Language Teaching (2024), **57**, 533–549
doi:10.1017/S026144482300037X

CAMBRIDGE UNIVERSITY PRESS

THINKING ALLOWED

Second language speech comprehensibility: A research agenda

Dustin Crowther and Daniel R. Isbell

University of Hawai'i at Mānoa, Honolulu, USA
Corresponding author: Dustin Crowther; Email: dcrowth@hawaii.edu

(Received 7 November 2022; revised 19 April 2023; accepted 7 August 2023)

Abstract

Interest in second language (L2) speech comprehensibility, or listeners' perceived ease or difficulty of understanding a given utterance, has seen continual growth since Munro and Derwing's (1995) seminal publication in *Language Learning*. While recognizing the body of knowledge that has developed in the 25 plus years since Munro and Derwing's study, in the present paper we look to the future as we consider what an active and informative research agenda might look like for the next 10–15 years. In this regard, we propose four primary areas of future research, including a need to: (1) extend inquiry beyond a primary focus on L2 English speech, (2) explore comprehensibility more thoroughly as a social construct, (3) investigate the effects of more targeted task manipulations, and (4) delve deeper into the effects of comprehensibility-oriented instruction. We additionally discuss the need for more replication in L2 comprehensibility research while also promoting several scholarly disciplines we feel can greatly inform future research, such as task-based literature.

1. Introduction

In the current landscape of second language (L2) pronunciation research, it is generally agreed upon that what Levis (2005) labeled as the Intelligibility Principle serves as a more appropriate guide for the purposes of both pronunciation acquisition and teaching than a so-called Nativeness Principle (see also Levis, 2020a). More simply, the target of L2 pronunciation learning should be the production of understandable rather than nativelike speech. Such a claim does not seem controversial in the broader literature, given increased emphasis on L2 users' multilingual repertoires (e.g., Ortega, 2019). As evident by a pair of previous *Language Teaching* research timelines (Crowther et al., 2022; Munro & Derwing, 2011), as well as the 2020 special issue of the *Journal of Second Language Pronunciation* (see Levis, 2020b), listeners' global evaluations of L2 speech have served as a primary source of data within L2 pronunciation research, particularly in reference to three key global dimensions made popular in Munro and Derwing (1995a): accentedness, intelligibility, and the focus of this paper, comprehensibility.

Though not the first reference to comprehensibility in L2 speech research (see Smith & Nelson, 1985; Varonis & Gass, 1982), Munro and Derwing (1995a) firmly established comprehensibility in the sense that is it investigated today; that is, as a listener's perceived ease or difficulty of understanding a given utterance (Derwing & Munro, 2015). As highlighted in Crowther et al. (2022), scholarly interest in L2 comprehensibility has grown since Munro and Derwing (1995a), with a particularly sharp increase post-2010. In our current paper, we first elaborate on the construct of comprehensibility, differentiate it from its sister construct of intelligibility, and provide support for the increased scholarly interest in the construct. After briefly highlighting existing trends in comprehensibility research, we set out what we see as key themes for investigating L2 speech comprehensibility over the next 10–15 years.

© The Author(s), 2023. Published by Cambridge University Press

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

In doing so, we aim to not only highlight WHAT might be done moving forward, but also HOW, building on recent publications at the forefront of these proposed agendas.

1.1 Why comprehensibility as a measure of understanding

When asked to judge an L2 utterance for degree of comprehensibility, listeners are essentially evaluating how easy or difficult to understand they perceive the said utterance to be. This has primarily been done through some form of scalar rating (Munro & Derwing, 2015; see Isaacs & Thomson, 2013, and O'Brien, 2016, for further discussion on rating procedures). Such evaluations are thought to reflect the degree of processing difficulty experienced by a listener (e.g., Munro & Derwing, 1995b). This perceptual nature is what differentiates comprehensibility, as a construct, from INTELLIGIBILITY, another measure of understanding similarly put forth in Munro and Derwing (1995a). Intelligibility represents "the degree of match between a speaker's intended message and the listener's comprehension" (Derwing & Munro, 2015, p. 5), or, more simply, intelligibility is the accuracy of a listeners' understanding. Levis (2018) highlighted how intelligibility, in this sense, may refer to both broad and narrow interpretations. Broadly speaking, intelligibility may be measured at the semantic level, frequently through listeners' summaries of what they understood or responses to comprehension questions. A narrower interpretation focuses on the extent to which listeners are able to decode the lexical items that make up a given utterance, measured primarily through transcription tasks (see Levis & Silpachai, 2021, and Kang et al., 2018, for further theoretical and methodological discussions, respectively). As comprehensibility and intelligibility are both measures of understanding, it is unsurprising that scores are often moderately-to-strongly correlated, though the unique focus of each allows for finer-grained examination of listener understanding (Derwing & Munro, 2015). Specifically, while increased intelligibility (i.e., when listeners indicate increased accuracy of understanding) can be associated with increased comprehensibility (i.e., when listeners indicate speech to be easier to understand), it is also possible that two speakers may be assessed as equally intelligible despite different ratings of comprehensibility (i.e., when a listener's accuracy scores for two speakers are equivalent, even if one speaker required greater listener effort to be understood). As comprehensible speech is more likely to be intelligible than unintelligible (Derwing & Munro, 2015), it would seem that the acquisition of intelligible speech may outpace the acquisition of comprehensible speech. As such, Thomson (2018) proposed that a pedagogical focus on comprehensible speech would likely provide intelligibility benefits as well. Such an interpretation could help to explain the substantial increase in L2 comprehensibility research highlighted in Crowther et al. (2022), though Saito (2021), in his meta-analytic report, also highlighted the construct's "strong ecological validity, as it is assumed to reflect the instant and impressionistic judgements by interlocutors during oral communication in real-life contexts" (p. 86). Without downplaying the overall importance of attaining intelligible speech to L2 spoken language development (see Levis, 2018), the current paper focuses on the increased scholarly interest in the comprehensibility of L2 speech.

To return to where we began this discussion, L2 pronunciation scholars now strongly advocate for understandable speech over nativelike speech as the primary target of acquisition and teaching (Levis, 2005, 2020a). A scholarly interest in comprehensibility is strongly aligned with this belief, especially given substantial evidence that despite possessing a foreign accent, L2 speakers are able to produce speech evaluated by listeners as highly comprehensible. This finding was first highlighted in Munro and Derwing (1995a) and has received additional support through both original (e.g., Crowther et al., 2015a; Trofimovich & Isaacs, 2012) and replication (Huensch & Nagle, 2021) research.

1.2 Developing a research agenda

At our invited colloquium for the 2023 *American Association for Applied Linguistics* (AAAL) conference, we highlighted trends in L2 comprehensibility research that have strongly defined the field. These trends included:

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

- disentangling the relationships among comprehensibility, intelligibility, and accentedness;
- identifying which linguistic features listeners attend to when assigning comprehensibility (vs. accentedness) ratings;
- examining speaker and listener differences in the production and perception of comprehensible speech;
- investigating pedagogical effects on comprehensibility development; and
- exploring methodological choices in assessing comprehensibility.

We did not argue at AAAL, nor do we argue here, that these topics have been exhausted. Rather, we propose that the body of knowledge gained from these lines of research now allows for new routes of inquiry to take shape, which can help to shore up empirical bases and chart new courses to advance theory and understanding in L2 comprehensibility research. In what follows, we provide a summary of four key themes that we see as exciting avenues for comprehensibility research. As the themes proposed cover theoretical, methodological, and pedagogical ground, we do not argue for the importance of one over another. Rather, each holds promise for advancing our understanding of comprehensibility. For each theme, we provide at least one concrete research task (summarized in Table 1). An additional fifth theme arguing the need for more replication research is also included. We believe these themes and associated research tasks could usefully guide comprehensibility research over the next 10–15 years.

2. Target language extension

In Crowther et al. (2022), one key finding reported was that ~80% of the studies identified in our initial search had listeners assessing the L2 English speech of participants, a trend frequent across different areas of L2 pronunciation research (see Levis, 2021). Of the 41 studies included in our timeline, only nine studies elicited comprehensibility ratings of languages other than English (LOTEs), with the first study not appearing until 2014. Among these studies, only four non-English target languages were considered: Spanish ($n = 4$), German ($n = 3$), French ($n = 1$), and Korean ($n = 1$), the last of which was published by the second author (Isbell et al., 2019).[1] Given the parameters of Crowther et al.'s search (peer reviewed journal publications), it is indeed likely that additional LOTE comprehensibility

Table 1. Overview of key research themes and task proposals

Key research themes	Proposed research tasks
Target language extension	A. Investigate LOTEs
Comprehensibility as a social construct	B. Investigate comprehensibility in interaction through interlocutor perceptions and actions
	C. Explore a wider range of social variables
Speaking task considerations	D. Investigate targeted manipulations of task complexity
	E. Investigate the consistency of speaker performance relative to their peers across tasks
Pedagogical practices	F. Investigate comprehensibility-oriented instruction with practicing classroom teachers
	G. Conduct more extensive longitudinal investigations of (instructed) comprehensibility development
Replication in L2 comprehensibility scholarship	

[1]We note that Crowther et al.'s (2022) timeline consisted of 41 studies purposefully selected from a more extensive collection of L2 comprehensibility studies. It is from this more extensive collection that we calculated ~80% of studies focused

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

research has been conducted, though LOTEs would still seem underrepresented in the comprehensibility literature. For further evidence of an L2 English bias, consider that Saito (2021), the only published comprehensibility-oriented meta-analysis at the time of writing, included only studies that focused on English-as-a-second-language speakers.

2.1 Research Task A: Investigate LOTEs

While a bias towards L2 English speech is understandable given the vast global use of English, such a hegemonic emphasis has also been a source of concern in L2 research (e.g., Ushioda, 2017), as it is unclear how well findings tied to L2 English learning are generalizable to LOTEs. The prevalence of English in global communication and the emergence of numerous regional/national varieties associated with diverse first languages (L1s) and L2s make its sociolinguistic landscape distinct from many other languages. To provide a concrete example, the second author's research on L2 Korean comprehensibility (Isbell et al., 2019; Isbell & Lee, 2022) has found very strong correlations between comprehensibility and accentedness ($r \geq .90$), which calls into question the basic distinguishability of the two constructs now taken for granted on the basis of L2 English research where weaker correlations are typically found. The linguistic structure of English, too, can hardly be considered representative of the world's languages. Research on L2 English has highlighted the importance of suprasegmental pronunciation features in comprehensibility (and in comprehensibility-oriented pedagogy), but some notable features that drive such findings in English, like lexical stress, are completely absent from many languages. Levis (2021) highlighted how "different languages have different phonological features, different combinations of phonological features, and different types of phonological learning have their own challenges" and provides the example that while "all languages seem to make use of pitch and length differences to communicate meaning in one way or another, L2 learners may find the challenges more or less difficult depending on how pitch and length are used" (p. 144). Given that such phonological differences exist across languages, it seems necessary to investigate the extent to which our understanding of L2 speech comprehensibility is generalizable beyond L2 English.

Any investigation into LOTEs would not necessitate a researcher to stray too far from existing methodological practices. Instead, an increased use of replication in comprehensibility research is suggested (see Porte & McManus, 2019). Very few replication studies exist amongst published comprehensibility research, though Huensch and Nagle (2021) can serve as an exemplary example of how conducting replication research can help to solidify our current understanding of comprehensibility while simultaneously addressing a lack of focus on LOTEs (see also Isbell & Lee, 2022). Framed as a conceptual replication of Munro and Derwing (1995a), Huensch and Nagle made several methodological adjustments, all informed by post-1995 scholarship. The key focus of their replication, though, was to shift away from a focus on "advanced L2 English speakers in second (rather than foreign) language contexts" (p. 627). Instead, they focused on L2 Spanish learners attending a US university (i.e., foreign language context), with the focus on Spanish motivated by the language's popularity and importance in the US. With Crowther et al. (2022) already comparing and contrasting the methods of their study with those used in Munro and Derwing (1995a), we here simply highlight that Huensch and Nagle's key findings provided additional support for those of Munro and Derwing. Essentially, by replicating Munro and Derwing (1995a), Huensch and Nagle provided evidence that the relationship between comprehensibility and the dimensions of intelligibility and accentedness were not restricted only to L2 English speech but held true for at least one other LOTE. Given that Huensch and Nagle focused only on L2 Spanish, further inquiry into other LOTEs seems necessary. We return to the topic of replication near the end of our paper, but we note now that a strong basis for replication in comprehensibility research can be simply seen as investigating to what extent findings for L2 English speech are generalizable to LOTEs.

on L2 English. However, within this larger collection, LOTEs additionally included Dutch, Japanese, Mandarin, Portuguese, Thai, and Turkish.

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

To more clearly articulate a direction forward, comprehensibility scholars might choose a high-quality comprehensibility study with English as a target language and replicate it with a different target language, as Huensch and Nagle (2021) did with Munro and Derwing (1995a); see also Saito and Akiyama's (2017) conceptual replication of Trofimovich and Isaacs (2012), in which they substituted Japanese for English as the target language. A more elaborative step would be to perform a replication of such high-quality studies with more than one target language, and to report findings in a single paper. For example, a replication that conducts parallel studies with both French and Chinese Mandarin as target languages would enable important comparisons to the original English findings. French differs from English in the sense of rhythmic timing (syllable- vs. stress-timed). Chinese Mandarin differs from English both rhythmically and in reference to the use of tone. In addition, comparisons between French and Chinese Mandarin, which are similar rhythmically but differ in reference to tone, would prove revealing. The selection of languages to be targeted may also be considered along sociolinguistic grounds, such as the perceived prestige of the target language in a given community. Ultimately, the purpose of any replication is to extend a body of knowledge that has been predominately focused on L2 English speech, a language that holds a particular global status that differs from most other languages in the world. Replication of existing studies can inform us on the extent to which what we know about what makes L2 English speech comprehensible is applicable to non-English L2 speech, which in turn informs both theory building and L2 pronunciation pedagogy. For an initial list of high-quality studies that might inform such replication, see Crowther et al. (2022).

3. Comprehensibility as a social construct

One reason to focus on comprehensibility over intelligibility as a measure of understanding is that it may serve as a window into listeners' instant and impressionistic judgments in real-life contexts (Saito, 2021). For example, even for intelligible speech, decreased comprehensibility may elicit more negative listener reactions towards an L2 speaker, which in turn may affect listeners' willingness to put in the effort necessary to understand a speaker (i.e., a speaker may simply disregard an L2 speaker, as seen in Vujinović, 2017). Until relatively recently, comprehensibility research has remained constrained to the rating of monologic speech (i.e., a one-way listening task), with listeners assigning a single scalar rating at the conclusion of an utterance. Nagle et al. (2019), arguing that comprehensibility should be viewed as a dynamic rather than static construct (i.e., listeners' perceptions are likely to change over time), asked listeners to upgrade and downgrade their evaluations of a speakers' comprehensibility over the course of ∼150–290-second utterances. While Nagle et al. (2019) still maintained a focus on monologic speech, Trofimovich et al. (2020) employed a similar approach across a series of collaborative tasks. They asked each member of a pair to evaluate their partner's comprehensibility at roughly equal intervals throughout three tasks. Over the course of the three tasks, interlocutors tended to move from mismatched ratings (i.e., one partner would provide a more positive evaluation than the other) to aligned ratings (i.e., the two partners rated each other similarly). This interactive approach has since been extended to consider how behavioral (linguistic and physical) and affective characteristics observable during interaction inform interlocutors' comprehensibility ratings (Nagle et al., 2022; Trofimovich et al., 2021). This initial wave of research serves as the impetus for Research Tasks B and C, the former of which focuses on within-interaction behavior while the latter considers how social context interacts with perceptions of comprehensibility.

3.1 Research Task B: Investigate comprehensibility in interaction through interlocutor perceptions and actions

As a subjective judgment of L2 speech, perceptions of comprehensibility are likely to exist beneath what is explicitly observable during interaction. For example, an analysis of breakdowns during communication (which would reflect a loss of intelligibility) does not inform us of the degree of effort that interlocutors exert in order to attain mutual intelligibility. Put simply, a lack of communication

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

breakdowns does not equate to highly comprehensible speech. Instead, it may reflect a particular interlocutor's ability to manage and negotiate difficult interactions (e.g., Galaczi & Taylor, 2018; Roever & Kasper, 2018). Beyond understanding what interactive and interlocutor characteristics inform the perception of comprehensibility during interaction (e.g., Nagle et al., 2022; Trofimovich et al., 2020, 2021), what should also be of interest are interlocutor adjustments to their behaviors and communication strategies to account for potentially less comprehensible speech. For our 2023 AAAL colloquium, presenter Charlie Nagle titled his talk "You're comprehensible, so what? Examining comprehensibility as a predictor of communicative outcomes". We agree with his sentiment here; beyond identifying what promotes/hinders the comprehensibility of L2 speech, we argue a need to now consider what L2 users do when they encounter less comprehensible speech. Here, we would argue that L2 comprehensibility researchers, as they push forward with a more interactive focus, build dialogue with the extensive body of literature on interactional competence (e.g., Salaberry & Kunitz, 2019).

Crowther (2020) proposed a need to consider not only interlocutors' perceptions of their speaking partner (in reference to comprehensibility), as was the case in Nagle et al. (2022) and Trofimovich et al. (2021), but to also consider these perceptions in relationship to the number and types of communication breakdowns during interaction and the ways in which such breakdowns are negotiated and resolved. Rather than recording and analyzing monologic speech, researchers can pursue interactive speech, following recent trends. While continuing to collect interlocutors' perceptions of their speaking partners' comprehensibility, analyses should also focus on the frequency and types of breakdowns that occur during interaction. By doing so, analyses can reveal whether comprehensibility ratings are associated with how frequently communication breaks down, as well as whether specific types of breakdowns (e.g., phonological, lexical, syntactic) elicit more or less positive perceptions compared to other types. The use of video recording can also allow for the consideration of nonverbal behavioral effects on comprehensibility judgments (e.g., gesture, eye gaze; see Nagle et al., 2022). Finally, the analysis of communication breakdowns, particularly those between interlocutors who may perceive the other as possessing low-comprehensibility speech, will enable an understanding of what interlocutors do to overcome associated difficulties. A primarily monologic emphasis has led to pedagogical implications that prioritize how to increase the comprehensibility of L2 speech as a means to prevent communication breakdowns. An additional pedagogical approach would be to highlight the types of communication strategies interlocutors employ that help to overcome breakdowns or strain resulting from limited comprehensibility. Such knowledge, in reference to L2 comprehensibility research, is currently limited. Triangulating across both etic (e.g., listener perceptions) and emic (e.g., interactional behavior) data would provide more in-depth knowledge on the relationship between comprehensibility and real-world interactive performance.

3.2 Research Task C: Explore a wider range of social variables

Building off of the previous task, we note how Levis (2005) made clear that attaining understanding (inclusive of both comprehensibility and intelligibility) is reliant on both speaker and listener. Roever and Kasper (2018) differentiated speaking, which they viewed as monologic responses to stimuli, from *talking*, or interactive language use that involved designing utterances for specific interlocutors and comprehending implied social actions. It is these interpersonal and social relationships that we have tended to overlook in comprehensibility research to this point. It has long been established that listeners bring with them personal biases when it comes to rating L2 speech. General bias-oriented effects, such as linguistic stereotyping and reverse linguistic stereotyping, are well attested to in the literature (e.g., Kang & Rubin, 2014; Rubin, 2012). L2 testing literature has similarly made clear the important role of addressing listener bias in speech assessment (e.g., Winke & Gass, 2013; Winke et al., 2013). Comprehensibility judgments, specifically, have been linked to less favorable emotional reactions and social judgments (e.g., Dragojevic & Giles, 2016; Lev-Ari & Keysar, 2010). What is unclear is to what extent ratings of comprehensibility are associated with social variables less

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

represented in current research. For example, O'Brien (2023) investigated the link between listeners' judgments of comprehensibility and perceived competence, with competence considered across three high prestige (doctor, lawyer, professor) and three low prestige (cleaner, server, retail worker) jobs. While O'Brien's participants, L1 Tagalog speakers of English, were deemed to be highly comprehensible, they were judged to be less competent than a native speaking comparison group, but only when the job in question was deemed to require higher skill abilities. For low ability jobs, these nonnative-English users were actually rated as being more competent than their native peers, raising interesting questions regarding the interaction between comprehensibility and occupational prestige (see also Hosoda & Stone-Romero, 2010). O'Brien's study is particularly interesting in that her findings carry high social implications, as Tagalog was the most widely spoken, non-English mother tongue in her context (Calgary, Canada), a context located within a country that uses immigration to address labor and skill shortages across the country. Flores and Rosa (2022) argued the need to move beyond universalizing a conception of linguistic competence, one generally defined by a population overrepresented as white. Might it be that our treatment of comprehensibility, or more specifically the underlying processes of judging comprehensibility, need to similarly be considered less universally? What may be more revealing is understanding how comprehensibility judgments are formed across different social contexts in which different social variables may be in play.

The task proposed here is not to upend current empirical practices. Rather, we argue a need to expand the range of listener and social variables included in our current investigations into comprehensibility. While we discuss target language and elicitation task elsewhere, here we highlight the need for research designs that include: (a) diverse listeners, (b) multiple communication contexts, or (c) both. By diverse listeners, we refer to greater consideration of those that exist beyond a university context (i.e., those who make up the listener population of the majority of comprehensibility research). Capturing listener diversity will require careful, advanced planning with deliberate recruiting efforts, as well as the use of questionnaires (or other means) of collecting information on relevant social variables. Social variables of interest should include not only those commonly reported in L2 research like age and linguistic background, but also those less reported, like race/ethnicity, occupation, socioeconomic status, political beliefs, and so forth. Though there does indeed exist some variation in communicative contexts investigated (such as O'Brien, 2023), greater expansion beyond academia is still needed. When designing research that features multiple communication contexts, researchers should target social variables like power, prestige, formality, and speaker/interlocutor characteristics like L1 background, race/ethnicity, age, and gender. Greater diversity in research designs in reference to both listeners and communicative contexts should allow for a richer understanding of how social variables operate during comprehensibility judgments. For example, the role of political views in listeners' perceptions of comprehensibility are under-researched (though work out of Montréal, Québec sheds some light in this area; see Reid et al., 2019). Researchers might distribute a background questionnaire that includes items measuring listeners' attitudes towards immigration, local/national policies on language, and so forth, to multiple populations in which a target language may or may not be perceived as favorable. A more explicit example in this sense would be to compare the perceptions of working-class populations across multiple US states where English is and is not listed as an official language. Would potential differences in political views associate with differences in the judgments of, for example, Spanish-accented English speech? If judgments of comprehensibility are socially formed, then investigations into the linguistic influences on such judgments may require a different interpretation (see Preston, 2018, for a discussion on linguistic regard).

4. Speaking task considerations

The evaluation of L2 comprehensibility requires appropriate speech elicitation practices. While Munro and Derwing (2015) highlighted the importance of "equipment quality, quiet recording environments, and post-collection processing of audio files" (p. 24), surprisingly less emphasis has been placed on the speaking tasks themselves. Here, we would argue that a greater interdisciplinary approach might

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

benefit comprehensibility research moving forward; specifically, we argue for an approach that considers task-based literature in greater detail. Within the body of comprehensibility research, a picture narration task, such as that used in Munro and Derwing (1995a), has frequently been employed, allowing for comparability with Munro and Derwing's original findings (e.g., Crowther et al., 2015b; Trofimovich & Isaacs, 2012). Though different from the picture narrative used in their 1995 study (see Munro & Derwing, 2020, for a note on their original tool), much comprehensibility research has made use of what has been affectionately dubbed the suitcase narrative, which is openly available through the IRIS Digital Repository under Derwing et al. (2009).

Crowther et al. (2015a) raised concerns regarding an overreliance on this single task and highlighted how research into task effects made clear that the complexity of a task has a direct effect on the language users produced, particularly in reference to complexity, accuracy, and fluency (e.g., Robinson, 2011; Skehan, 2009). Following earlier research linking listeners' evaluation of comprehensibility to specific characteristics of the speech stream (i.e., Trofimovich & Isaacs, 2012), Crowther et al. (2015a) compared listeners' judgments of comprehensibility for 60 speakers between two tasks: a long turn and an integrated task. Crowther et al. (2018) extended this inquiry to include a picture narrative and Crowther (2020) additionally included an interactive task. In each of these studies, findings have been relatively consistent: as the complexity of a task increased, listeners tended to make harsher evaluations of comprehensibility while simultaneously attending to a wider range of characteristics in the speech stream (e.g., listeners appear to consider lexical and grammatical production more so on more complex tasks). Research Tasks D and E aim to extend this line of inquiry.

4.1 Research Task D: Investigate targeted manipulations of task complexity

The treatment of task manipulations in comprehensibility studies, when compared with studies more fully based in task literature, is relatively simplistic. In addressing task complexity, Crowther et al. (2015a, 2018) and Crowther (2020) compared complexity across wholly distinct tasks. The demands and expected outcomes of a picture narrative differ substantially from those of either a long turn or integrated task. Such studies lacked the more targeted manipulations characteristic of much task-based research. While a full synthetic review is beyond the scope of the current paper, several key manipulations in the task literature that appear to influence learners' oral production include planning time (e.g., Foster & Skehan, 1996; Mehnert, 1998), topic familiarity (e.g., Qiu, 2020), and task repetition (e.g., Lambert et al., 2017). If such focused task manipulations can influence learners' oral production along the lines of complexity, accuracy, and fluency, all of which can inform listeners' perceptions of comprehensibility, then investigating the interaction between task complexity and comprehensibility directly seems warranted.

Whereas Crowther et al. (2015a, 2018) and Crowther (2020) compared speech across wholly distinct tasks, the extent to which more targeted task manipulation impacts comprehensibility remains underexplored. One example comes from Choi (2021), who asked users to describe five different six-frame picture narratives. Through manipulating the number of elements in each narrative and the clarity of time change across pictures in the focal story, Choi placed her five narratives on a continuum of simple to complex. Interestingly, her findings not only indicated that listeners (both native and nonnative) assessed the most complex task as most comprehensible, but their evaluations became less favorable as task complexity decreased (this pattern held true for accentedness and fluency as well). Choi argued, in line with Robinson's (2011) Cognition Hypothesis, that the higher cognitive load of the more complex version of the picture narrative pushed speakers to draw upon a wider range of linguistic resources that simultaneously allowed them to produce more comprehensible speech. Choi's findings, where a more complex version of the picture narrative elicited more comprehensible speech, contrasted with those of Crowther et al. (2015a, 2018), where the more complex but markedly different tasks elicited less comprehensible speech. As made clear by Choi, there remains significant room for further investigation of task manipulation and how it may relate to the comprehensibility of L2 speech.

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

As examples of manipulations within a specific task, we provide two potential avenues for extending the common picture narrative task. First, manipulations might consider the provision (and removal) of planning time, which has been shown to benefit L2 users in both picture- and non-picture-based tasks (O'Grady, 2019), particularly in the complexity and fluency of speech (see Ellis, 2005). As complexity and fluency of speech appear to inform listeners' judgments, manipulating planning time may lead to greater variability in listeners' evaluations of comprehensibility. Second, a more elaborate manipulation, in reference to the commonly used picture narrative, might include presenting speakers with only one image of the story at a time, requiring them to create links between pictures in real time, or, as was done in Kormos et al. (2011), provide learners with unrelated pictures, requiring them to formulate an original story on the spot. Unlike in Crowther et al. (2015a, 2018) and Crowther (2020), in Choi (2021) and the two hypothetical examples above, the core task of picture narration has not changed, even if the parameters around the task have. Understanding how users' performance, at the level of perceived comprehensibility, varies across different conditions of a task may be as informative as understanding variations in their performance between wholly different tasks. Additional manipulations, both for the picture narration as well as other common tasks used to elicit speech for comprehensibility ratings (e.g., integrated, long turn), might make use of the range of manipulable features included in Robinson's (2011) Cognition Hypothesis (e.g., number of elements, extent of perspective taking).

4.2 Research Task E: Investigate the consistency of speaker performance relative to their peers across tasks

While evidence indicates that the complexity of a task is likely to influence the comprehensibility of speech, an additional area of interest related to tasks we want to promote is the extent to which speakers may also demonstrate consistency in performance across tasks. Typically, comprehensibility research comparing tasks has focused on how certain variables, like task complexity, influence listener judgments of speaker performance. While it seems that task, in one way or another, impacts speaker performance, less emphasis has been placed on how such tasks might inform the degree to which speakers' performance might (not) vary across tasks relative to their peers. For example, may we assume that a speaker who is judged as more comprehensible on a less complex task, relative to their peers, will remain more comprehensible, relative to their peers, on a more complex task? That is, even if their overall comprehensibility, however assessed, decreases, would we still expect them to be judged as relatively more comprehensible than their peers? Little attention has been paid to such within-speaker consistency across tasks, though Isbell et al.'s (2019) study of L2 Korean and Huensch and Nagle's (2021) study of L2 Spanish provided some initial evidence that variation associated with speaker performance across tasks can meaningfully account for differences in comprehensibility ratings across tasks. In one of our most recent projects (Crowther et al., 2023), speakers completed four tasks, which were then rated for comprehensibility by listeners. While we found differences in comprehensibility ratings across tasks, we also found consistency *within* speakers across the tasks, suggesting that while it is fair to characterize comprehensibility as a co-determined outcome of an encounter between a speaker and a listener, speakers also have some qualities and/or competencies that make them generally more or less comprehensible in a range of situations.

To investigate consistencies in speaker comprehensibility and related aspects of spoken performance across tasks, researchers will need to: (a) design studies that elicit multiple spoken performances and (b) conduct and report analyses that address within-person associations. Such studies would require at least two tasks to elicit speech, but preferably more – four to ten tasks would allow for much stronger inferences about consistency across tasks. In terms of analyses, bivariate or intraclass correlations are good starting points, and with enough tasks, factor analytic approaches are also a possibility. For this research task, the same speaking tasks should be administered to all speakers, and tasks should vary according to dimensions of interest such as topic, social context, and complexity or some notion of difficulty. Variation in task characteristics could be minor or more dramatic.

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

Selecting tasks with relatively minor variations in characteristics would, in theory, reveal the potential for a speaker's observed listener-based comprehensibility scores to be similar across tasks. Selecting tasks that vary more markedly, especially in terms of complexity or expected difficulty, could lead to better understanding of how powerfully a speaker-specific comprehensibility trait might generalize across a range of communicative demands and contexts.

Including markedly different tasks would also yield understanding of how task features like complexity might serve to differentiate speaker comprehensibility. Here, we touch briefly on Weir's (2005) sociocognitive framework for validating language tests, in which performance is linked to task demands. More specifically, Weir proposed that task demands interact with a speaker's internal processes, which, in turn, influence response quality. Differences in processing capacity between learners (i.e., in reference to linguistic resources available and speech processing capabilities) should lead to differences in the quality of responses to a given task. In short, a less complex task, that may require less complex processing, may elicit less of a difference in response quality between two speakers than a more complex task, where a speaker with more efficient internal processing is more likely to outperform those with less efficient internal processing. If the perception of comprehensibility is linked to a range of linguistic features of speech, a more complex task may allow more advanced speakers to differentiate themselves from less-advanced peers more so than a less complex task. Such a view assumes that the gap between two speakers' comprehensibility is not constant across tasks, and is more likely to manifest itself as the complexity of the speech performance increases. However, emphasis on investigating listeners' perceptions across tasks as opposed to delving deeper into similarities in individual speakers' performance has left this possibility an open question. Such inquiry likely has significant implications for pedagogical practices, where we turn to next.

5. Pedagogical practices

Saito and Plonsky's (2019) meta-analysis found that while pronunciation instruction was often beneficial, findings were still in need of further refinement. Importantly, only ~30% of studies considered pronunciation gains at a global level such as comprehensibility (e.g., Derwing et al., 1998; Isbell et al., 2019; Zhang & Yuan, 2020), indicating a greater pedagogical focus on development of specific segmental/suprasegmental features. In another meta-analysis of 17 instructional studies with comprehensibility as an outcome measure, Saito (2021) found that treatment, inclusive of segmental, prosodic, and/or fluency foci, was effective in terms of comprehensibility development. Yet, of concern, as highlighted in both Saito (2012) and Saito and Plonsky (2019), is that most gains are limited to the production of controlled speech in which learners have limited demands in terms of conceptualizing and formulating their utterances. The extent to which this controlled improvement transfers over to spontaneous speech appears limited. To address this overarching concern, we propose two tasks. Research Task F proposes conducting research with practicing classroom teachers, which can help with understanding the feasibility of classroom interventions and the types of gains we might expect. Research Task G calls for more extensive longitudinal research that can place a focus on long-term improvement, as opposed to short term gains (Levis, 2018).

5.1 Research Task F: Investigate comprehensibility-oriented instruction with practicing classroom teachers

The extent to which comprehensibility-oriented teaching is implemented in L2 classrooms appears limited (Foote et al., 2011, 2016) and under-researched. While teachers in countries where English is spoken as a first language appear to agree with a focus on understandable over nativelike speech (Foote et al., 2011), their classroom practices do not always echo the types of instruction that support such a focus (Foote et al., 2016). This suggests a gap between research and classroom practice. Calls for greater dialogue between researchers and teachers to close such gaps are nothing new in L2 acquisition research (e.g., McKinley, 2019), nor for L2 pronunciation research more specifically (Derwing &

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

Munro, 2005), with a growing body of scholarship now focused specifically on L2 pronunciation teaching training and implementation (e.g., Burri & Baker, 2021). This is where we argue that greater direct collaboration between L2 comprehensibility researchers and in-service language teachers is of value. As an example, consider the second author's investigation into L2 Korean pronunciation training and the effects on comprehensibility (Isbell et al., 2019). In this study, he worked in collaboration with L2 Korean teachers (co-authors Park and Lee) to implement an instructional treatment that was informed by both research and teachers' observations. The benefits of such collaboration are that the pedagogical implications of laboratory-based studies (such as Crowther et al., 2015b) can be interpreted, adapted, and implemented in practical contexts and reappraised if needed. This sort of classroom-based research is likely to highlight contextual factors and practical constraints that teachers often cite as barriers to implementing pronunciation instruction.

As a starting point, researchers interested in bringing laboratory-based findings into the L2 classroom first need to identify a teaching context in which these findings could be applied. Given L2 teachers' sometimes hesitance to accept research into their classroom practices (e.g., Borg, 2010), we strongly encourage these researchers to work collaboratively with teachers within this context, as opposed to mandating a pedagogical approach to be implemented (see Isbell et al., 2019, for one such collaborative example). Through collaboration, a theoretically grounded but practical pedagogy for addressing comprehensibility can be pursued. Any study should adhere to common characteristics of an intervention study, such as the provision of pre- and post-treatment testing and inclusion of a control group (e.g., Derwing et al., 1998; Isbell et al., 2019). What has generally been absent from previous comprehensibility classroom studies, however, are tools traditionally associated with teacher research, such as reflection journals, classroom observations, and student surveys (Borg & Sanchez, 2015). We strongly encourage future pedagogically-oriented studies to triangulate change (or lack of change) in students' comprehensibility with both teachers' and students' experiences with a given treatment. A key objective of teacher research is to provide insight into the on-the-ground challenges of implementing change into L2 pedagogy. As such, it becomes crucial that future pedagogically-oriented comprehensibility research reports not only the effects of a given treatment, but also the experiences in developing and implementing a given treatment. Such reporting provides insight into context-grounded work that can offer essential understandings of both teaching and learning (e.g., Burns, 2010; Freeman, 1998). In addition, a triangulated understanding of how well a given treatment worked in one context can inform in what ways fellow researchers and teachers may go about implementing (and adapting) the same treatment for another. From such research, a one-size-fits-all model of L2 comprehensibility teaching is not the goal; instead, continued in-depth reporting of intervention studies across different contexts should allow for the creation of an instruction toolbox from which instructors can draw when promoting comprehensible speech in their classrooms.

5.2 Research Task G: Conduct more extensive longitudinal investigations of (instructed) comprehensibility development

Large-scale and extensive longitudinal investigations of comprehensibility are rare, with Derwing and Munro's (2013) 7-year investigation into the pronunciation development of L1 Slavic and L1 Chinese English learners in Canada a primary exception. While some recent studies have made strides in examining comprehensibility development over more extended periods of time in instructional settings (e.g., Nagle, 2018; Saito et al., 2018), we see longer-term longitudinal research as a useful approach to understanding the extent to which pronunciation instruction can truly impact comprehensibility.

As noted earlier, Saito and Plonsky (2019) found that pronunciation gains resulting from instruction were most visible in controlled speech targeting discrete pronunciation features as opposed to global constructs like comprehensibility in spontaneous speech. Given the limited duration of most studies included in their analysis, these authors argued, based on Skill Acquisition Theory (see DeKeyser, 2017), that learners may need more time and practice to proceduralize and automatize

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

the pronunciation knowledge gained from the more explicit instruction they have received. With more modest gains seen for comprehensibility as a result of instruction (with effects characterized as small in Saito, 2021), including some studies where medium-length interventions yielded unclear or no improvements (e.g., 17.5 hours in Levis et al., 2016, and 8 hours in Isbell et al., 2019), we see more and more extensive longitudinal research as critically important. Such research would yield a better understanding of how much instruction is needed to move the needle for comprehensibility, and it would also provide an opportunity to investigate the efficacy and efficiency of different instructional approaches and/or techniques.

Saito (2012), in reviewing a number of pronunciation instruction studies, found that gains in spontaneous speech occurred primarily in studies that employed focus-on-form approaches, highlighting the importance of incorporating different types of meaning-oriented speaking tasks mixed with the provision of corrective feedback. Communicatively-oriented practices for pronunciation instruction have been well documented, including Gurzynski-Weiss et al.'s (2017) special issue on task-based language teaching and pronunciation instruction (see also Mora & Levkina's, 2017, special issue epilogue). Though the target of such studies is frequently feature-based (though see Gordon, 2021), it is well established that such segmental and suprasegmental features inform listener evaluations of comprehensibility. As such, the apparent next step would be to consider to what extent such communicatively-oriented instructional practices lead to visible comprehensibility gains and, relatedly, the length and amount of practice needed for such gains to occur.

Following Task F and given the objective to investigate communicatively-oriented instructional practices, establishing a relationship with potential teachers from the outset of any longitudinal study is key. In designing a study, several variables are of primary interest. Most notably would, of course, be length of treatment (e.g., one month vs. one semester vs. one year). Following students across multiple semesters poses challenges, as they transition from one class to another, further highlighting the need for researcher-teacher collaboration. Length of treatment, however, should not be seen as the sole predictor of pedagogical effectiveness. As already discussed, it may be that students need substantial practice to incorporate gains in explicit pronunciation knowledge/ability into their spontaneous use of the language. The number of hours of practice within a given treatment should also be documented, as well as ways in which this practice is distributed (e.g., blocked vs. interleaved; see Suzuki, 2021). Types of communicative tasks employed, frequency with which these tasks are repeated throughout training, and extent to which these tasks reflect those used for pre- and post-testing are additionally of interest. Importantly, as made clear in Derwing and Munro (2013), a means of tracking students' out-of-class language contact and use is necessary, as differences outside of class may impact in-class learning. In this sense, investigating classroom gains in second vs. foreign language contacts continues to be of interest.

6. Replication in L2 comprehensibility scholarship

Earlier, we discussed the usefulness of replication to grow beyond the English-centric focus of much comprehensibility research up until now. Though less prevalent in applied linguistics than desired (Marsden et al., 2018), we have seen a positive trend in recent years towards how replication is viewed in the field (McManus, 2022). One reason put forth for the lack of replication research in general has been that such research may be seen as lacking originality and/or innovation (see Porte, 2012, and Porte & McManus, 2019, for in-depth discussions on replication research). However, and as evident from Huensch and Nagle's (2021) study discussed earlier, replication research can also serve to consolidate and strengthen our knowledge base (Marsden et al., 2018; Porte, 2012; see also Nagle & Hiver, 2023, for a study specific to L2 pronunciation research). That replication of previous L2 comprehensibility is limited beyond a handful of studies, including Huensch and Nagle (2021, of Munro & Derwing, 1995a), Zhang and Yuan (2020, of Derwing et al., 1998), and Isbell and Lee (2022, of Trofimovich et al., 2016), may not be surprising given that such practices are still gaining acceptance in the field. Yet knowing that replication can both support and extend what we already know regarding

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

the construct of comprehensibility should be seen as a key scholarly direction moving forward. And importantly, many journals known for publishing comprehensibility research, such as *Language Learning* and *Studies in Second Language Acquisition*, now dedicate sections specifically to replication, helping to alleviate concerns regarding perceived lack of originality/innovation as a barricade to publication.

7. Conclusion

The approach to L2 speech comprehensibility established by Munro and Derwing (1995a) has persisted to the present day, with a sharp increase in scholarly attention post-2010 (Crowther et al., 2022). This recent body of research has considered both original (e.g., Crowther et al., 2015a; Trofimovich & Isaacs, 2012) and metaanalytic work (Saito, 2021), though with limited replication of earlier studies (e.g., Huensch & Nagle, 2021). Building off of existing empirical trends and methodological practices, we proposed several themes for future research that we believe will help advance comprehensibility research over the next 10–15 years. These themes include suggestions that we believe will push us beyond what might be viewed as well tread ground: considering LOTEs and more thoroughly exploring speaker and listener populations beyond the university context. We additionally advocated for greater interdisciplinary consideration in comprehensibility and identified several scholarly disciplines that we believe can greatly inform future directions in comprehensibility research, including task-based literature. The expanded scope of comprehensibility research we have sketched out, especially when addressing more than one research task in tandem, will likely require more elaborate research designs and sophisticated analyses. We see this as a sign that L2 comprehensibility research is maturing, and also as an opportunity for the research community to pool resources and collaborate, such as multi-site data collection.

We note that the themes included here are not exhaustive. Indeed, other avenues for L2 comprehensibility research, such as the role of individual differences (e.g., Saito et al., 2020) and the place of comprehensibility within language assessment practices (e.g., Isaacs & Harding, 2017; Isaacs et al., 2018), are likely to produce valuable insights. Although it is difficult to predict how comprehensibility research will grow and change over the next 10–15 years, what seems certain is that as interest in L2 pronunciation (and L2 speaking more generally) continues to move beyond native norms towards greater emphasis on broad intelligibility (or different forms of understanding), interest in L2 speech comprehensibility is likely to persist, with scholarly inquiry holding implications at both theoretical and pedagogical levels.

Competing interest. The authors declare none.

References

Borg, S. (2010). Language teacher research engagement. *Language Teaching*, *43*(4), 391–429. doi:10.1017/S0261444810000170

Borg, S., & Sanchez, H. S. (2015). Key issues in doing and supporting language teacher research. In S. Borg & H. S. Sanchez (Eds.), *International perspectives on teacher research* (pp. 1–13). Palgrave Macmillan.

Burns, A. (2010). *Doing action research in English language teaching: A guide for practitioners*. Routledge.

Burri, M., & Baker, A. (2021). 'I Feel… Slightly out of Touch': A longitudinal study of teachers learning to teach English pronunciation over a six-year period. *Applied Linguistics*, *42*(4), 791–809. doi:10.1093/applin/amab009

Choi, J. S. (2021). The effects of task complexity on comprehensibility in second language speech. *Applied Language Learning*, *31*(1&2), 50–70.

Crowther, D. (2020). Rating L2 speaker comprehensibility on monologic vs. interactive tasks: What is the effect of speaking task type? *Journal of Second Language Pronunciation*, *6*(1), 96–121. doi:10.1075/jslp.19019.cro

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

Crowther, D., Isbell, D. R., & Nishizawa, H. (2023). What characterizes 'acceptable' academic English speech? Examining linguistic/temporal influences on university stakeholders' judgments across tasks. *Applied Psycholinguistics*. Published online 11 August 2023.

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015a). Does a speaking task affect second language comprehensibility? *Modern Language Journal*, *99*(1), 80–95. doi:10.1111/modl.12185

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

Crowther, D., Trofimovich, P., Saito, K., & Isaacs, T. (2015b). Second language comprehensibility revisited: Investigating the effects of learner background. *TESOL Quarterly*, *49*(4), 814–837. doi:10.1002/tesq.203

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. doi:10.1017/S027226311700016X

DeKeyser, R. M. (2017). Knowledge and skill in ISLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of second language acquisition* (pp. 15–32). Routledge.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379–397. doi:10.2307/3588486

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. doi:10.1111/lang.12000

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., & Rossiter, M. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(4), 533–557. doi:10.1017/S0272263109990015

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*(3), 393–410. doi:10.1111/0023-8333.00047

Dragojevic, M., & Giles, H. (2016). I don't like you because you're hard to understand: The role of processing fluency in the language attitudes process. *Human Communication Research*, *42*(3), 396–420. doi:10.1111/hcre.12079

Ellis, R. (2005). Planning and task-based performance: Theory and research. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 3–34). John Benjamins.

Flores, N., & Rosa, J. (2022). Undoing competence: Coloniality, homogeneity, and the overrepresentation of whiteness in applied linguistics. *Language Learning*. Published online 31 October 2022. doi:10.1111/lang.12528

Foote, J. A., Holtby, A. K., & Derwing, T. M. (2011). Survey of the teaching of pronunciation in adult ESL programs in Canada, 2010. *TESL Canada Journal*, *29*(1), 1–22. doi:10.18806/tesl.v29i1.1086

Foote, J. A., Trofimovich, P., Collins, L., & Urzúa, F. S. (2016). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal*, *44*(2), 181–196. doi:10.1080/09571736.2013.784345

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*(3), 299–323. doi:10.1017/S0272263100015047

Freeman, D. (1998). *Doing teacher research: From inquiry to understanding*. Heinle & Heinle.

Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, *15*(3), 219–236. doi:10.1080/15434303.2018.1453816

Gordon, J. (2021). Pronunciation and task-based instruction: Effects of a classroom intervention. *RELC Journal*, *52*(1), 94–209. doi:10.1177/0033688220986919

Gurzynski-Weiss, L., Long, A., & Solon, M. (2017). TBLT and L2 pronunciation: Do the benefits of tasks extend beyond grammar and lexis? *Studies in Second Language Acquisition*, *39*(2), 213–224. doi:10.1017/S0272263117000080

Hosoda, M., & Stone-Romero, E. (2010). The effects of foreign accents on employment-related decisions. *Journal of Managerial Psychology*, *25*(2), 113–132. doi:10.1108/02683941011019339

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. doi:10.1111/lang.12451

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

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. doi:10.1080/15434303.2013.769545

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. doi:10.1177/0265532217703433

Isbell, D. R., & Lee, J. (2022). Self-assessment of comprehensibility and accentedness in second language Korean. *Language Learning*, *72*(3), 806–852. doi:10.1111/lang.12497

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. doi:10.1075/jslp.17010.isb

Kang, O., & Rubin, D. (2014). Listener expectations, reverse linguistic stereotyping, and individual background factors in social judgments and oral performance assessment. In J. M. Levis & A. Moyer (Eds.), *Social dynamics in second language accent* (pp. 239–254). De Gruyter Mouton.

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. doi:10.1111/lang.12270

Kormos, J., Trebits, A., & Robinson, P. (2011). Working memory capacity and narrative task performance. In P. Robinson (Ed.), *Second language task complexity: Rethinking the cognition hypothesis of language learning and performance* (pp. 257–285). John Benjamins.

Lambert, C., Kormos, J., & Minn, D. (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, *39*(1), 167–196. doi:10.1017/S0272263116000085

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. doi:10.1016/j.jesp.2010.05.025

Levis, J. (2020a). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation*, *6*(3), 310–328. doi:10.1075/jslp.20050.lev

Levis, J. (2020b). Changes in L2 pronunciation: 25 years of intelligibility, comprehensibility, and accentedness. *Journal of Second Language Pronunciation*, *6*(3), 283–309. doi:10.1075/jslp.20054.lev

Levis, J. (2021). L2 pronunciation research and teaching: The importance of many languages. *Journal of Second Language Pronunciation*, *7*(20), 141–153. doi:10.1075/jslp.21037.lev

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

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation*. Cambridge.

Levis, J. M., & Silpachai, A. O. (2021). Speech intelligibility. In T. M. Derwing, M. J. Munro, & R. I. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking* (pp. 160–173). Routledge.

Levis, J. M., Sonsaat, S., Link, S., & Barriuso, T. A. (2016). Native and nonnative teachers of L2 pronunciation: Effects on learner performance. *TESOL Quarterly*, *50*(4), 894–931. doi:10.1002/tesq.272

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. doi:10.1111/lang.12286

McKinley, J. (2019). Evolving the TESOL teaching–research nexus. *TESOL Quarterly*, *53*(3), 875–884. doi:10.1002/tesq.509

McManus, K. (2022). Are replication studies infrequent because of negative attitudes? Insights from a survey of attitudes and practices in second language research. *Studies in Second Language Acquisition*, *44*(5), 1410–1423. doi:10.1017/S0272263121000838

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*(1), 83–108. doi:10.1017/S0272263198001041

Mora, J., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition*, *39*(2), 381–399. doi:10.1017/S0272263117000183

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. doi:10.1111/j.1467-1770.1995.tb00963.x

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. doi:10.1177/002383099503800305

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

Munro, M. J., & Derwing, T. M. (2015). A prospectus for pronunciation research in the 21st century: A point of view. *Journal of Second Language Pronunciation*, *1*(1), 11–42. doi:10.1075/jslp.1.1.01mun

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

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. doi:10.1111/modl.12461

Nagle, C., & Hiver, P. (2023). Optimizing second language pronunciation instruction: Replications of Martin and Sippel (2021), Olson and Offerman (2021), and Thomson (2012). *Language Teaching*. Published online 12 April 2023. doi:10.1017/S0261444823000083

Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*, *41*(4), 647–672. doi:10.1017/S0272263119000044

Nagle, C. L., Trofimovich, P., O'Brien, M. G., & Kennedy, S. (2022). Beyond linguistic features: Exploring behavioral and affective correlates of comprehensible second language speech. *Studies in Second Language Acquisition*, *44*(1), 255–270. doi:10.1017/S0272263121000073

O'Brien, M. G. (2016). Methodological choices in rating speech samples. *Studies in Second Language Acquisition*, *38*(3), 587–605. doi:10.1017/S0272263115000418

O'Brien, M. G. (2023). *Comprehensibility and client interactions*. American Association for Applied Linguistics. https://issuu.com/aaal_applied_linguistics/docs/aaal_2023_pdf_1

O'Grady, S. (2019). The impact of pre-task planning on speaking test performance for English-medium university admission. *Language Testing*, *36*(4), 505–526. doi:10.1177/0265532219826604

Ortega, L. (2019). SLA and the study of equitable multilingualism. *Modern Language Journal*, *103*(S1), 23–38. doi:10.1111/modl.12525

Porte, G., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.

Porte, G. K. (2012). *Replication research in applied linguistics*. Cambridge University Press.

Preston, D. (2018). Language regard: What, why, how, whither? In B. E. Evans, E. J. Benson, & J. N. Stanford (Eds.), *Language regard: Method, variation, and change* (pp. 3–28). Cambridge.

Qiu, X. (2020). Functions of oral monologic tasks: Effects of topic familiarity on L2 speaking performance. *Language Teaching Research*, *24*(6), 745–764. doi:10.1177/1362168819829021

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

Reid, K. T., Trofimovich, P., & O'Brien, M. G. (2019). Social attidues 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. doi:10.1017/S0272263118000244

Robinson, P. (2011). Second language task complexity, the cognition hypothesis, language learning, and performance. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 3–37). John Benjamins.

Roever, C., & Kasper, G. (2018). Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Language Testing*, *35*(3), 331–355. doi:10.1177/0265532218758128

Rubin, D. (2012). The power of prejudice in accent perception: Reverse linguistic stereotyping and its impact on listener judgments and decisions. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference*, Sept. 2011. (pp. 11–17). Iowa State University.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, *46*(4), 842–854. doi:10.1002/tesq.67

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, *55*(3), 866–900. doi:10.1002/tesq.3027

Saito, K., & Akiyama, Y. (2017). Linguistic correlates of comprehensibility in second language Japanese speech. *Journal of Second Language Pronunciation*, *3*(2), 199–217. doi:10.1075/jslp.3.2.02sai

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. doi:10.1111/lang.12297

Saito, K., Macmillan, K., Mai, T., Suzukida, Y., Sun, H., Magne, V., Ilkan, M., & Murakami, A. (2020). Developing, analyzing and sharing multivariate datasets: Individual differences in L2 learning revisited. *Annual Review of Applied Linguistics*, *40*, 9–25. doi:10.1017/S0267190520000045

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652–708. doi:10.1111/lang.12345

Salaberry, M. R., & Kunitz, S. (Eds.). (2019). *Teaching and testing L2 interactional competence: Bridging theory and practice*. Routledge.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532. doi:10.1093/applin/amp047

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

Suzuki, Y. (2021). Optimizing fluency training for speaking skills transfer: Comparing the effects of blocked and interleaved task repetition. *Language Learning*, *71*(2), 285–325. doi:10.1111/lang.12433

Thomson, R. I. (2018). Measurement of accentedness, intelligibility and comprehensibility. In O. Kang & A. Ginther (Eds.), *Assessment in second language pronunciation* (pp. 11–29). Routledge.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905–916. doi:10.1017/S1366728912000168

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, *19*(1), 122–140. doi:10.1017/S1366728914000832

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. doi:10.1075/bct.121.08tro

Trofimovich, P., Tekin, O., & McDonough, K. (2021). Task engagement and comprehensibility in interaction: Moving from what second language speakers say to what they do. *Journal of Second Language Pronunciation*, *7*(3), 435–461. doi:10.1075/jslp.21006.tro

Ushioda, E. (2017). The impact of global English on motivation to learn other languages: Toward an ideal multilingual self. *Modern Language Journal*, *101*(3), 469–482. doi:10.1111/modl.12413

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

Vujinović, S. (2017). *How do three immigrant women from former Yugoslavia perceive the role of accent in intelligibility and comprehensibility in the Canadian workplace?* [Unpublished manuscript]. York University, Toronto, ON. Retrieved from http://hdl.handle.net/10315/34196

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave Macmillan.

Winke, P., & Gass, S. (2013). The influence of second language experience and accent familiarity on oral proficiency rating: A qualitative investigation. *TESOL Quarterly*, *47*(4), 762–789. doi:10.1002/tesq.73

Winke, P., Gass, S., & Myford, C. (2013). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, *30*(2), 231–252. doi:10.1177/0265532212456968

Zhang, R., & Yuan, Z.–m. (2020). Examining the effects of explicit pronunciation instruction on the development of L2 pronunciation. *Studies in Second Language Acquisition*, *42*(4), 905–918. doi:10.1017/S0272263120000121

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press

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 R. Isbell (Ph.D. Second Language Studies, Michigan State University) is an Assistant Professor in the Department of Second Language Studies at the University of Hawai`i at Mānoa, where he teaches courses in language assessment and quantitative research methods. In addition to assessment, his research interests include L2 speaking and technology in language learning. His work has been published in journals such as *Applied Linguistics*, *Modern Language Journal*, *Language Assessment Quarterly*, *Language Testing*, *Language Learning*, and *Studies in Second Language Acquisition*.

Cite this article: Crowther, D., & Isbell, D. R. (2024). Second language speech comprehensibility: A research agenda. *Language Teaching*, *57*(4), 533–549. https://doi.org/10.1017/S026144482300037X

https://doi.org/10.1017/S026144482300037X Published online by Cambridge University Press