 12 categories in total, including two global measures (accentedness, comprehensibility) and 10 specific linguistic variables, described in Table 3. To rate the two global dimensions and five pronunciation and fluency variables, listeners used audio recordings. To rate the five remaining variables, they used transcripts of each audio file, which minimized the influence of pronunciation and fluency variables on ratings of lexis, grammar, and discourse (Crossley, Salsbury, & McNamara, 2015). Each of the three sets of measures were rated simultaneously within their associated sets (global; pronunciation & fluency; lexis, grammar, & discourse) because there are few differences between measures rated simultaneously or consecutively (O'Brien, 2016). Listeners were trained by the first author on all measures (using three practice recordings) prior to evaluating each speech sample. Listener ratings were obtained in four individual two-hour sessions within a three-week span. Samples were blocked and counterbalanced by task and presented in unique randomizations. Subjective measures were chosen, as Saito et al (2015)—using the dataset featured in Isaacs and Trofimovich (2012)—found that listener ratings of these 10 linguistic measures aligned closely with the measures derived through coding by trained coders in the original study. Further details about the rating categories and assessment procedure can be found in Crowther et al. (2015a, 2015b).

TABLE 3

Reliability

Listeners were consistent in their judgements of all rated measures across tasks (Table 4), exceeding the benchmark value of Cronbach's $\alpha > .70–.80$ (Larson-Hall, 2009). Therefore,

single mean scores per speaker were derived by averaging across all listener ratings for each rated measure.

TABLE 4

Results

Global Ratings

The first analysis considered the relationship between accentedness and comprehensibility (Table 5). The two constructs were strongly correlated in the picture ($r = .80$), IELTS ($r = .79$), and TOEFL ($r = .74$) tasks. A two-way repeated measures ANOVA with global rating (accentedness, comprehensibility) and task (picture, IELTS, TOEFL) as within-groups factors yielded a significant main effect of global rating, $F(1, 59) = 231.02$, $p < .001$, $\eta_\rho^2 = .80$, and task, $F(2, 59) = 3.21$, $p = .044$, $\eta_\rho^2 = .05$, along with a significant interaction, $F(1, 59) = 31.46$, $p < .001$, $\eta_\rho^2 = .35$.

TABLE 5

Post hoc (Bonferroni adjusted) analyses revealed that comprehensibility was rated significantly higher than accentedness in all tasks: picture ($M_{diff} = 112.45$, $p < .001$, $d = 0.69$, 95% CI [85.15, 139.75]), IELTS ($M_{diff} = 227.58$, $p < .001$, $d = 1.46$, 95% CI [200.93, 254.24]), and TOEFL ($M_{diff} = 175.73$, $p < .001$, $d = 1.09$, 95% CI [145.28, 206.18]). Additionally, speakers were perceived to be less accented in the picture, compared to the IELTS ($M_{diff} = 31.50$, $p = .032$, $d = 0.19$, 95% CI [2.04, 60.96]) or TOEFL ($M_{diff} = 30.43$, $p = .017$, $d = 0.18$, 95% CI [4.28, 56.58]) tasks, with no difference between the IELTS and TOEFL tasks ($M_{diff} = 1.07$, $p = 1.00$, $d = 0.01$, 95% CI [–30.93, 28.79]). For comprehensibility, speakers were rated higher in the IELTS, compared to either the picture ($M_{diff} = 83.63$, $p < .001$, $d = 0.55$, 95% CI [45.01, 122.26]) or TOEFL ($M_{diff} = 50.78$, $p = .008$, $d = 0.33$, 95% CI [10.92, 90.65]) tasks, with no difference

between the picture and TOEFL tasks (M_{diff} = 32.85, p = .087, d = 0.21, 95% CI [–3.32, 69.02]).

Pronunciation and Lexicogrammar as Predictors of Global Ratings

The next analysis targeted the relationship between accentedness and comprehensibility and the 10 rated linguistic measures. For each task, an exploratory principal component analysis (PCA) with Oblimin rotation was conducted to determine if the 10 linguistic measures showed any patterns based on their clustering. Kaiser-Meyer-Olkin values (picture = .85; IELTS = .91; TOEFL = .86) and Bartlett's tests of sphericity (picture: $\chi^2(60) = 692.09$, $p < .001$; IELTS: $\chi^2(60) = 822.95$, $p < .001$; TOEFL: $\chi^2(60) = 700.57$, $p < .001$) suggested excellent factorability of the correlation matrix, despite the relatively small sample size (N = 60). As shown in Table 6, the PCA yielded a clear two-factor solution for each task, accounting for a substantial amount of variance per task (picture = 82%; IELTS = 82%; TOEFL = 87%). In all tasks, Factor 1 was labeled *pronunciation* and Factor 2 *lexicogrammar*. Aside from speech rate loading on both factors in the picture task, the factors were identical and distinct across tasks.

TABLE 6

Pearson correlations, which were computed between accentedness and comprehensibility and the two PCA factor scores (pronunciation, lexicogrammar), derived through the Anderson-Rubin method of obtaining noncorrelated factor estimates (see Table 7), revealed that comprehensibility was strongly associated with both factors in all tasks, featuring strong associations (r > .60), following Plonsky and Oswald's (2014) guidelines. Accentedness was strongly correlated with the pronunciation factor across tasks. While the relationships between accentedness and lexicogrammar were overall weaker, they increased in strength from the picture and IELTS tasks, where associations were weak (r > .25), to the TOEFL task, where the association was medium in strength (r > .40).

TABLE 7

To investigate the relative contribution of the pronunciation and lexicogrammar factors to explaining the variance in accentedness and comprehensibility ratings across tasks, hierarchical multiple regression analyses were carried out, separately per task, with accentedness or comprehensibility as criterion variables. Considering the strength of factor associations (see Table 7), pronunciation was entered as a predictor first, followed by lexicogrammar, which allowed for determining if lexicogrammar could explain any additional variance in accentedness or comprehensibility beyond that accounted for by the pronunciation factor. As shown in Table 8, both pronunciation and lexicogrammar were significant predictors of comprehensibility in the picture (78% of total variance explained), IELTS (74%), and TOEFL (87%) tasks. However, only pronunciation predicted accentedness scores in the picture (77%), IELTS (74%), and TOEFL (76%) tasks.

TABLE 8

Individual Linguistic Measures and Global Ratings

Though the lexicogrammar factor was not found to be a significant predictor of accentedness, as task complexity increased, there was a gradual increase in the likelihood of lexicogrammar predicting accenteness across tasks: picture (p = .988), IELTS (p = .250), and TOEFL (p = .053). Considering this trend toward lexicogrammar being more relevant to explaining accentedness ratings with increasing task difficulty, the final analysis explored associations between accentedness and comprehensibility and the full set of 10 linguistic measures. Following Plonsky and Oswald's (2014) guidelines, a series of Pearson correlations (r) were conducted (Table 9).

TABLE 9

Individual correlations confirmed that comprehensibility in all tasks was strongly linked to all pronunciation and lexicogrammar measures, with strong associations ($r > .60$), and that accentedness across tasks was associated with the five pronunciation measures ($r > .40$). However, there was a gradual increase in the relevance of lexicogrammar measures to accentedness. Whereas only two lexicogrammar measures (lexical appropriateness, grammatical accuracy) featured associations with accentedness in the picture task (small-to-medium association strength), all five lexicogrammar measures were linked to accentedness in the IELTS and TOEFL tasks. These latter two tasks differed in the strength of these correlations. In the IELTS task, four associations were weak ($r < .40$). However, in the TOEFL task, two of the five (lexical appropriateness, grammatical accuracy) featured medium-strength associations.

Discussion

The goal of this report was to investigate whether the partial independence between accentedness and comprehensibility—established in prior research in reference to linguistic measures of L2 speech associated with each construct (e.g., Trofimovich & Isaacs, 2012)—would hold when analyzed across speaking task. Three speech production tasks (picture, IELTS, TOEFL) were considered, with task complexity determined through an analysis of resource-directing variables (Robinson, 2005). Comprehensibility across all tasks was strongly associated with measures of L2 phonology and lexis/grammar, suggesting (in line with prior research) that listeners draw on multiple linguistic dimensions of L2 speech when evaluating comprehensibility (e.g., Crowther et al., 2015a; Isaacs & Trofimovich, 2012). While linguistic correlates of comprehensibility remained consistent across task, linguistic correlates of accentedness did not. For the picture task, listener perception of accentedness was associated almost exclusively with phonology measures, replicating prior findings for picture narratives, including for participants

from a different dataset (e.g., Trofimovich & Isaacs, 2012). With increased task complexity, however, associations between accentedness and measures of L2 lexis/grammar strengthened (see Table 9), implying that aspects of lexis and grammar could serve as predictors of accentedness. Linguistic distinctions between accentedness and comprehensibility were thus clearest in the picture task but were most blurred in the more complex TOEFL task. These findings extend, but do not contradict, prior results: Because previous reports either did not include a focus on both accentedness and comprehensibility (Crowther et al., 2015a) or targeted only one task (Crowther et al., 2015b), a task-specific pattern was not established.

As predicted by task complexity frameworks, increased task difficulty likely elicits more elaborate language as L2 speakers strive to meet the increased demands of the task (Robinson, 2005). At least for the TOEFL task, one reason for these greater demands is the greater flexibility provided in how to respond. In the picture task, speakers had a predetermined structure to follow (describe pictures in order), and in the IELTS task, they dealt with (usually) a familiar topic. They thus likely felt little pressure to engage all available linguistic resources, apart from the need to retrieve and use vocabulary drawn from a restricted lexical set (pictures objects, actions) or a familiar subject (future job, favorite sports), which might have been their primary source of difficulty (Hilton, 2008). By contrast, the TOEFL task offered more flexibility in how speakers chose to respond. With this flexibility of having multiple ways to make themselves understood—for example, when integrating the content of the listening and reading passages—speakers likely relied on all linguistic resources in their possession, making it likely that issues of lexis and grammar (in addition to being relevant to comprehensibility) also contributed to accentedness. Put simply, increased task demands bring both constructs—which are already highly interrelated (r = .74–.80)—into greater alignment with respect to linguistic dimensions associated with each,

such that similar linguistic dimensions contribute to the perceptual signatures of accentedness and comprehensibility.

Whereas increased task demands ostensibly minimized distinctions between accentedness and comprehensibility (in terms of the linguistic dimensions associated with each), task complexity appeared to produce the opposite effect on each rated construct. As task difficulty increased, L2 speakers were generally rated as *more* accented (less nativelike) but also *more* comprehensible (see Table 5). However, this interpretation must be considered in light of the small effect sizes associated with differences in accentedness and comprehensibility ratings between tasks. For accentedness, despite the fact that speakers' performances were rated as being more heavily accented in the IELTS and TOEFL tasks, compared to the picture task, the actual effect was minimal (d = .01–.17). For comprehensibility, while the effect for a difference in ratings between the picture and IELTS tasks was notable (d = .55), it was still relatively small (Plonsky & Oswald, 2014). One way to interpret these general patterns is to suggest that increased task difficulty was linked to a broader range of linguistic dimensions feeding into listeners' perception of accent. Yet greater accentedness, in more complex tasks, was not associated with speech that was also harder to understand. This result likely reflects similar effects of the greater flexibility provided by more complex tasks—this time, for listeners. As part of the training procedure, listeners had the opportunity to view the picture task prompt, an approach used to minimize task familiarity effects for the few initial samples rated. Listeners thus likely developed clear expectations of what they were about to hear, such that ease or difficulty of understanding was related to the extent to which each story conformed to these expectations. In contrast, with no clear path for task completion in the IELTS and especially the TOEFL task, there were many more avenues for speakers to make themselves understood.

Paradoxically, for listeners, this may make more complex tasks easier in regards to understanding the message. Listeners could, for example, rely on multiple linguistic cues (i.e., not just a handful of lexical items called for by the task) to help them process the message. Yet with increased linguistic freedom comes greater opportunity for producing language which might diverge from what listeners would consider nativelike. Essentially, for the IELTS and TOEFL tasks, while speakers had more ways to get it *right* in terms of comprehensibility, they also had more ways to get it *wrong* in regards to how accented (nativelike) they sounded.

Conclusion

Theoretically speaking, these findings question the strength of one specific previously reported distinction between accentedness and comprehensibility, namely, with respect to the linguistic dimensions of speech associated with each construct. Because the two constructs were largely overlapping in the TOEFL task, this distinction appears to be task specific, such that the linguistic variables relevant to each construct vary with task demands. This finding implies a multifaceted relationship between linguistic correlates of accentedness and comprehensibility, one that must be situated within task differences and likely also listener expectations. In terms of pedagogical implications, the finding that linguistic distinctions between accentedness and comprehensibility are blurred in a complex task might be encouraging for both language learners and teachers. Through the use of complex tasks, L2 speakers might be able to practice a range of linguistic targets (phonology, fluency, lexis, and grammar), with most having a bearing on both accentedness and comprehensibility. From an assessment perspective, as pointed out by an anonymous reviewer, pronunciation as a specific criterion of a speaking test may be best isolated in a less complex task, where non-phonological linguistic measures are less likely to influence a rater's assessment.

Of course, the above interpretations imply that further research is necessary to strengthen our understanding of the role of task in listener perceptions of L2 speech. The findings are based on the ratings of 10 highly educated and experienced, native-speaking L1 English listeners who may not be representative of listeners that L2 speakers are likely to encounter in everyday interactions. The perceptions of listeners with less formal linguistic training, different levels of L2 speech exposure, and L1 backgrounds would help elaborate on the findings presented here. Additionally, there might have been effects of both pretask and online planning time on the accuracy and complexity of language produced across the tasks. While planning time was not controlled in the picture task (such that speakers could start narrating their story as soon as they had familiarized themselves with the images), both the IELTS and TOEFL tasks differed in the amount of time allotted before speaking and in the time pressure imposed on speakers during production. Needless to say, amount of planning time and also availability of task materials (access to prompt images or texts) need to be carefully considered in future research exploring task effects on judgments of accentedness and comprehensibility. Last but not least, while the targeted tasks were representative of both existing literature (picture narrative) and common L2 assessment tools (IELTS, TOEFL), they do not encompass the full scope of task complexity. For instance, what happens when task goals are no longer individual, but require a speaker to take into consideration the views and actions of another? If we learn L2s for communication, then explorations of linguistic correlates of accentedness and comprehensibility in authentic communication would be a logical, and necessary, direction of future work. Clearly, a consideration of task complexity has led to a more complex understanding of the linguistic measures that define L2 pronunciation.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning, 59*, 249-306.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, *36,* 570-590.

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.

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

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

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

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31*, 553-557.

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.

Educational Testing Service. (2006). *The official guide to the new TOEFL iBT*. North America: McGraw Hill.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language

performance. *Studies in Second Language Acquisition*, *18*, 299-323.

Harding, L. (2017). What do raters need in a pronunciation scale? The users' view. In T. Isaacs & P. Trofimovich (Eds.), *Second language pronunciation assessment*. Bristol, UK: Multilingual Matters.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *The Language Learning Journal, 36*, 153-166.

IELTS. (2009). *Official IELTS practice materials*. Los Angeles: IELTS International.

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

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 in pronunciation teaching. *TESOL Quarterly, 39*, 369-377.

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

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*, 237-241.

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

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

Révész, A., Michel, M., & Gilabert, R. (2015). Measuring cognitive task demands using dual-

task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition*. Published online 14 September 2015.

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.

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*. Published online 29 September 2015.

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2015). Lexical profiles of comprehensible second language speech. *Studies in Second Language Acquisition*, *38*, 677-701.

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

Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics, 4*, 142-164.

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

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

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

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics, 24*, 1-27.

Table 1

L2 Speakers' Background Characteristics

Variable	*M*	*SD*
Age	22.78	3.03
Years in Canada	0.53	0.31
Years of English study	12.25	12.45
Speaking ability (1–9)[a]	6.08	1.46
Listening ability (1–9)[a]	7.10	1.28
English use at home (0–100%)	26.83	29.05
English use at school (0–100%)	71.33	27.81
TOEFL iBT score	85.85	16.92
IELTS total score	6.64	0.55

Notes. [a]1 = *extremely poor*, 9 = *extremely fluent*.

Table 2

Summary of Task Complexity Variables for the Three Speaking Tasks

	Picture	IELTS	TOEFL
Few elements	+	+	–
Spatial reasoning	+	+	–
Here/now	+	–	–
Causal reasoning	–	–	+
Intentional reasoning	–	–	+
Perspective taking	–	–	+

Table 3

Rated Categories with Scalar Endpoint Descriptors (0-1000) and Category Summary

Rated measure	Left endpoint	Right endpoint	Category Summary
Global			
Accentedness	Heavily accented	No accent at all	How different a speaker sounds from a native English speaker
Comprehensibility	Hard to understand	Easy to understand	Ease or difficulty of raters' understanding of L2 speech
Pronunciation & fluency			
Segmental errors	Frequent	Infrequent or absent	Errors in production of individual consonants and vowels within a word
Word stress errors	Frequent	Infrequent or absent	Errors in the placement of word stress
Intonation	Unnatural	Natural	Appropriateness of pitch moves within speech, such as rising tones in yes/no questions
Rhythm	Unnatural	Natural	Difference in stress (emphasis) in content and function (grammatical) words
Speech rate	Too slow or too fast	Optimal	Speakers overall pacing and speed of utterance delivery
Lexis, grammar, & 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	Varied vocabulary	Sophistication of the vocabulary used by a speaker
Grammatical accuracy	Poor grammar accuracy	Excellent grammar accuracy	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

Table 4

Interrater Reliability Across Tasks (Cronbach's α)

Rated measure	Picture	IELTS	TOEFL
Accentedness	.93	.94	.95
Comprehensibility	.86	.91	.92
Segmentals	.92	.93	.93
Word stress	.78	.86	.84
Intonation	.78	.87	.87
Rhythm	.85	.84	.88
Speech rate	.90	.85	.91
Lexical appropriateness	.81	.84	.86
Lexical richness	.88	.85	.90
Grammatical accuracy	.80	.87	.87
Grammatical complexity	.89	.89	.90
Discourse richness	.90	.90	.90

Table 5

Means (Standard Deviations) for Accentedness and Comprehensibility Across Tasks

Task	Accentedness	Comprehensibility
Picture	473.13 (170.48)	585.58 (157.49)
IELTS	441.63 (165.95)	669.22 (145.42)
TOEFL	442.70 (165.01)	618.43 (159.79)

Note. 0–1000 scale (0 = *heavily accented*, 1000 = *no accent at all*; 0 = *hard to understand*, 1000 = *easy to understand*).

Table 6

Factor Loadings (> .40) for PCA of Linguistic Measures Across Tasks

	Pronunciation factor			Lexicogrammar factor		
Linguistic measure	Picture	IELTS	TOEFL	Picture	IELTS	TOEFL
Segmentals	.91	.93	1.01			
Word stress	.89	.89	.90			
Intonation	.94	.95	.97			
Rhythm	.89	.93	.88			
Speech rate	.46	.61	.64	.58		
Lexical appropriateness				.73	.87	.73
Lexical richness				.97	.95	1.02
Grammatical accuracy				.79	.82	.85
Grammatical complexity				.97	.91	.98
Discourse richness				.98	.95	.99

Table 7

Correlations Between Accentedness, Comprehensibility, and Two PCA Factors Across Tasks

	Pronunciation factor			Lexicogrammar factor		
Measure	Picture	IELTS	TOEFL	Picture	IELTS	TOEFL
Accentedness	.88**	.86**	.88**	.30*	.37**	.42**
Comprehensibility	.76**	.78**	.84**	.69**	.72**	.83**

Note. *$p < .05$, p** $< .005$.

Table 8
Summary of Regression Analyses for Accentedness and Comprehensibility Across Tasks

Task	Predictors	R^2	ΔR^2	*B*	95% CI	*t*	*p*
Accentedness							
Picture	Pronunciation	.77	.77	.15	.13, .17	13.91	< .001
IELTS	Pronunciation	.74	.74	.15	.13, .18	11.71	< .001
TOEFL	Pronunciation	.76	.76	.16	.13, .18	12.54	< .001
Comprehensibility							
Picture	Pronunciation	.57	.57	.09	.07, .11	9.15	< .001
	Lexicogrammar	.78	.21	.08	.06, .10	7.45	< .011
IELTS	Pronunciation	.61	.61	.08	.06, .10	7.14	< .001
	Lexicogrammar	.74	.14	.06	.04, .09	5.55	< .001
TOEFL	Pronunciation	.71	.71	.09	.07, .11	9.40	< .001
	Lexicogrammar	.87	.16	.08	.06, .10	8.75	< .001

Table 9

Correlations Between Accentedness, Comprehensibility, and 10 Linguistic Measures Across Tasks

	Picture		IELTS		TOEFL	
Linguistic measure	Accent	Compr.	Accent	Compr.	Accent	Compr.
Segmentals	.95*	.82*	.93*	.78*	.93*	.77*
Word stress	.82*	.66*	.75*	.69*	.75*	.79*
Intonation	.72*	.58*	.78*	.64*	.82*	.79*
Rhythm	.78*	.76*	.78*	.74*	.82*	.85*
Speech rate	.56*	.77*	.55*	.72*	.62*	.83*
Lexical appropriateness	.40*	.64*	.39*	.63*	.56*	.78*
Lexical richness	.20	.59*	.38*	.69*	.34**	.76*
Grammatical accuracy	.39*	.65*	.44*	.65*	.53*	.79*
Grammatical complexity	.20	.59*	.31**	.64*	.35**	.79*
Discourse richness	.12	.55*	.28**	.66*	.35**	.78*

Note. *$p < .05$, p** $< .005$.