L2 LEARNERS' SELF-ASSESSMENT OF COMPREHENSIBILITY AND ACCENTEDNESS: OVER/UNDER-ESTIMATION, EFFECTS OF RATING PEERS, AND ATTENTION TO SPEECH FEATURES

Mireia Ortega, Universitat de Barcelona
Joan C. Mora, Universitat de Barcelona
Ingrid Mora-Plaza, Universitat de Barcelona

Comprehensibility and accentedness of L2 speech are often assessed through native listeners' ratings, but there is little research on learners' speech self-assessments. This study investigates the extent to which learners' self-assessment matched native listeners' evaluations and whether their self-assessments were influenced by having previously assessed the performance of peers. In addition, we asked learners to identify the speech features they related to comprehensibility and accentedness when self-assessing their speech. Advanced L2 English learners (*N*=56) performed a picture-description oral narrative task, which they then self-assessed for comprehensibility and accentedness under two conditions: having previously evaluated 20 speech samples of peers (*N*=24) or not (*N*=32). Native English listeners (*N*=14) assessed the 56 learners' narratives for the same dimensions. Results indicated that learners self-assessed their speech inaccurately for comprehensibility and accentedness, by either overestimating or underestimating their own speech, in accordance with previous research on comprehensibility (Trofimovich et al., 2016). Nevertheless, previous rating experience did not lead to more accurate self-assessments of comprehensibility and accentedness, suggesting that extended practice in the assessment of comprehensibility may be necessary for learners to be able to calibrate their speech self-assessments. When assessing their speech for comprehensibility, learners reported paying attention to pronunciation and the story plot rather than grammar or fluency, whereas for accentedness they focused on segmentals and overall accent rather than suprasegmentals.

Cite as: Ortega, M., Mora, J. C., & Mora-Plaza, I. (2022). L2 learners' self-assessment of comprehensibility and accentedness: Over/under-estimation, effects of rating peers, and attention to speech features. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. https://doi.org/10.31274/psllt.13354.

INTRODUCTION

Learning the phonology of a second language (L2) is an arduous task, since learners' L1 phonological categories are already well-established and L1-based perception hinders attunement to the phonetic features and distributional properties of L2 sounds (Kuhl et al., 2008). Consequently, few learners attain native-like pronunciation in the L2 (Abrahamson & Hyltenstam, 2009) and most speak the L2 with various degrees of foreign accent. Although some L2 learners might wish to acquire native-like pronunciation in their L2 (Moyer, 2014), most would prefer to attain the more realistic goal of comprehensible speech. Comprehensibility refers to listeners' perceptions of how easy or difficult L2 speech is to understand, whereas accentedness is related to listeners' perceptions of how close L2 learners' speech is to that of native speakers. These perceptual constructs of L2 speech are related to one another but also partially independent, as heavily accented speech may still be highly comprehensible to listeners (Munro & Derwing, 1995, 2020), and some underlying linguistic features are common to both (e.g., pronunciation accuracy) while others (e.g., lexical richness and accuracy) would influence comprehensibility but not accentedness (Trofimovich & Isaacs, 2012). Over the past

20 years, comprehensibility has been favoured over nativeness as a teaching goal for L2 speech development (Levis, 2005, 2020; Saito et al., 2017). Consequently, much L2 speech research has been devoted to identifying those pronunciation features that underlie successful comprehensibility regardless of accentedness (Isaacs & Trofimovich, 2012), demonstrating that comprehensibility is a multidimensional construct, linked to pronunciation, speaking fluency, and grammatical and lexical accuracy and richness, whereas accentedness is mainly determined by segmental and suprasegmental production accuracy.

Most studies to date have examined comprehensibility and accentedness through native listeners' (NL) ratings. However, fewer studies have investigated L2 learners' perceptions of their own speech or the speech of other L2 learners. This would be valuable knowledge, as in general, self-assessment ability is related to L2 performance (Li & Zhang, 2021) and inaccurate speaking self-assessments could indicate that learners are not able to notice the pronunciation, lexical and grammatical errors affecting the comprehensibility of their speech and the speech of others, which could have a negative impact on their L2 pronunciation development (O'Brien, 2019; Wrembel, 2015). In addition, assessment of comprehensibility of L2 speech by listeners of diverse L1 backgrounds would be useful in determining a comprehensibility benchmark in communicative interaction among L2 speakers of various L1s.

L2 Speech Self-Assessment

The few existing studies on L2 speech self-assessment have shown a mismatch between self-assessments and NLs' assessments on comprehensibility and accentedness (Strachan et al., 2019), and some of the findings are consistent with the Dunning-Kruger effect (Johnson et al., 2003), whereby learners whose speaking performance is less proficient (less comprehensible and more strongly accented) tend to overestimate their speech quality, whereas those whose speech is more proficient tend to underestimate it (Trofimovich et al., 2016; Saito et al., 2020). This suggests that learners tend to misjudge their own speech and may be unable to identify the linguistic features that influence their comprehensibility (Strachan et al., 2019). This means they are unaware of what they need to focus on to improve the comprehensibility of their speech. The magnitude of the misalignment between learners' self-assessments and ratings by others, however, has been shown to depend on the dimension being rated and various other factors. For example, for comprehensibility, the linguistic features that underlie this dimension make a difference, as misalignments are stronger in relation to segmental accuracy, word stress, rhythm, intonation and speech rate than in relation to lexical, grammatical or discourse aspects (Trofimovich et al., 2016). L2 learners' proficiency levels, including pronunciation accuracy, also impacts self-assessments and how they might differ from others' ratings. For example, in Saito et al. (2020), more comprehensible speakers, who had initially underestimated their performance, were more successful at aligning their self-assessments with the perceptions of others over time than less comprehensible speakers, who had initially overestimated their comprehensibility. For accentedness, learners generally perceive less of an accent in their speech than in the speech of others (Mitterer et al., 2020), as this is the L2 voice they are more frequently exposed to and the accent they find more intelligible. This cognitive bias would add to the negative impact of the Dunning-Kruger effect on L2 acquisition, as L2 learners' failure to notice in their speech the accent they perceive in other learners suggests they are unaware of errors they would need to notice to be able to improve their pronunciation. Self-assessment practice might be able to enhance learners' ability to notice their own pronunciation errors.

Training in self-assessment has been found to help L2 learners develop awareness of their speech comprehensibility, as shown by stronger correlations between self- and other-assessments (Kissling & O'Donnell, 2015). Methods of improving self-assessment skills include discussing learners' own performance, familiarizing learners with rating criteria, the use of self-testing exercises, benchmarking, and peer assessment (Dunning et al., 2004). However, little research to date has investigated the effectiveness of these methods. Tsunemoto et al. (2021) implemented a treatment based on benchmarking (asking learners to discuss speech evaluation criteria) and peer assessment (evaluating the speaking performance of peers) to raise L2 learners' awareness of their own comprehensibility. They found the treatment to effectively reduce the misalignment between self-assessments of comprehensibility and NLs' ratings, suggesting L2 learners' increased awareness of their own pronunciation.

A related, under-researched question concerning speech self-assessment is the extent to which L2 learners are able to accurately identify the linguistic features that previous research (e.g., Isaacs & Trofimovich, 2012; Saito et al., 2016, 2017; Trofimovich & Isaacs, 2012) has shown to differentially underlie the perceptual dimensions of comprehensibility (pronunciation accuracy, speaking fluency, lexical and grammatical accuracy and richness) and accentedness (pronunciation accuracy at the segmental and prosodic levels and speaking fluency). Whether learners are able to pay attention to those features that would allow them to improve comprehensibility and accentedness also remains an under-researched question. For example, Strachan, et al. (2019) asked L2 learners to reflect on what they would do to make their speech more comprehensible to listeners. The L2 learners mentioned pronunciation and lexical and grammatical features, but awareness of such features varied greatly across learners, while content and organization were consistently mentioned.

The present study follows this line of research by examining L2 learners' self-assessments and NLs' ratings of L2 speech for both comprehensibility and accentedness. More specifically, we investigate the impact of peer assessment on the alignment between L2 learners' self-assessments and NLs' ratings (i.e., the Dunning-Kruger effect). In addition, we provide qualitative data by examining which speech features learners focus on when performing speech self-assessments.

Research Questions

The present study has three aims. We first investigated the extent to which L2 learners' self-assessments of comprehensibility and accentedness matched NLs' ratings. Secondly, we examined how learners' speech self-assessments were influenced by having previously evaluated the performance of peers on the same task for the same dimensions. Finally, we identified the speech features learners linked to comprehensibility and accentedness when self-assessing their speech. This study was guided by three research questions (RQ):

RQ1. Do L2 learners' self-assessments of comprehensibility and accentedness match those of NLs?

RQ2. Does peer assessment influence learners' alignment of self-assessments and NLs' ratings?

RQ3. What are the linguistic features learners linked to comprehensibility and accentedness when self-assessing their speech?

METHODS

Participants

Fifty-six Catalan-Spanish bilingual adult EFL undergraduate learners of English participated in the present study for course credit (see Table 1 for demographics). They belonged to two intact English phonetics and phonology classes, each of which was assigned to either an experimental or a control group. While participants in the control group (G1, *N*=32) only evaluated their own speech, participants in the experimental group (G2, *N*=24) did so after having evaluated 20 speech samples from peers performing the same task. Participant groups were comparable in terms of L2 learning history and proficiency as measured by X/Y Lex vocabulary size test (Table 1).

Table 1

Participants' demographics

	G1		G2	
Measure	*M*	*SD*	*M*	*SD*
Age at testing (years)	20.0	3.0	19.0	1.9
Age of onset of L2 learning (years)	6.6	2.6	5.5	1.5
L2 instruction (years)	13.2	0.4	13.3	0.4
Spoken L2 input / output (h/week)	14.1/ 6.0	17.5/7.2	24.8/3.4	27.9/5.9
L2 use [a] (h/week)	3.3	1.2	2.7	1.1
Vocabulary size (0-10,000 words) [b]	5845.3	961.2	5945.8	1305.8
Self-estimated proficiency (1=very poor-7=native-like) [c]	4.9	0.9	4.8	1.0

[a] L2 use with native and non-native speakers in hours per week.
[b] Obtained through the X/Y Lex (Meara & Milton, 2003; Meara & Miralpeix, 2006).
[c] Averaged self-estimated ability to speak spontaneously, understand, read, write and pronounce English.

Expert native English listeners (*N*=14; 6 males, 8 females) with a mean age of 33.5 (*SD*= 8.03) evaluated the learners' oral narratives individually for both comprehensibility and accentedness. All listeners were experienced EFL teachers born and raised in monolingual English families, and spoke either British (86%) or American (14%) English varieties. They were living in Barcelona at the time of data collection and reported being very familiar with Spanish/Catalan-accented English (*M*=8.15, *SD*= 1.67) on a 9-point Likert scale (1 = "not familiar at all"; 9 = "very familiar"). Finally, all NLs reported conversational knowledge of either Spanish or Catalan.

Materials

First, a language background questionnaire and the X/Y Lex vocabulary size test were administered to gather learners' information on their L2 learning history and English proficiency level. A picture-based story telling task, the suitcase story (Derwing et al., 2009) – a task previously used in similar studies (e.g. Trofimovich & Isaacs, 2012; Saito et al., 2016; Saito et al., 2017) – was used to elicit an oral narrative.

The participants carried out a comprehensibility and an accentedness self-rating task, where they had to self-assess their own speech using nine-point scales after listening twice to their full speaking performance. In this task, learners were first given a definition of comprehensibility and then asked to rate their own speech sample (1 = "very difficult to understand"; 9 = "very easy to understand). The same procedure was used to obtain accentedness self-assessments (1 = "no foreign accent"; 9 = "very strong accent") (see Appendix). Immediately after rating their own speech, they were prompted to qualitatively describe the speech dimensions they had paid attention to ("Please explain what you were focusing on when rating your oral narrative"). The fourteen NLs performed a similar rating task over headphones in Praat (Boersma and Weenink, 2020) to evaluate the L2 learners' oral productions. In this task, there were three speech samples for practice and fifty-six speech samples of approximately one minute each that listeners rated twice, first for comprehensibility and then for accentedness. Listeners could play the speech samples up to three times.

Procedure

Learners filled out an online language background questionnaire at home before coming to the lab, where they performed the vocabulary size test and the picture description task. One week later, they were sent their own recordings and asked to evaluate their own speech sample for comprehensibility and accentedness. Participants in G2 were asked to rate twenty speech samples from peers immediately before assessing their own recordings. Participants in G1 completed all tasks in 45 minutes, whereas those in G2 took 60 minutes. Native English listeners then evaluated the learners' full productions for comprehensibility and accentedness through the same interface (60-90 min.).

Data analysis

Inter-rater reliability (Cronbach's alpha intra-class correlation coefficients) of NLs' ratings was high for comprehensibility (α = .93) and accentedness (α = .95), so we computed average comprehensibility and accentedness ratings across listeners for each learner.

Following Trofimovich et al. (2016) and Saito et al. (2020), comprehensibility and accentedness overconfidence scores were calculated by subtracting, for each learner, the mean NLs' rating (1-9) from the learners' self-assessment (1-9). In this way, positive scores represented self-assessments that overestimated performance (e.g. self-assessment 7 minus NLs' rating 5.5 = +1.5 overconfidence score), whereas negative scores corresponded to self-assessments that underestimated performance, and scores around zero indicated alignment between self-assessments and NLs' ratings. In addition, we computed a proportion measure of alignment of self-assessments with NLs' ratings (0-1) independent of under- and over-estimation where 0 represented complete alignment (identical learner's and NLs' ratings) and 1 complete misalignment (i.e. maximum difference of 8 between learner's and NLs' ratings).

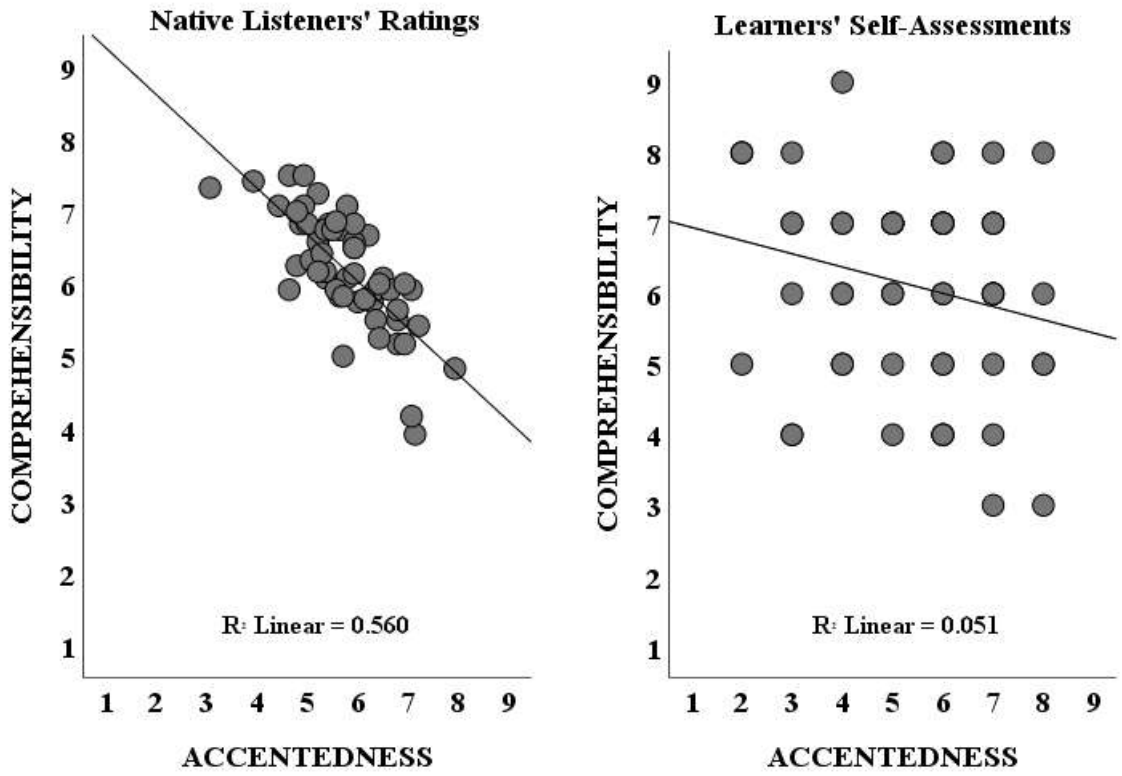
Qualitative analyses for the third research question were based on the eleven linguistic correlates of comprehensibility and accentedness in Saito et al. (2017): segmental errors, word stress, intonation, rhythm, speech rate, lexical appropriateness, lexical richness, grammatical accuracy, grammatical complexity, story richness and story cohesion, which we grouped under five broad categories in our analyses (pronunciation, fluency, vocabulary, grammar, and the way the story is told).

RESULTS

As expected, *Pearson's-r* correlations between comprehensibility and accentedness assessments conducted by native listeners were significantly and negatively related (r (54)= -.748, p < .001), suggesting that the speech samples NLs found easier to understand were also the least accented. This relationship was non-significant for learners' self-assessments (*Spearman Rho*: r_s (54) = -.210, p = .121), suggesting that these perceptual dimensions are not inversely related to one another for L2 learners, though one would expect them to understand their own speech perfectly irrespective of the strength of their Spanish/Catalan accent (Figure 1).

Figure 1

Scatterplot of comprehensibility and accentedness ratings for NLs' ratings (left) and learners' self-assessments (right).



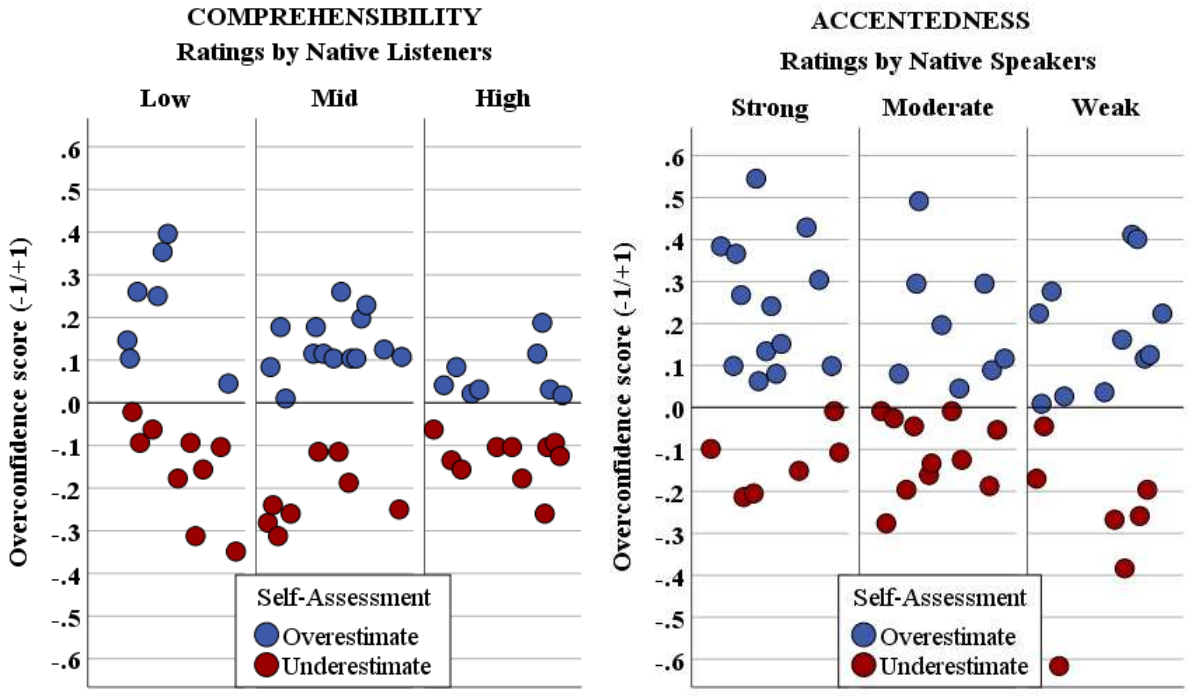
Research question 1 examined the overall relationship between learners' self-assessments and NLs' ratings of comprehensibility and accentedness. The results indicated, in line with previous research, that learners' self-assessments and NLs' ratings matched only in one case. More specifically, statistical analyses showed that *Spearman's rho* correlations revealed a moderate association between self-assessments and NLs' evaluations for comprehensibility (r_s(54)= .40, p=.002), but no association for accentedness (r_s(54) =.19, p =.147). This suggests that learners' and NLs' assessments shared a significant 16% of the variance when rating comprehensibility, but only 3.6% of the variance when rating accentedness.

As regards the relationship between learners' overconfidence scores and NLs' evaluations of comprehensibility, a negative weak association emerged (*Pearson's r* (54) = -.27, p= .04). The more comprehensible learners were judged to be by NLs, the lower their overconfidence score, that is, the

more they underestimated the comprehensibility of their speech (Figure 2, left). A similar picture emerged for accentedness. For this construct, we reversed the scale to make the ratings and scores comparable to those for comprehensibility. Overconfidence scores were negatively associated with accentedness ratings of the native listeners (r (54) = -.41, p= .002). The association was not strong, but it indicated that learners whose speech was more strongly accented (and therefore less native-like) tended to overestimate the degree of accent in their speech, whereas those whose speech was less strongly accented tended to under-rate themselves (Figure 2, right).

Figure 2

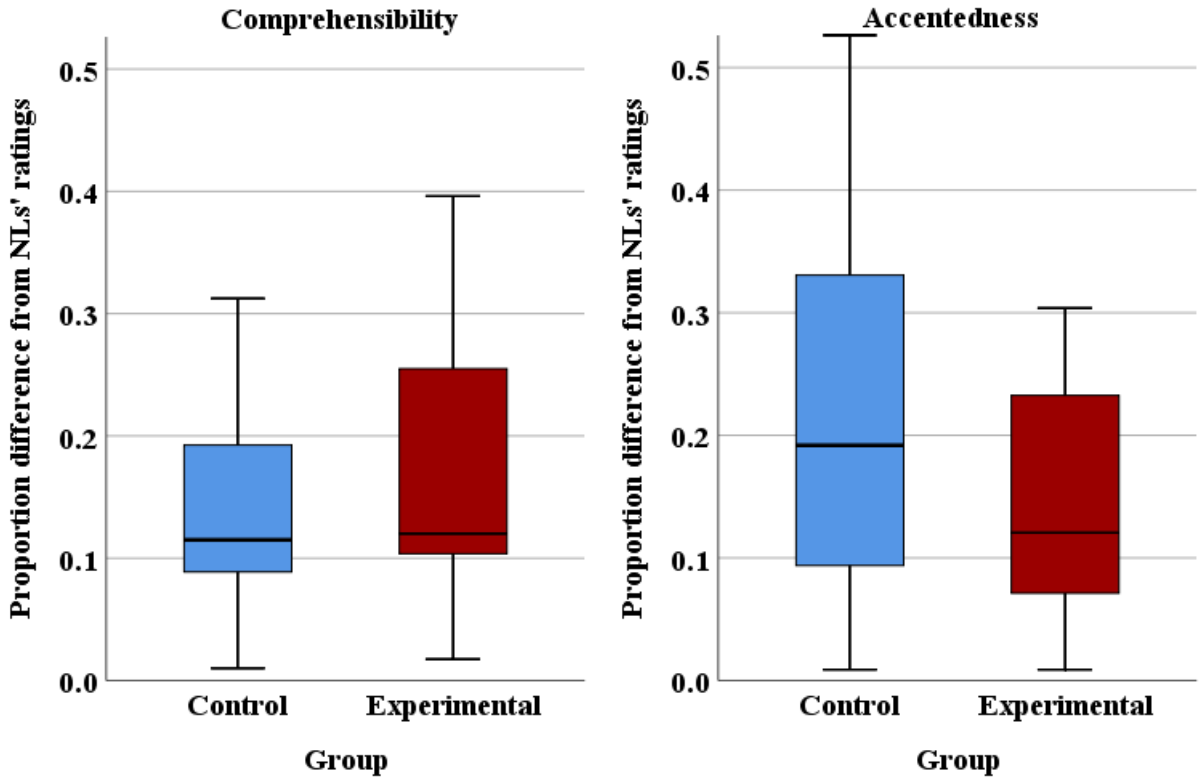
Scatterplot between overconfidence scores from self-assessment ratings and NLs' ratings for comprehensibility (left) and accentedness (right).



Research question 2 assessed whether learners' self-assessments were affected by having previously evaluated their peers' performance. Comprehensibility self-assessments (Figure 3, left) did not align better with NLs ratings for participants who had engaged in peer assessment (G2, experimental) than for those (G1, control) who had not (U = 436, p = .384). Descriptively, accentedness self-assessments seemed to align with NLs' ratings for G2 (Figure 3, right), but the magnitude of the difference between the participants' and NLs' assessment did not reach significance (U = 309, p = .214).

Figure 3

Alignment between learners' self-assessments and NLs' ratings (0=complete alignment).



Our third research question asked about the linguistic dimensions learners reported paying attention to when rating their own speech. For comprehensibility, G1 (self-assessment only) participants identified overall pronunciation (29%) and how well the story had been told (25%) as the main dimensions they focused on when assessing their own speech samples. This was followed by vocabulary (15%) and ease of understanding in general (17%); grammar and fluency were referred to only occasionally. G2 participants (assessment of others before self-assessment) also reported pronunciation (30%) and ease of understanding (30%) as the main dimensions they were focusing on while rating their own speech for comprehensibility. This was followed by fluency (17%) and vocabulary (13%). However, the way the story had been told was mentioned occasionally (7%). Irrespective of the group, all learners viewed the comprehensibility of their own speech as depending to a large extent on their pronunciation and their ability to explain a coherent story.

Regarding accentedness, learners mainly reported paying attention to segmental aspects of pronunciation (G1: 53%, G2: 54%), that is, accuracy in the production of vowels and consonants, often specifying which specific sounds they were paying attention to. Both G1 and G2 reported paying attention to the degree of nativeness of their own productions (30% vs. 38%, respectively). Suprasegmental and prosodic aspects of speech production such as rhythm (G1: 7%, G2: 0), fluency (G1: 7%; G2: 4%) and stress (G1: 3%; G2: 4%) were mentioned in fewer cases. Answers in both groups were very similar.

To sum up, learners reported focusing mainly on pronunciation and ease of understanding when assessing their own speech for comprehensibility, whereas for accentedness they reported focusing on segmental accuracy and how native-like their productions were. Participants in both groups reported paying attention to the same linguistic dimensions.

DISCUSSION

This study revealed that L2 learners' self-assessments of their own speech for accentedness did not match the ratings by NLs, as in previous research. However, unlike previous findings, the learners' and NLs' assessments for comprehensibility were moderately related to one another. It is possible that learners found rating accentedness accurately harder than rating comprehensibility, as they might have lacked the necessary phonological awareness to do so or they might have been unaware of their own pronunciation mistakes. On the other hand, they might have found rating comprehensibility accurately easier, especially if when rating their own speech, they were thinking of how their speech might be easy or difficult to understand to NLs. A hypothesis we could not test with the current data set. Moreover, it appeared that more comprehensible and less accented learners under-rated themselves, while less comprehensible and more strongly accented learners over-rated themselves, which lends support to earlier data (Trofimovich et al., 2016; Saito et al., 2020).

In addition, as shown in Figure 3, those learners who conducted peer assessments were slightly more accurate in their self-assessments for accentedness than those who had not, but this difference did not reach significance. Such a tendency was not observed for comprehensibility (cf. Saito et al., 2020). More extended self-assessment practice or performing peer assessment benchmarking training over time (as in Tsunemoto et al., 2021) may be needed for learners to calibrate their self-assessments and to raise L2 learners' awareness of their own comprehensibility.

Finally, we found learners' intuitive reports of the linguistic dimensions underlying comprehensibility and accentedness to differ considerably from what we know native speakers pay attention to when rating non-native speech. For example, in our sample, participants did not report paying much attention to grammar and fluency when rating comprehensibility, or stress, rhythm and intonation when rating accentedness. This bias towards a focus on speech features related to pronunciation may have been caused by the explicit teaching on segmental features of English speech received in their English phonetics and phonology course.

To conclude, the present study provides further evidence of the existence of a Dunning-Kruger effect in L2 speech. However, it does not offer conclusive support for peer assessment helping sensitize learners to the pronunciation features that make their speech less comprehensible and more strongly accented. Lastly, learners identify pronunciation and content as the key linguistic features underlying ease of understanding whereas accentedness was mainly linked to segmental accuracy and nativelikeness.

Future research should investigate whether learners would focus on different dimensions when rating others than when assessing their own speech and the extent to which their focus might be guided through listening and speech assessment instruction. It would also be interesting to investigate the extent to which NLs' assessments of comprehensibility match those of other non-native speakers with a variety of L1 backgrounds. In addition, longitudinal studies would help explore factors and instructional techniques that minimize the mismatch between self-assessments and listeners' ratings

of speech for comprehensibility and accentedness. Finally, more research is needed to assess the effects of speech assessment training on developing L2 pronunciation.

ABOUT THE AUTHOR

Mireia Ortega is currently a lecturer and researcher at the University of Barcelona, where she has lectured on phonetics and phonology, linguistics, and second language acquisition. Her interests include the development of speech perception and production in foreign language acquisition, the use of phonetic training in improving learners' pronunciation and comprehensibility, and the effects of individual differences on L2 pronunciation development.

Department of Modern Languages and Literatures and English Studies.
Faculty of Philology and Communication, University of Barcelona.
Gran Via de les Corts Catalanes, 585.
08007-Barcelona, Spain.
m.ortega@ub.edu

Joan C. Mora is associate professor in the Department of Modern Languages and Literatures and English Studies in University of Barcelona. His research has examined the role of contextual factors and individual differences in the development of L2 speech and oral fluency, and the acquisition of L2 phonology.

Department of Modern Languages and Literatures and English Studies.
Faculty of Philology and Communication, University of Barcelona.
Gran Via de les Corts Catalanes, 585.
08007-Barcelona, Spain.
mora@ub.edu

Ingrid Mora-Plaza is postgraduate researcher and lecturer at the University of Barcelona, where she teaches English phonetics and phonology courses. Her research focuses on L2 pronunciation learning and teaching and the effects of individual differences on L2 speech acquisition. She is interested in the potential benefits of task-based pronunciation teaching and high-variability phonetic training on L2 speech development in instructional settings.

Department of Modern Languages and Literatures and English Studies.
Faculty of Philology and Communication, University of Barcelona.
Gran Via de les Corts Catalanes, 585.
08007-Barcelona, Spain.
imoraplaza@ub.edu

ACKNOWLEDGEMENTS

This research was funded through Grant PID2019-107814GB-I00 from the Spanish Ministry of Science, Innovation and Universities.

REFERENCES

Abrahamson, N., & Hyltenstam, K. (2009). Age of acquisition and nativelikeness in second language – listener perception vs. linguistic scrutiny. *Language Learning, 59,* 249-306.

Boersma, P., & Weenink, D. (2020). Praat: doing phonetics by computer (Version 6.1.09) [Computer program]. http://www.praat.org/ (accessed 1 January 2020).

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition, 31*(4), 533-557.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*(3), 69–106.

Kissling, E. M., & O'Donnell, M. E. (2015). Increasing language awareness and self-efficacy of FL students using self-assessment and the ACTFL proficiency guidelines. *Language Awareness, 11*, 283-302.

Kuhl, P., Conboy, B., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society. B 363*, 979-1000.

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

Johnson, K., Ehrlinger, K., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12,* 83-87.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *Tesol Quarterly, 39*(3), 369-377.

Levis, J. M. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation, 6*(3), 310-328.

Li, M., & Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing, 38*(2), 189-218.

Meara, P. M., & Milton, J. (2003). X_Lex: The Swansea advanced vocabulary levels test. (V2.05.) [Computer software]. Lognostics. https://www.lognostics.co.uk/tools/

Meara, P. M., & Miralpeix, I. (2006). Y_Lex: The Swansea advanced vocabulary levels test. (V2.05.) [Computer software]. Lognostics. https://www.lognostics.co.uk/tools/

Mitterer, H., Eger, N. A., & Reinisch, E. (2020). My English sounds better than yours: Second-language learners perceive their own accent as better than that of their peers. *PloS One, 15*(2), e0227643.

Moyer, A. (2014). Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics, 35*(4), 418-440.

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

Munro, M. J., & Derwing, T. M. (2020). Foreign accent, comprehensibility and intelligibility, redux. *Journal of Second Language Pronunciation, 6,* 283-309.

O'Brien, M. G. (2019). Attending to second language lexical stress: exploring the roles of metalinguistic awareness and self-assessment. *Language Awareness*, *28*(4), 310-328.

Saito, K., Trofimovich, P., Abe, M., & In'nami, Y. (2020). Dunning-Kruger effects in second language speech learning: How does self-perception calibrate with other perception over time? *Individual Differences and Learning, 79,* 101849.

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439-462.

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical correlates of comprehensibility versus accentedness in second language speech. *Bilingualism: Language and Cognition, 19*(3), 597-609.

Strachan, L., Kennedy, S., & Trofimovich, P. (2019). Second language speakers' awareness of their own comprehensibility. *Journal of Second Language Pronunciation*, *5*(3), 347-373.

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

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.

Tsunemoto, A., Trofimovich, P., Blanchet, J., Bertrand, J., & Kennedy, S. (2021). Effects of benchmarking and peer-assessment on French learners' self-assessments of accentedness, comprehensibility, and fluency. *Foreign Language Annals, 55*, 135-154.

Wrembel, M. (2015). Metaphonological awareness in multilinguals: A case of L3 Polish. *Language Awareness*, *24*(1), 60-83.

APPENDIX

Rating instructions given to participants

In order to do this task you need to have Praat installed on your computer. You can download it from this link: http://www.fon.hum.uva.nl/praat/

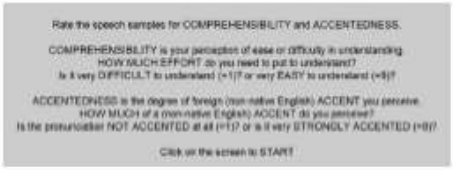
In this task we are asking you to rate some speech samples for comprehensibility and accentedness. There are 3 speech samples for practice and 20 speech samples of approximately 1 minute each that you need to rate twice, once for comprehensibility and once for accentedness. The speech samples were obtained through the picture-based retelling task you recorded in the booth in the lab.

Comprehensibility is how much effort it takes to understand what someone is saying. When you listen to someone's speech and you can understand her/him with ease, that is, without putting any effort, then a speaker is highly comprehensible. However, if you struggle with it and must listen very carefully, putting a great amount of effort in listening, or in fact you cannot understand what is being said at all, then a speaker has low comprehensibility.

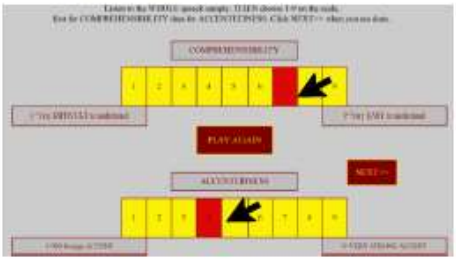
Accentedness is how much of a non-native accent you can perceive in someone's speech. For example, when a Spanish speaker speaks English you may notice that he/she does not sound like a native speaker of English because of the Spanish accent in his/her speech. We would like to know how much of a non-native English accent you perceive in the speech samples you are asked to rate.

The whole task will approximately take 45 minutes of your time and it will be done in Praat. You need to do the task, save your ratings and send it to us by uploading the Praat output in the Phonetics I course Campus Virtual.

1. Click on Run and read the instructions on the first page carefully,



then click on the screen to start the task.



1. FIRST pay attention to HOW EASY THE SPEECH SAMPLE IS TO UNDERSTAND. Listen to the speech sample carefully and rate it for COMPREHENSIBILITY by clicking on a number 1-9 on the scale.

2. Click on PLAY AGAIN to listen to the speech SAMPLE AGAIN.

3. THEN pay attention to HOW STRONGLY ACCENTED THE SPEECH SAMPLE IS. Listen to the speech sample carefully and rate it for ACCENTEDNESS by clicking on a number 1-9 on the scale.

4. CLICK on NEXT to listen to the next speech sample, and rate it in the same way.