Language Teaching (2022), **55**, 470–489
doi:10.1017/S0261444821000537

CAMBRIDGE UNIVERSITY PRESS

RESEARCH TIMELINE

Second language speech comprehensibility

Dustin Crowther* , Daniel Holden† and Kristen Urada†

University of Hawai'i at Mānoa, Honolulu, Hawai'i
*Corresponding author. Email: dcrowth@hawaii.edu

(Received 6 July 2021; revised 20 December 2021; accepted 23 December 2021)

1. Introduction

Published in *Language Learning* in 1995, Munro and Derwing's* investigation of foreign accent, comprehensibility, and intelligibility in second language (L2) speech instigated significant change in L2 pronunciation research (Levis, 2020). A key finding was that despite the presence of a foreign accent, listeners could indeed comprehend L2 speech. Within their framework, comprehension of L2 speech was assessed along two dimensions. The first, intelligibility, was an assessment of actual listener comprehension, measured through listener transcriptions of a given utterance. The second, comprehensibility, was a scalar measure of how easy to understand listeners perceived an utterance to be. While these two measures of listener comprehension (i.e., understanding) have been shown to correlate, they have also been shown to measure different forms of understanding (Derwing & Munro, 2015). This means that while increased intelligibility can be associated with increased comprehensibility, it is still common for listeners to accurately transcribe nonnative speech while simultaneously indicating the speech to be hard to understand (i.e., evidence indicates that intelligibility outpaces comprehensibility in the development of L2 pronunciation). Research published in the 25 years since has repeatedly demonstrated that accentedness, comprehensibility, and intelligibility are partially independent dimensions of L2 speech (Munro & Derwing, 2020). Whereas Munro and Derwing's (2011) research timeline attended to the concepts of accent and broad intelligibility (i.e., inclusive of both actual and ease of understanding), the past decade has seen an increased scholarly emphasis specifically on the global speech dimension of COMPREHENSIBILITY. Given this increased scholarly interest, our timeline is presented with the goal of tracing the post-Munro and Derwing (1995*) development of L2 speech comprehensibility research.

Comprehensibility is most commonly defined in reference to listeners' perceived ease or difficulty of understanding a given utterance (Derwing & Munro, 2015). In this sense, understanding has been associated with the degree of processing difficulty experienced by the listener. For example, Munro and Derwing (1995b*) found that listeners demonstrated longer reaction times in a truth judgment task for less comprehensible speech (see also Ludwig & Mora, 2017*). While our focus on comprehensibility begins with Munro and Derwing (1995a*), we briefly note here several earlier studies that have been influential in our contemporary understanding of comprehensibility.[1] This earlier influence can most clearly be seen in Varonis and Gass (1982), who defined comprehensibility as 'how easy it is to interpret the message' (p. 125). Their 5-point scale included end points of 'I understand the sentence easily' and 'I didn't understand this sentence at all' (p. 126). Interestingly, in a later study, Gass and Varonis (1984) would again refer to the dimension of comprehensibility, though their use of listener transcriptions was more characteristic of contemporary measures of intelligibility (see Kang et al., 2018). Scalar approaches towards assessing L2 speech remained common prior to Munro and Derwing (1995a*), and indeed had been in use well before for a range of sociolinguistic interests (e.g., Lambert et al., 1960). Yet, most late-1980/early-1990 studies emphasized listeners' perceptions of foreign language

†Holden and Urada are joint second authors.
*Indicates full reference appears in the subsequent timeline.

© The Author(s), 2022. Published by Cambridge University Press

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

accent (e.g., Flege, 1988; Flege & Fletcher, 1992), or in the case of Anderson-Hsieh et al. (1992), a construct that combined both accent and intelligibility into a single scale. Utilizing a definition similar to that presented in Varonis and Gass (1982), Munro and Derwing (1995a*) firmly established comprehensibility in the sense that it is generally investigated today.

Before transitioning into the specifics of our timeline, we feel it is important to make clear the importance of previous and ongoing research into different dimensions of L2 speech. The partial independence of comprehensibility, intelligibility, and accentedness has served as the basis for proposed revisions for both how L2 pronunciation is taught and assessed. Levis (2005) described both a NATIVENESS PRINCIPLE and an INTELLIGIBILITY PRINCIPLE as present in language curriculums and discussed the need for pedagogical practices and materials that better reflect both researchers' and teachers' increasing recognition of intelligible over nativelike speech as the target of L2 pronunciation learning (see Levis, 2020, for a more elaborative discussion). Levis's proposal has served to motivate much of the research into comprehensibility over the past decade. Echoing this pedagogical shift from nativelike speech to intelligible speech, recent scholarship aiming to bridge second language acquisition (SLA) and assessment-focused research has extensively discussed the importance of emphasizing intelligible over nativelike speech in pronunciation assessment (e.g., Isaacs & Trofimovich, 2017; Kang & Ginther, 2018; see also Isaacs & Harding, 2017, whose research timeline on pronunciation assessment includes multiple references to comprehensibility). As we continue to push intelligible speech as a target of pronunciation teaching and assessment, it is necessary to understand where the field currently stands in reference to understanding relevant dimensions of intelligible speech. In the current timeline, this relevant dimension is comprehensibility.

2. Timeline methodology

The current timeline uses Munro and Derwing (1995a*) as a starting point, following up on the *Journal of Second Language Pronunciation*'s 2020 special issue entitled '25 years of intelligibility, comprehensibility and accentedness'. We present 41 empirical studies drawn from 14 peer-reviewed applied linguistics journals,[2] as a means to show how research into comprehensibility has both grown and transformed our understanding of the dimension. In line with common practices in meta-analytic research (e.g., Marsden et al., 2018; Plonsky et al., 2019), we here focus on empirical journal publications rather than book chapters, dissertations, theses, and conference proceedings to ensure included research has been through the journal peer review system. An initial review conducted by the first author identified 99 empirical studies published between 1995 and 2020 that included a measure of comprehensibility (spread across 19 journals). Only 21 studies were published prior to 2011, showing the significant growth in scholarly interest over the last decade ($N = 78$). We additionally considered post-2020 publications published during the drafting of our timeline.

Given space constraints, and in keeping with the essence of this strand in *Language Teaching*, we selectively identified 41 studies that demonstrate how contemporary research builds on, and is enriched by, that which preceded it, or which provide a novel approach towards investigating and understanding comprehensibility. Though we have strived to include a range of scholars to the extent possible, a review of the timeline will clearly reveal several prevalent names post-2010 (in addition, of course, to numerous studies published by Murray Munro, Tracey Derwing, and colleagues). Such inclusions are necessary to best represent the growth in comprehensibility research over the past 25 years.

3. Existing trends in L2 speech comprehensibility research

As previously stated, the 41 studies in our timeline either build on previous comprehensibility research or take a novel approach to investigate comprehensibility. We specifically highlight several key themes we see as prevalent across 25 years of comprehensibility research, each of which we summarize below.

As articulated in Levis (2005), both speaker and listener are essential elements in the production of intelligible speech, and thus it is no surprise that both have received focus in L2 comprehensibility research. Whereas Munro and Derwing (1995a*) featured one set of listeners (native English) assessing

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

the speech of one set of speakers (first language (L1) Mandarin learners of English), subsequent research has considered how varying both speaker L1 (e.g., Crowther et al., 2015a*; Derwing & Munro, 1997*) and listener L1 (e.g., Foote & Trofimovich, 2018*; Munro et al., 2006*) may inform comprehensibility ratings. Beyond potential L1 effects, speakers' degree of comprehensibility appears tied to a range of speaker-based variables, including length of residence (Saito, 2015*), motivation (Nagle, 2018*; Saito et al., 2018*), and aptitude (Saito et al., 2019*).

Extending directly out of listener-based considerations, and given the strong support that comprehensibility was a partially independent dimension from accentedness (e.g., Derwing & Munro, 1997*; Munro & Derwing, 1995a*, b*), attention turned to understanding what listeners attended to when rating comprehensibility. While initial research focused on segmental and suprasegmental influences (e.g., Munro & Derwing, 2006*; Kang, 2010*; Kang et al., 2010*), subsequent research revealed that listeners attended to not only phonological but also lexicogrammatical characteristics of speech (e.g., O'Brien, 2014*; Trofimovich & Isaacs, 2012*), with a strong emphasis on lexical use (Saito et al., 2016b*). However, the linguistic profile that informs listeners' perception of L2 comprehensibility appears to vary depending on a range of variables, including speaker L1 (Crowther et al., 2015a*), listener status (e.g., Foote & Trofimovich, 2018*; O'Brien, 2014*), task (e.g., Bergeron & Trofimovich, 2017*; Crowther et al., 2015b*), and various individual differences (e.g., Saito et al., 2019*).

Of additional interest has been, of course, understanding to what extent speakers are able to improve the comprehensibility of their speech, and what effect pronunciation instruction may have on this development. Longitudinal research, ranging from a few semesters of study (e.g., Saito et al., 2019*) through to seven years of immersive contact (see Derwing et al., 2006, 2008, 2013*) has shown that comprehensibility can indeed continue to develop, though such development seems highly tied to the extent to which the language is used (Derwing et al., 2013*; Zielinski & Prior, 2020*). Extending back to Derwing et al. (1998*), pronunciation instruction has fairly consistently been shown to benefit the development of comprehensible speech, with positive benefits found for instruction provided in both second and foreign language contexts (Zhang & Yuan, 2020*), as well as for both English and non-English languages (e.g., Derwing et al., 1998*; Isbell et al., 2019*).

Drawing from the above discussion, we highlight the following themes throughout our timeline:

- SPE – Speaker-based variables
- LIST – Listener-based variables
- LC – Linguistic correlates of comprehensibility
- LONG – Longitudinal development
- PED – Pedagogical inquiry
- MTD – Methodological practices

Studies thematically labelled as MTD explicitly identified as investigating/validating some aspect of methodological practice in comprehensibility research. In this sense, the types of scales used to rate comprehensibility have been the focus of methodological inquiry, whether as a means to validate current rating procedures (e.g., Isaacs & Thomson, 2013*; O'Brien, 2014*), or to introduce alternative rating options (Saito et al., 2017*). In recent years, there has been a push to investigate listeners' perception of comprehensibility as dynamic, rather than static (e.g., Nagle et al., 2019*), as well as in interactive, as opposed to monologic, performance (Trofimovich et al., 2020*). The use of MTD for specific studies does not indicate that additional methodological innovation does not exist elsewhere in our timeline, but only that such studies did not forefront their innovation. To help with framing comprehensibility within the larger body of global speech research (see Munro & Derwing, 2011), we make note of additional global dimensions of speech considered in each study:

- ACC – Accentedness
- INT – Intelligibility
- FLU – Fluency

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

Though not included in Munro and Derwing (1995a*), interest in perceived fluency, or the ease of flow of L2 speech, typically in reference to the presence/absence of pauses and other dysfluency markers (Derwing & Munro, 2015), has since often been considered in relationship to comprehensibility (e.g., Derwing et al., 1998*; Suzuki & Kormos, 2020*), and is thus considered in this timeline similarly to accentedness and intelligibility. Finally, given that ~80% of identified studies featured English as the target language, we have explicitly identified studies in which a non-English language was the target of analysis.

Notes

[1] Larry E. Smith's overarching framework of *International Intelligibility* (see Smith & Nelson, 1985) included a comprehensibility dimension. Defined as a listener's ability to understand a word or utterance, Smith's dimension is more reflective of actual understanding, aligning it more closely with Munro & Derwing's (1995a*) dimension of intelligibility.

[2] An initial list of 15 journals was taken from VanPatten & Williams (2002), which was then supplemented by an additional four journals featured in Munro & Derwing's (2011) timeline (post-1995 publications), two primary language testing journals, and the *Journal of Second Language Pronunciation* (established 2015).

References

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentais, prosody, and syllable structure. *Language Learning*, *42*(4), 529–555. https://doi.org/10.1111/j.1467-1770.1992.tb01043.x.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. 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*(3), 359–380. https://doi.org/10.1093/applin/amm041.

Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fluency development in mandarin and slavic speakers. *System*, *34*(2), 183–193. https://doi.org/10.1016/j.system.2006.01.005.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *The Journal of the Acoustical Society of America*, *84*(1), 70–79. https://doi.org/10.1121/1.396876.

Flege, J. E., & Fletcher, K. L. (1992). Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America*, *91*(1), 370–389. https://doi.org/10.1121/1.402780.

Gass, S., & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, *34*(1), 65–89. https://doi.org/10.1111/j.1467-1770.1984.tb00996.x.

Isaacs, T., & Harding, L. (2017). Pronunciation assessment. *Language Teaching*, *50*(3), 347–366. https://doi.org/10.1017/S0261444817000118.

Issacs, T., & Trofimovich, P. (Eds.) (2017). *Second language pronunciation assessment*. Multilingual Matters.

Kang, O., & Ginther, A. (Eds.) (2018). *Assessment in second language pronunciation*. Routledge.

Kang, O., Thomson, R. I., & Moran, M. (2018). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning*, *68*(1), 115–146. https://doi.org/10.1111/lang.12270.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology*, *60*(1), 44–51. https://doi.org/10.1037/h0044430.

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–377. https://doi.org/10.2307/3588485.

Levis, J. (2020). Changes in L2 pronunciation: 25 years of intelligibility, comprehensibility, and accentedness. *Journal of Second Language Pronunciation*, *6*(3), 277–282. https://doi.org/10.1075/jslp.20054.lev.

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, *68*(2), 321–391. https://doi.org/10.1111/lang.12286.

Munro, M. J., & Derwing, T. M. (2011). The foundations of accent and intelligibility in pronunciation research. *Language Teaching*, *44*(3), 316–327. https://doi.org/10.1017/S0261444811000103.

Munro, M. J., & Derwing, T. M. (2020). Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation*, *6*(3), 283–309. https://doi.org/10.1075/jslp.20038.mun.

Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2019). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*, *36*(4), 583–621. https://doi.org/10.1177/0267658319828413.

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

Smith, L. E., & Nelson, C. L. (1985). International intelligibility of English: Directions and resources. *World Englishes*, *4*(3), 333–342. https://doi.org/10.1111/j.1467-971X.1985.tb00423.x.

VanPatten, B., & Williams, J. (2002). *Research criteria for tenure in second language acquisition: Results from a survey of the field*. Unpublished manuscript, University of Illinois at Chicago.

Varonis, E. M., & Gass, S. M. (1982). The comprehensibility of nonnative speech. *Studies in Second Language Acquisition*, *4* (2), 114–136. https://doi.org/10.1017/S027226310000437X.

Dustin Crowther is an Assistant Professor in the Department of Second Language Studies at the University of Hawai'i at Mānoa. His research agenda emphasizes the attainment of intelligible speech for additional language speakers, inclusive of speaker- and listener-based variables. Specifically, he takes into account linguistic and intercultural considerations that define native-nonnative and nonnative-nonnative interaction. Given the increased spread of English, much of his research is informed by scholarship derived from Global Englishes. As an experienced language instructor, his long-term scholarly objective is to link research to pedagogy. Dr. Crowther additionally emphasizes the promotion of methodological rigor within applied linguistics research.

Daniel Holden is currently a Ph.D. candidate in the Department of Second Language Studies at the University of Hawai'i at Mānoa. His main research interests are critical administration, second language teaching and learning, language assessment, materials development, and the impact of edu-business and globalization of international English programs. His long-term research goal is to foster a stronger connection between administrators and the members of their school communities through critical dialogue. He has most recently been doing assessment-related research for a short-term intensive English program, focusing on the revision of its writing and speaking placement exams.

Kristen Urada is a Ph.D. student in the Department of Second Language Studies at the University of Hawai'i at Mānoa. Her main research interests are instructed second language acquisition, task-based language teaching, Hawai'i Creole, and speech perception. As a teacher-researcher, she aims to provide research-informed pedagogy, thus supporting the linkage of research to pedagogy, as well as making research accessible to the multilingual community through the Multi'ōlelo initiative (https://multiolelo.com).

Cite this article: Crowther, D., Holden, D., & Urada, K. (2022). Second language speech comprehensibility. *Language Teaching*, *55*(4), 470–489. https://doi.org/10.1017/S0261444821000537

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

Year	References	Annotations	Theme
1995	Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*(1), 73–97.	To determine whether targeting accent was an appropriate emphasis in pronunciation instruction, **Munro and Derwing** sought to establish if there was a relationship between accentedness, comprehensibility, and intelligibility. Native Mandarin speakers completed a picture narration. Native listeners used a 9-point scale to rate comprehensibility (1 = easy to understand, 9 = difficult to understand) and accentedness (1 = no foreign accent, 9 = strong foreign accent), and transcribed 36 short utterances as a measure of intelligibility. Despite an apparent accent, many speakers were still rated as highly comprehensible, which led the authors to propose that comprehensibility was a better predictor of intelligibility than accentedness. These findings provided evidence to show that presence of a foreign accent does not necessarily equate to low comprehensibility and intelligibility. Munro and Derwing suggested language educators should no longer consider accentedness, comprehensibility, and intelligibility on a single scale and that the dimension to be assessed should depend on the overall goals of the student and their program.	ACC, INT
1995	Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289–306.	After initially confirming the findings of Munro and Derwing (1995a)[3] in reference to the partial independence of comprehensibility and accentedness, **Munro and Derwing** considered whether degree of comprehensibility and accentedness impacted response latency (i.e., the length of time it took listeners to assess the truth value of a statement). For sentences rated as being low in comprehensibility, listeners took longer to make their judgment than they did for sentences with moderate or high comprehensibility. No such effect was found for degree of accentedness. In short, it appeared that when listeners required extra time to process an utterance, they tended to perceive the utterance as less comprehensible.	LIST

(*Continued*)

[3]Authors' names are shown in small capitals when the study referred to appears elsewhere in this timeline.

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
1997	Derwing, T. M. & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*(1), 1–16.	**Derwing and Munro** extended their earlier work, which featured L1 Mandarin learners of English (Munro & Derwing, 1995a, 1995b), by recruiting speakers from four different L1 backgrounds (Cantonese, Japanese, Polish, Spanish). L2 speakers from each L1 with intermediate proficiency were more harshly rated for accentedness than for comprehensibility, which in turn was assessed less harshly than intelligibility. Listeners' comments indicated that when rating comprehensibility, they paid attention to enunciation, grammar, speech rate, and prosody. The authors reiterated how the three dimensions of comprehensibility, accentedness, and intelligibility appeared 'closely related, but not equivalent' (p. 11).	SPE LC ACC, INT
1998	Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favour of a broad framework for pronunciation instruction. *Language Learning*, *48* (3), 393–410.	**Derwing et al.** compared the effectiveness of two pronunciation treatments over a 10-week period on comprehensibility, accentedness, and fluency development. Segmental treatment emphasized individual sounds, while global treatment focused on prosodic considerations (stress, intonation, rhythm). For a read aloud task, both treatment groups demonstrated increased comprehensibility and accentedness. However, for a picture narration, only the global group showed improvement, and only for comprehensibility and fluency. Derwing et al. reasoned that the global students more successfully carried over prosodic elements to their spontaneous speech, where segmental errors were likely less salient. The authors highlighted the benefits of including both segmental and global features in pronunciation pedagogy targeting comprehensible L2 speech.	PED ACC, FLU
2006	Munro, M., & Derwing, T. (2006). The functional load principle in ESL pronunciation instruction: An exploratory study. *System*, *34*(4), 520–531.	With an interest in identifying pedagogical targets for addressing comprehensibility, **Munro and Derwing** initiated a chain of research that investigated the extent to which different linguistic features of L2 speech informed listeners' perceptions of comprehensibility (and accentedness). In their study, they focused on a feature of segmental production, functional load (FL). They found that high FL errors had a greater effect on comprehensibility and accentedness than did low FL errors, though this effect was only cumulative for accentedness. The authors reasoned that because comprehensibility was affected not by the total number of FL errors, but by the type of FL errors made (i.e., high), listeners' perceptions were based more on the nature of the errors committed.	LC ACC

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2006	Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 111–131.	Using the same speech samples and rating procedures, **Munro et al.** extended the findings of Derwing and Munro (1997) to consider the effects of listeners' L1 background on comprehensibility, accentedness, and intelligibility assessments. Listeners representing four L1 backgrounds (English, Cantonese, Japanese, Mandarin) transcribed a set of utterances from speakers representing four L1s (Cantonese, Japanese, Polish, Spanish) while providing comprehensibility and accentedness ratings for each. In general, listeners demonstrated similarities in both their accuracy of transcription and ratings of speech across L1 backgrounds, with only weak evidence for a shared L1 effect. The authors proposed that some listeners some of the time may understand speech that shares their accent better than speech that does not, and that linguistic properties of speech may play a more significant role in listeners' understanding.	SPE LIST ACC, INT
2010	Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, *38*(2), 301–315.	**Kang** built upon Munro & Derwing (2006) by extending a focus on linguistic correlates from segmental to suprasegmental influences on comprehensibility ratings. Four temporal measures explained 33% of variation in listeners' comprehensibility ratings of L2 speech. In contrast, 39% of variation in accentedness ratings was explained by a combination of five prosodic and temporal variables. Only one measure (articulation rate) was shared by the two dimensions. For comprehensibility, listeners' perceived speakers as more comprehensible as they spoke faster (with the caveat that no speaker ever spoke at a rate faster than that of a native speaker).	LC ACC
2010	Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal*, *94*(4), 554–566.	Rather than using a scalar rating (e.g., Derwing & Munro, 1997; Munro & Derwing, 1995a), **Kang et al.** operationalized accentedness through a set of suprasegmental measures (spanning speech rate, pause, stress, and pitch considerations) to investigate the effect of accent on comprehensibility. Similar to Kang (2010), the combined suprasegmental measures accounted for ∼50% of variation in listeners' ratings of comprehensibility, a total the authors considered extraordinary given that segmental production was not included in the analysis. These findings reinforced the importance of suprasegmental production in producing comprehensible speech.	LC

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2012	Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences of listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition, 34*(3), 475–505.	After Munro & Derwing's (2006) segmental and Kang's (2010) and Kang et al.'s (2010) suprasegmental foci, **Isaacs and Trofimovich** gave additional consideration to grammatical, lexical, and discourse measures as influences on listeners' comprehensibility ratings. Making use of correlation data between 19 linguistic measures and comprehensibility ratings, along with in depth reports elicited from three experienced English language teachers, the authors proposed an initial L2 comprehensibility scale with three levels of performance (accompanied by descriptors), a first step in the development of a formative assessment instrument for L2 speech comprehensibility.	LIST LC
2012	Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905–916.	**Trofimovich and Isaacs** drew upon the data similarly featured in Isaacs and Trofimovich (2012), though with emphasis on how the linguistic profile that informed listeners' ratings of comprehensibility differed from that of accentedness. Listeners' judgments indicated a stronger grammatical and lexical influence on comprehensibility ratings than for accentedness ratings. The authors' findings served as strong evidence that listeners' judgments of comprehensibility relied on a wider range of linguistic measures than accentedness, including non-phonological measures.	LIST LC ACC
2013	Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*(2), 163–185.	One of a series of related studies targeting longitudinal pronunciation development (see also Derwing et al., 2006, 2008), **Derwing and Munro** measured the comprehensibility, accentedness, and fluency development of L1 Mandarin and L1 Slavic speakers of English over a 7-year period. While Mandarin speakers demonstrated little to no improvement in their L2 speech, Slavic speakers showed continued improvement for both comprehensibility and fluency, even after formal language instruction had been removed. The authors proposed that Mandarin speakers' greater ties with their L1 community resulted in less opportunities/willingness to engage with native users, which in turn may have limited their pronunciation development in comparison with the Slavic speakers. As these results demonstrated contrasting differences in pronunciation development between the two L1 groups, the authors made a call for more longitudinal research into L2 speech development with learners from different L1 groups.	SPE LONG ACC, FLU

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2013	Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*(2), 135–159.	Given the common use of rating scales for assessing comprehensibility (e.g., Derwing & Munro, 1997; Munro & Derwing, 1995a), **Isaacs and Thomson** investigated the effects of scale length (5-point, 9-point), with an additional consideration of listener experience (novice, experienced). While 5-point scales were used with more clarity than 9-point scales (i.e., raters distinguished between speakers more clearly), comprehensibility ratings indicated a concerning trait. For comprehensibility, 76% of speakers assigned scores between 2–7 and listeners were highly inconsistent in the score assigned to each speaker (i.e., the same speaker might receive a 3 from one listener and a 5 from another). The authors argued the need for a better understanding of the perceptions that underlie listeners' ratings of global speech dimensions.	LIST MTD ACC, FLU
2014	Derwing, T.M., Munro, M.J., Foote, J.A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, *64*(3), 526–548.	**Derwing et al.** extended pedagogical research on L2 comprehensibility development beyond the ESL classroom (see Derwing et al., 1998) to the workplace. They provided employees of a window factory with three months of pronunciation instruction. Employees demonstrated increased comprehensibility and intelligibility from training, though the extent of improvement varied per employee. While noting the difficulties of providing pronunciation training in the workplace, the authors highlighted how despite significant time spent in the target language community (~19 years), employees' pronunciation could still benefit from explicit training.	PED ACC, INT, FLU
2014	O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, *64*(4), 715–748.	Following Munro et al. (2006), **O'Brien** additionally elicited speech ratings from nonnative listeners, extending knowledge of linguistic correlates beyond just native listeners' perceptions. Nonnative listeners found L2 speakers to be comprehensible despite possessing an accent, with comprehensibility ratings drawing on a range of phonological and lexicogrammatical influences, as previously found for native listeners in Trofimovich and Isaacs (2012). Furthermore, listener characteristics, such as language background, courses taken, and exposure to nonnative speakers, contributed to the differences in comprehensibility ratings of native versus nonnative German listeners.	LIST LC ACC, FLU L2: German

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2015	Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2015a). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly, 49*(4), 814–837.	**Crowther et al.** extended Trofimovich and Isaacs's (2012) focus on which linguistic measures inform listener perception of comprehensibility and accentedness by including speakers from different L1 backgrounds (see also Derwing & Munro, 1997). While listeners considered similar phonological features when rating three L1 groups for accentedness, group differences existed for comprehensibility. The authors argued these findings indicated there is no single linguistic measure universally predictive of speakers' comprehensibility.	SPE LC ACC
2015	Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015b). Does a speaking task affect second language comprehensibility? *Modern Language Journal, 99*(1), 80–95.	With an aim to extend consideration of linguistic correlates of comprehensibility beyond a picture narration (i.e., O'Brien, 2014; Trofimovich & Isaacs, 2012), **Crowther et al**. compared listeners' ratings across a less complex IELTS-inspired task and a more complex TOEFL-inspired task. Listeners rated IELTS-elicited speech as more comprehensible, while TOEFL-elicited speech ratings appeared to rely on a wider range of phonological and lexicogrammatical measures. These findings were moderated by speakers' L1 status, as seen previously in Crowther et al. (2015a).	SPE LC
2015	Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, *65*(3), 563–595.	Following Derwing and Munro's (2013) call for more longitudinal research into pronunciation development, **Saito** investigated the effects of length of residence (LOR) on L2 oral development. LOR (short, mid, long) was a significant predictor of comprehensibility, but not accentedness, ratings, though comprehensibility development appeared to plateau after ∼3 years of residence. As seen previously (Crowther et al., 2015a, 2015b; Trofimovich & Isaacs, 2012), comprehensibility ratings were associated with both phonological and lexicogrammatical measures, though different patterns were identified based on speakers' LOR.	SPE LC ACC
2016	Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation*, *2*(1), 56–92.	Building on previous intervention research (e.g., Derwing et al., 1998, 2014), **Gordon and Darcy** evaluated to what extent providing a modest amount of short-term explicit pronunciation instruction can lead to improved comprehensibility in English as a Second Language (ESL) classrooms. Results indicated that even short-term explicit pronunciation instruction can lead to increased comprehensibility, though this was only the case for a group receiving suprasegmental-based instruction (as opposed to vowel-based and nonexplicit instructional groups).	PED

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2016	O'Brien, M. G. (2016). Methodological choices in rating speech samples. *Studies in Second Language Acquisition*, *38*(3), 587–605.	**O'Brien** extended Isaacs and Thomson's (2014) methodological inquiry, which focused on scale length, to rater condition. While one listener group rated all speakers for comprehensibility, accentedness, and fluency at the same time, another group rated each speaker three separate times, once for each dimension. While native and nonnative listeners considered as a single group indicated no difference between the two rating conditions, when considered separately, nonnative listeners who rated each dimension together were more favorable in their ratings of comprehensibility than those nonnative listeners who rated each dimension separately. Given that listeners, both native and nonnative, differentiated between all three dimensions, O'Brien argued in support of Munro and Derwing's (1995a) suggestion to make clear which dimension of L2 speech is the target of assessment.	LIST MTD ACC, FLU L2: German
2016	Saito, K., Trofimovich, P., & Isaacs, T. (2016a). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, *37*(2), 217–240.	To further investigate which linguistic measures listeners' pay attention to when rating L2 speech (e.g., Crowther et al., 2015a, 2015b; O'Brien, 2014; Saito, 2015), **Saito et al.** considered ratings assigned to L2 speakers with different levels of perceived comprehensibility (low beginner, high beginner, intermediate, advanced). Ratings of word stress and intonation clearly differentiated between groups, with additional features distinguishing between other ability pairings. The authors proposed listeners' perception of comprehensibility drew upon different linguistic measures at different levels of oral ability, with only word stress and intonation equally important across levels.	SPE LC ACC
2016	Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016b). Lexical correlates of comprehensibility versus accentedness in second language speech. *Bilingualism: Language and Cognition*, *19*(3), 597–609.	**Saito et al.** provided a more nuanced measure of the effect of lexicon on comprehensibility ratings, building off of initial findings in Trofimovich and Isaacs (2012) and Crowther et al. (2015a, 2015b). Several lexical measures made significant contributions to listeners' perceptions of comprehensibility beyond that of pronunciation. Specifically, listeners' perceptions were informed by fluency, lexical appropriateness, lexical abstractness, and sense relations (i.e., range of meanings). The authors argued that, when rating comprehensibility, listeners likely drew on all available linguistic information, whether phonological or lexical.	LC ACC

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2017	Bergeron, A., & Trofimovich, P. (2017). Linguistic dimensions of L2 accentedness and comprehensibility: Exploring task and listener effects in second language French. *Foreign Language Annals*, *50*(3), 547–566.	Similar to O'Brien (2014), who considered linguistic correlates of comprehensibility beyond English, **Bergeron and Trofimovich** placed emphasis on L2 French speech, with speaking task included as a moderating variable (see Crowther et al., 2015b). As had been established (e.g., Crowther et al., 2015a; Trofimovich & Isaacs, 2012), comprehensibility for a picture narration was associated with a wider range of linguistic measures than was accentedness. However, for a more cognitively complex interview task, a similar profile of linguistic correlates existed for both comprehensibility and accentedness.	LIST LC ACC L2: French
2017	Ludwig, A., & Mora, J. C. (2017). Processing time and comprehensibility judgments in non-native listeners' perception of L2 speech. *Journal of Second Language Pronunciation*, *3*(2), 167–198.	**Ludwig & Mora** investigated the extent to which comprehensibility ratings were informed by the degree of processing required when hearing familiar (L1 match) versus unfamiliar (L1 mismatch) L2 speech. L1-Catalan, L1-German, and L1-English speakers produced word- and sentence-level stimuli, which were presented to nonnative listeners that had differing L1 backgrounds and L2 proficiency levels, as well as a native speaking control group. When comprehensibility ratings were considered in light of listeners' reaction times during a word identification task and a sentence verification tasks, listeners tended to react more quickly to speech they also rated as more comprehensible, echoing the findings of Munro and Derwing (1995b). However, this association was stronger for native versus nonnative listeners, and high-proficiency versus low-proficiency listeners. These findings reinforced not only the relationship between comprehensibility and processing difficulty, but also the important role of listeners' background characteristics in nonnative speech perception (see also Crowther et al., 2015).	SPE LIST

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2017	Saito, K., Trofimovich, P., Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*(4), 439–462.	Whereas Isaacs and Thomson (2013) reviewed the frequently employed 5-point and 9-point scales for speech rating, **Saito et al.** aimed to validate the use of 1000-point sliding scales for rating both global (i.e., accentedness, comprehensibility) and linguistic (i.e., phonological, lexicogrammatical) performance. They elicited ratings from NS listeners, half with linguistic and/or language teaching experience and half without. They found high correlations between 18 coded linguistic measures first reported for the same speech samples in Trofimovich and Isaacs (2012) and 11 scalar measures included in their current study (though less in number, the authors' measures were drawn from the same range of phonological and lexicogrammatical considerations), though these correlations were slightly more consistent for experienced listeners. As in Trofimovich and Isaacs (2012), comprehensibility ratings were informed by both pronunciation and lexicogrammatical measures. The authors discussed how their findings indicated that raters, trained and untrained, could evaluate complex and intuitive linguistic categories, though experienced raters rated more leniently and consistently, the latter indicating a better understanding of the rating scale.	LIST LC MTD ACC
2018	Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2018). Linguistic dimensions of L2 accentedness and comprehensibility vary across speaking tasks. *Studies in Second Language Acquisition, 40*(2), 443–457.	As a reanalysis of data presented in Crowther et al. (2015a, 2015b), the authors revisited the partial independence of comprehensibility and accentedness in reference to task complexity. Across tasks, lexicogrammar measures were associated with comprehensibility ratings. However, as task complexity increased (long turn -> picture narration -> skill integration), so did the influence of lexicogrammatical measures on listeners' ratings of speakers' accentedness. **Crowther et al.** proposed that the partial distinction between comprehensibility and accentedness appeared task specific, and this distinction blurred as task complexity increased, a finding that echoed Bergeron and Trofimovich (2017), this time for L2 English speech.	LC ACC

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2018	Foote, J. A., & Trofimovich, P. (2018). Is it because of my language background? A study of language background influence on comprehensibility judgments. *Canadian Modern Language Review*, *74* (2), 253–278.	To extend previous inquiry into both speaker L1 (e.g., Crowther et al., 2015a, 2015b) and listener L1 (e.g., O'Brien, 2014), Foote and Trofimovich first investigated which linguistic measures contributed to listeners' rating of comprehensibility. They then considered whether the contributing measures would change based on the L1 of the listener or the speaker. Though Mandarin listeners (as opposed to French and Hindi) were the only group who tended to downgrade non-L1 Mandarin speakers compared with L1 Mandarin speakers, all listeners made more positive comments when rating speakers with a shared L1. In addition, different profiles of linguistic correlates existed for different speaker-listener groups. The authors argued that given similar limited shared L1 effects in previous research (see also Munro et al., 2006), scholarly emphasis should be placed more on which linguistic components of speech contribute to comprehensibility within different speaker-listener combinations.	SPE LIST LC
2018	Isaacs, T., Trofimovich, P., & Foote, J. A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing*, *35*(2), 193–216.	In support of an increased emphasis on producing understandable over nativelike speech in language testing, **Isaacs et al.** pursued further the development of an L2 comprehensibility scale for the purposes of formative assessment. Employing an iterative process drawing upon input from experienced English language teachers, the authors revised the scale first proposed in Isaacs and Trofimovich (2012). The final iteration of the scale featured five levels of performance for four elaboratively described criteria categories (pronunciation, fluency, grammar, vocabulary). The authors indicated their intent to validate their scale in future research.	LIST MTD
2018	Nagle, C. (2018). Motivation, comprehensibility, and accentedness in L2 Spanish: Investigating motivation as a time-varying predictor of pronunciation development. *Modern Language Journal*, *102*(1), 199–217.	As O'Brien (2014) extended inquiry from English to German, **Nagle** did so as well by considering L2 pronunciation development in Spanish. Whereas Derwing and Munro's (2013) longitudinal investigation considered development beyond formal classroom training, Nagle looked at the extent to which pronunciation (comprehensibility, accentedness) improved over one year of communicative-based instruction. In addition, he continued a trend towards understanding how individual differences, specifically motivation, affected learners' development over time (see Saito, 2015). Over this one-year period, while the learners' motivation was associated with decreased accentedness, no relationship was found for improved comprehensibility.	SPE LONG ACC L2: Spanish

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2018	Saito, K., Dewaele, J. M., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, *68*(3), 709–743.	Following Saito (2015) and Nagle (2016), **Saito et al.** focused on the extent to which individual differences (i.e., emotion, motivation, experience) predicted comprehensibility development over time. Changes in speakers' comprehensibility over three months were linked to motivational (Ideal L2 Self) and emotional (Private Enjoyment and Anxiety) states, as well as recent extracurricular L2 use (e.g., formal after school study, conversation activities).	SPE LONG
2019	Isbell, D. R., Park, O. S., & Lee, K. (2019). Learning Korean pronunciation: Effects of instruction, proficiency, and L1. *Journal of Second Language Pronunciation*, *5*(1), 13–48.	While previous pronunciation instruction research focused primarily on English (Derwing et al., 1998, 2014; Gordon & Darcy, 2016), **Isbell et al.** extended inquiry to L2 Korean development. In line with previous findings for English, pronunciation instruction was found to improve comprehensibility over time, with second year Korean learners benefiting more than first year learners (who had improved whether they received explicit pronunciation instruction or not).	PED ACC L2: Korean
2019	Nagle, C. (2019). Developing and validating a methodology for crowdsourcing L2 speech ratings in Amazon Mechanical Turk. *Journal of Second Language Pronunciation*, *5*(2), 294–323.	With the majority of speech rating research having occurred in a face-to-face environment, **Nagle** investigated the potential for eliciting crowdsourced online speech ratings. He used Amazon Mechanical Turk (AMT), a remote, crowdsourcing platform that provided small monetary compensation to AMT users for completing an assigned task. Raters' assessment of all global dimensions proved highly reliable (> . 89), on par with laboratory-based studies (e.g., Derwing and Munro, 2013). However, raters used comprehensibility and fluency scales more consistently than the accentedness scale. Nagle suggested that the use of AMT can allow for efficient and reliable speech ratings, assuming the inclusion of appropriate quality control measures, such as attention and validity checks.	MTD ACC, FLU L2: Spanish

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2019	Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*, *41*(4), 647–672.	Comprehensibility research thus far had focused on comprehensibility perceptions as static or unchanging. As such, **Nagle et al.** considered listeners' moment-to-moment judgments. Native listeners rated speech by upgrading and downgrading comprehensibility as often as they deemed necessary over the length of ∼150–290 second utterances (as opposed to providing a singular scalar rating representative of the entire utterance). Analyses indicated only 25% of listeners demonstrated a clear dynamic profile (i.e., most only upgraded or downgraded speakers once or twice across a sample). Increases in comprehensibility were linked to discourse considerations, while decreases were attributed to pronunciation, grammar, and vocabulary. The authors raised the question of how listeners established an initial baseline of comprehensibility, specifically when tasked to upgrade or downgrade a speaker.	LC MTD L2: Spanish
2019	Saito, K., Suzukida, Y., & Sun, H. (2019). Aptitude, experience, and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition*, *41*(1), 201–225.	Extending their earlier work looking at individual differences in L2 pronunciation development (see Saito, 2015; Saito et al., 2018), **Saito et al.** specifically investigated language learning aptitude. Whereas comprehensibility gains during an initial semester of study were associated primarily with in-class language use (and increased syllable accuracy), gains during a second semester of study were associated with sound sequence recognition (along with increased segmental accuracy). Saito et al. proposed that comprehensibility may develop more rapidly initially under new classroom conditions, but long-term development of difficult pronunciation features is tied to learners' language learning aptitude.	SPE LC LONG

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2019	Strachan, L., Kennedy, S., & Trofimovich, P. (2019). Second language speakers' awareness of their own comprehensibility: Examining task repetition and self-assessment. *Journal of Second Language Pronunciation*, *5*(3), 347–373.	Placing emphasis on the speaker, **Strachan et al.** investigated to what extent comprehensibility could be explicitly manipulated by learners. Speakers formed two groups (unprompted, prompted) and completed a graph interpretation task twice. Between performances, speakers self-rated their comprehensibility. Prior to their second performance, the prompted group were asked to make their speech easier to understand. Both speakers and native listeners rated each recording for comprehensibility, and ratings indicated neither group improved their comprehensibility between tasks. Alignment between self- and listener-ratings appeared stronger for speakers with lower comprehensibility. When speakers were asked how they could make their speech more comprehensible, they placed an emphasis on vocabulary and organization. Given limited alignment between self- and listener-ratings, Strachan et al. discussed a pedagogical need to raise speakers' awareness of, and provide strategies for, improving comprehensibility.	SPE LC
2019	Taylor Reid, K., Trofimovich, P., & O'Brien, M. G. (2019). Social attitudes and speech ratings: Effects of positive and negative bias on multiage listeners' judgments of second language speech. *Studies in Second Language Acquisition*, *41*(2), 419–442.	Introducing a new listener variable of interest, **Taylor Reid et al.** manipulated listeners' social biases prior to speech rating. Native English speakers from the anglophone population of Québec, Canada formed three experimental groups. In group A, listeners heard a rehearsed opinion from a researcher praising the L2 English skills of francophone speakers. In group B, listeners heard a more critical opinion. Group C served as a baseline who received no biasing anecdote. For comprehensibility, positively biased listeners rated speakers significantly higher than the baseline. Only older, negatively biased listeners provided lower ratings compared with the baseline. Interestingly, despite hearing a negatively biasing anecdote, younger listeners still rated comprehensibility higher than the baseline. The authors discussed the importance of raising listeners' awareness of preexisting biases prior to rating.	LIST LC ACC

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2020	Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*(1), 143–167.	**Suzuki and Kormos** investigated the extent to which listeners distinguished between the dimensions of comprehensibility and fluency. Listeners' ratings of comprehensibility were associated with several temporal measures of speech, including both speed and breakdown measures (along with morphological accuracy). The authors argued that comprehensibility and fluency are conceptually overlapping dimensions, a claim similar to that put forth by Munro and Derwing (1995a) regarding the relationship between comprehensibility and accentedness.	LC ACC, FLU
2020	Taylor Reid, K., O'Brien, M. G., Trofimovich, P., & Bajt, A. (2020). Testing the malleability of teachers' judgments of second language speech. *Journal of Second Language Pronunciation*, *6*(2), 236–264.	As a follow up to Taylor Reid et al.'s (2019) investigation into how social biases of community members impacted L2 speech ratings, **Taylor Reid et al.** placed emphasis on potential biases amongst native and nonnative teachers of German. Following a similar procedure to the predecessor study, teachers were divided into two groups: baseline and negative bias. For comprehensibility, when no negative bias was provided, native listeners were more lenient in their ratings than nonnative listeners. However, when a negative bias was presented, nonnative listeners provided similar ratings to their non-biased counterparts, while native listeners tended to downgrade speakers compared to their counterparts. Just as Taylor Reid et al. (2019) argued for the importance of raising listeners' awareness of their biases, the authors highlighted the importance of language teachers being made aware of their own biases and how these biases may be affected by comments provided by others.	LIST ACC L2: German
2020	Trofimovich, P., Nagle, C. L., O'Brien, M. G., Kennedy, S., Taylor Reid, K., & Strachan, L. (2020). Second language comprehensibility as a dynamic construct. *Journal of Second Language Pronunciation*, *6*(3), 430–457.	Whereas Nagle et al. (2019) focused on listeners' changing perceptions while assessing pre-recorded, monologic speech (i.e., a one-way listening task), **Trofimovich et al.** considered how an interlocutor's perception of their speaking partner changed across three collaborative tasks (i.e., during real time interaction). Interlocutors provided ratings during the three tasks at roughly equal intervals. Across tasks, comprehensibility ratings were U-shaped, with the first and third task rated relatively equally in terms of comprehensibility, and the middle task rated more severely. Additionally, where ratings appeared to be mismatched initially (one interlocutor was rated as more comprehensible), interlocutors appeared to align in ratings relatively quickly (i.e., over time, interlocutors began to rate each other similarly).	MTD

(*Continued*)

https://doi.org/10.1017/S0261444821000537 Published online by Cambridge University Press

(*Continued*)

Year	References	Annotations	Theme
2020	Zhang, R., & Yuan, Z. (2020). Examining the effects of explicit pronunciation instruction on the development of L2 pronunciation. *Studies in Second Language Acquisition*, *42*(4), 905–918.	**Zhang and Yuan** partially replicated Derwing et al.'s (1998) pedagogical intervention, with a focus on a foreign language learning environment (China). L1 Mandarin learners of English received either segmental or global instruction, with only the global group improving from pretest through delayed posttest on a picture narration (both groups improved on a sentence reading task). These findings continued to point towards the overall positive effects of pronunciation instruction, no matter the learning environment, and more specifically, the effects of global pronunciation instruction (i.e., prosody-based) on the production of comprehensible spontaneous speech.	PED
2020	Zielinski, B., & Pryor, E. (2020). Comprehensibility and everyday English use: An exploration of individual trajectories over time. *Journal of Second Language Pronunciation*, *6*(3), 352–379.	**Zielinski and Pryor** investigated the relationship between language use and comprehensibility development over a 10-month period amongst Australian migrants. While all but one participant demonstrated a change in comprehensibility (whether positive or negative), no two participants demonstrated the same trajectory. A relationship, though tenuous, existed between comprehensibility development and migrants' self-reported language use across a range of contexts (e.g., classes, friends, work); those with comprehensibility increases tended to demonstrate higher English use. This finding aligned well with Derwing and Munro's (2013) proposed explanation for differences between their L1 Mandarin and L1 Slavic learners.	SPE LONG
2021	Huensch, A., & Nagle, C. (2021). The effect of speaker proficiency on intelligibility, comprehensibility, and accentedness in L2 Spanish: A conceptual replication and extension of Munro and Derwing (1995a). *Language Learning, 71*(3), 626–668.	**Huensch and Nagle** presented a conceptual replication of Munro and Derwing (1995a), with a focus on L2 Spanish. Native Spanish speakers rated comprehensibility and accentedness on sliding scales (see Saito et al. 2017), while also transcribing each utterance as a measure of intelligibility (using AMT following procedures established in Nagle, 2019). The authors found comprehensibility ratings more aligned with intelligibility scores than accentedness ratings. As in Saito et al. (2016a), proficiency appeared to mediate which linguistic measures were associated with each global speech measure. Huensch and Nagle's study, as a conceptual replication, serves to demonstrate how despite extensions in theoretical interests and advances in methodological procedures over ∼25 years, the findings of Munro & Derwing (1995a) still stand today; specifically, that comprehensibility, accentedness, and intelligibility appear to be partially independent dimensions.	SPE LC ACC, INT L2: Spanish