

Research Report

BEYOND LINGUISTIC FEATURES

EXPLORING BEHAVIORAL AND AFFECTIVE CORRELATES OF COMPREHENSIBLE SECOND LANGUAGE SPEECH

Charles L. Nagle* 

Iowa State University

Pavel Trofimovich 

Concordia University

Mary Grantham O'Brien 

University of Calgary

Sara Kennedy

Concordia University

Abstract

Comprehensibility, or ease of understanding, has emerged as an important construct in second language (L2) speech research. Many studies have examined the linguistic features that underlie this construct, but there has been limited work on behavioral and affective predictors. The goal of this study was therefore to examine the extent to which anxiety and collaborativeness predict interlocutors' perception of one another's comprehensibility. Twenty dyads of L2 English speakers

This study was supported by grants from the Social Sciences and Humanities Research Council of Canada (SSHRC) to the second, third, and fourth authors. We are grateful to Kym Taylor Reid, Lauren Strachan, and Clinton Hendry for help with data collection; Aki Tsunemoto for help with data analyses; and the anonymous reviewers and the journal editor for the insightful comments and suggestions that helped us refine this article.   The experiment in this article earned an Open Materials and Open Data badge for transparent practices. The materials and data are available at <https://osf.io/kgev/> and at <https://www.iris-database.org/iris/app/home/detail?id=york:938889>

* Correspondence concerning this article should be addressed to Charles L. Nagle, Department of World Languages and Cultures, Iowa State University, 3102 Pearson Hall, 505 Morrill Road, Ames, Iowa 50011. E-mail: cnagle@iastate.edu

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

completed three interactive tasks. Throughout their 17-minute interaction, they were periodically asked to evaluate their own and each other's anxiety and collaborativeness and to rate their partner's comprehensibility using 100-point scales. Mixed-effects models showed that partner anxiety and collaborativeness predicted comprehensibility, but the relative importance of each predictor depended on the nature of the task. Self-collaborativeness was also related to comprehensibility. These findings suggest that comprehensibility is sensitive to a range of linguistic, behavioral, and affective influences.

INTRODUCTION

To communicate successfully in a second or additional language (L2), speakers must convey their message in a way that listeners can understand. Listeners may both understand a speaker and find the speaker easy to understand or may understand a speaker while needing to expend considerable effort. This is the basis for the distinction between intelligibility, a measure of actual understanding, and comprehensibility, listeners' perceived ease of understanding (Munro & Derwing, 1995; Nagle & Huensch, 2020). While intelligibility is a sensible baseline, most L2 speakers want their speech to be easy to understand, a goal that is more closely aligned with the notion of comprehensibility. Comprehensibility is also an intuitive evaluation that can be assessed through simple rating scales (e.g., *very difficult*–*very easy to understand*). Comprehensibility has therefore emerged as a particularly useful construct. Multiple lexical, grammatical, and phonological features underpin comprehensibility (Saito et al., 2017; Trofimovich & Isaacs, 2012), which means that L2 speakers' comprehensibility is likely to change as they produce varying levels of accuracy and complexity in each of these dimensions (Nagle et al., 2019). Comprehensibility and the linguistic features that underlie it also depend on the characteristics of the communicative task. For instance, when speakers engage in cognitively demanding tasks, their comprehensibility may decrease, and the use of accurate and sophisticated grammar and vocabulary may take on greater importance as they strive to convey complex ideas and relationships (Crowther et al., 2015).

What is missing from this body of work is a nuanced understanding of how comprehensibility unfolds over time in interactive scenarios, as speakers and listeners react and adapt to each other in real time. Recent work, which is compatible with dynamic views of language learning and use (de Bot et al., 2007), has begun to address this challenge, showing that comprehensibility is at least partially coconstructed (Trofimovich et al., 2020). Speakers and listeners appear to calibrate their speech to one another, resulting in a dynamic coupling of their comprehensibility. In interaction, however, comprehensibility is about more than just linguistic features. Comprehensibility might also have a strong affective and behavioral dimension. Just as listeners make a range of interpersonal evaluations based on a speaker's pronunciation (Fuentes et al., 2012), so too can the affective and behavioral dimensions of interpersonal dynamics influence interlocutors' comprehensibility. If L2 research is to achieve a transdisciplinary perspective (The Douglas Fir Group, 2016), then speech ratings, such as comprehensibility, must be coordinated with other socioaffective and behavioral measures that address the multi-dimensional nature of L2 communication.

Two socioaffective and behavioral components of communication that might have relevance to comprehensibility are interlocutors' anxiety and engagement, both of which can be conceived of as person-specific traits (i.e., some individuals are more anxious or engaged than others) and as states that emerge depending on the characteristics of the communicative setting (i.e., in certain situations, an individual may become more or less anxious or engaged). Broadly defined as a person's negative emotional reaction experienced in a situation in which a language is used (Gardner & MacIntyre, 1993), anxiety has been linked to lower levels of language achievement, with a medium-size effect ($r = -.36$), as shown in a recent meta-analysis of 97 studies (Teimouri et al., 2019). Increased levels of anxiety appear to inhibit the processing of linguistic stimuli at the input stage and to interfere with language production (MacIntyre & Gardner, 1994). As a construct with a strong socioaffective component, anxiety has also been argued to impact L2 speaker attitudes and motivational dispositions (Gardner & MacIntyre, 1993) and to undermine language development by disrupting communication processes (Dewaele, 2010). More recent research investigating state- or situation-specific aspects of anxiety has linked it on a dynamic timescale to speakers' individual experiences, such as their topic choice, their knowledge of vocabulary, and listeners' verbal and nonverbal reactions to speakers (Gregersen et al., 2014). Overall, then, the accumulated body of work on anxiety suggests that experiencing high levels of anxiety in general or at particular points in an interaction might distract interlocutors, interfering with the cognitive processes that are necessary for producing and comprehending speech. This interference could then lead to decreased comprehensibility.

Another dimension relevant to comprehensibility is speaker engagement, which broadly refers to people's degree of interest and participation in an activity (Philp & Duchesne, 2016). To date, various components of engagement—including cognitive (e.g., sustained attention or effort), behavioral (e.g., quantity of task-relevant talk), and social (e.g., reciprocity shown by speakers, as in turn-taking)—have been linked to contextual and situational variables in L2 communication. Speaker engagement is high when interlocutors communicate about familiar topics, rather than repeat the same task and content (Qiu & Lo, 2017). Engagement is also high when speakers discuss content relevant to their lives and experiences, compared to externally imposed topics (Lambert et al., 2017), and engagement is greater when speakers communicate with interlocutors of higher proficiency (Dao & McDonough, 2018). Unlike computer-mediated communication, face-to-face interaction is particularly conducive to eliciting higher levels of engagement in L2 speakers, especially in complex tasks (Baralt et al., 2016). Seen from this perspective, engagement (broadly defined) might therefore shape interlocutors' perception of each other's comprehensibility. That is, whereas anxiety might interfere with the cognitive and behavioral processes that are necessary for successful (L2) communication, speaker engagement might lead to greater understanding, especially in an interactive context where one partner's comprehensibility is at least partially dependent on the other's.

THE PRESENT STUDY

Comprehensibility has to date been researched nearly exclusively in relation to linguistic elements of interaction, focusing on how speakers' comprehensibility is shaped by

various phonological, fluency, grammatical, and discursive features in their speech, typically across different tasks (Crowther et al., 2015; Saito et al., 2017). The goal of this exploratory study was to extend this work beyond a strictly linguistic realm by investigating comprehensibility in interaction as a function of the affective and behavioral dimensions of anxiety and engagement. Anxiety and engagement are clearly multi-dimensional constructs, with multiple measures offering insight into their different facets, such as heart rate and galvanic skin response for anxiety or display of positive emotion and turn-taking frequency for engagement. Nevertheless, due to lack of systematic prior work linking comprehensibility to anxiety and engagement, we operationalized anxiety and engagement broadly, using scalar ratings, to elicit interaction-centered measures for these constructs from L2 speakers. Anxiety was defined as perceived stress, worry, or nervousness that a speaker is feeling while completing a task. Engagement was operationally defined as the perceived degree of a speaker's collaborativeness.

To explore links between interlocutors' comprehensibility and their perceived anxiety and collaborativeness, we revisited our dataset featuring paired interactions between L2 English speakers in three tasks, where the speakers carried out repeated assessments of themselves and each other (2.5 minutes apart) during 17 minutes of interaction. In our prior publication (Trofimovich et al., 2020), we tracked the speakers' comprehensibility ratings across time, exploring whether the ratings converged or diverged over time and task. For this report, we analyzed previously unpublished data targeting the speakers' self- and partner-specific ratings of anxiety and collaborativeness in relation to comprehensibility.

Because of the exploratory nature of this study, we made no specific predictions regarding the nature and strength of the relationships for comprehensibility, beyond anticipating a negative association with anxiety (a higher degree of anxiety might be associated with lower comprehensibility) and a positive association with collaborativeness (greater collaboration might co-vary with higher comprehensibility). However, because speaking task appears to impact situation-specific anxiety and engagement (Gregersen et al., 2014; Lambert et al., 2017; Qiu & Lo, 2017), we anticipated differences in associations across the different tasks performed by the speakers. With the overarching goal of understanding L2 speech as a dynamic, coconstructed system where socio-affective, linguistic, and behavioral factors interact to shape interlocutors' mutual impressions, we asked the following exploratory research question: To what extent do L2 interlocutors' impressions of one another's anxiety and collaborativeness predict their comprehensibility ratings in interaction?

METHOD

PARTICIPANTS

The interaction data came from a corpus of L2–L2 conversations between 40 (14 female, 26 male) university-level speakers at an English-medium university in Canada (Trofimovich et al., 2020). The speakers ($M_{\text{age}} = 25.85$ years, $SD = 2.89$), who represented 17 ethnolinguistic backgrounds, had begun learning English on average at 8.18 years ($SD = 4.58$) through primary and secondary instruction in their home countries and were recently accepted first-year graduate students in eight academic disciplines. Because all

speakers were studying in a university with a large cohort of international students, they reported substantial daily use of English ($M = 56.75\%$, $SD = 19.79$; 0–100% scale) and fairly high familiarity with accented English ($M = 6.33$, $SD = 1.67$; 1–9 scale). As part of university admission requirements, the speakers reported IELTS (31) or TOEFL (9) scores. When the nine TOEFL scores were replaced by equivalent IELTS values through validated conversion metrics (Educational Testing Service, 2017; Taylor, 2004), the speakers' IELTS performance was at a mean of 6.84 ($SD = 0.62$) for speaking and 7.60 ($SD = 0.95$) for listening. To contextualize these proficiency values among other established metrics, the mean IELTS speaking score of 6.84 roughly corresponds to TOEFL iBT speaking scores in the 20–23 range and the C1 Common European Framework of Reference for Languages (CEFR) band, whereas the mean IELTS listening score of 7.60 corresponds to TOEFL iBT listening scores in the 27–28 range and the C1 CEFR band. To encourage the use of English, the 40 speakers were randomly assigned to 20 pairs, such that the paired speakers were previously unfamiliar with each other and came from different backgrounds (see online Supplementary Materials).

SPEAKING TASKS AND TARGET RATINGS

The corpus included three task performances per pair, with all tasks completed in the same order. During the first (warm-up) task, the speakers were asked to discover three things they had in common with their partner (e.g., a favorite movie). For the second task, the speakers were asked to develop a coherent shared narrative using a set of 14 scrambled pictures, with seven images randomly distributed to each partner and partners unable to see one another's images. The 14 images told a story of a man who won the lottery but subsequently experienced a misfortune that made him realize that wealth does not always equal happiness. For the final task, the speakers were asked first to share some of the challenges they experienced as international students adjusting to life in a new academic environment (e.g., gaining access to health care, obtaining work permits) and then to provide common solutions for these challenges. The warm-up task lasted 3 minutes; the remaining two tasks lasted 7 minutes each.

During the 17-minute interaction, each speaker provided seven sets of ratings for comprehensibility (reported in Trofimovich et al., 2020) and for anxiety and collaborativeness (previously unpublished, analyzed here as time-sensitive predictors of comprehensibility). The seven sets of ratings occurred at comparable intervals: after each task (Times 1, 4, and 7) and approximately 2.5 minutes and 5 minutes into Task 2 (Times 2 and 3) and Task 3 (Times 5 and 6). The speakers used a paper booklet to record their ratings, with continuous scales (100-millimeter lines) printed next to each dimension, one labeled "me" for the self-rating and the other labeled "my partner" for the rating of the speaker's partner. Each scale included only endpoint labels, and the speakers marked a point on each line corresponding to their impression.

Although comprehensibility has typically been measured through 7- or 9-point Likert scales (e.g., Munro & Derwing, 1995), researchers have occasionally opted for continuous scales over ordinal ones, using a straight line bounded by endpoint descriptors in a paper-and-pencil format (e.g., Isaacs et al., 2015), as in this study, or a slider to record the rating in a computer or online interface (e.g., Saito et al., 2017). Existing scale validation and scale comparison work indicates that there is little difference in the ratings of

comprehensibility obtained through scales of various lengths and resolutions (Isaacs & Thomson, 2013), through different scale types (Munro, 2018), or through static or dynamic assessments (Nagle et al., 2019), which implied that the choice of the comprehensibility scale in this study was unlikely to have impacted rating validity. Comprehensibility was defined for the speakers as a judgment of how much effort it takes to understand what someone is saying. Anxiety was introduced as the level of stress, worry, or nervousness that someone is feeling while completing a task. Collaborativeness referred to the action of working with someone to produce or create something. Collaborating implied active participation and working together as a team, whereas not collaborating involved lack of participation and acting as an individual rather than a team member (see online Supplementary Materials).

PROCEDURE

The two speakers in each pair, participating in one audio-recorded session, were seated at opposite sides of a table, with seating determined randomly upon speaker arrival. A low barrier was placed between the speakers to prevent them from seeing one another's materials while allowing for an unobstructed view of gestures and facial expressions. After completing a background questionnaire, the speakers heard a research assistant (RA) define each rated dimension and explain how to use the rating booklet, which included instructions for each task and seven sets of scales (one per page). The speakers were told that they would engage in repeated assessments, evaluating the immediately preceding 2–3 minutes of interaction, and that their ratings would be private. They were also reminded that, during Tasks 2 and 3, the RA would stop the interaction briefly to allow for mid-task assessments. Specific task instructions were given before each task, always in the same manner. The speakers read the instructions, then summarized the instructions to the RA as a comprehension check, and finally asked clarification questions. The speakers were reminded that Task 1 would be stopped after 3 minutes and Tasks 2 and 3 after 7 minutes even if the discussion were ongoing, and that the RA would be using a timer to keep task duration and assessment intervals comparable.

DATA ANALYSIS

TARGET MEASURES AND COVARIATES

The criterion variable was the speakers' ratings of their partner's comprehensibility. As in Trofimovich et al. (2020), these ratings were recorded per speaker at the seven rating episodes and expressed numerically (out of 100), by measuring the distance with a ruler (to the nearest millimeter) between the anchor point and the speaker's mark (the intersection of the cross or angle point of the checkmark) on the 100-millimeter scale. The predictors were each speaker's self- and partner-specific ratings of anxiety and collaborativeness, on the assumption that the speakers' impressions of comprehensibility might be shaped not only by how they view their partner's anxiety and collaborativeness (henceforth, partner-anxiety and partner-collaborativeness) but also by how the speakers perceived their own anxiety and collaborativeness (henceforth, self-anxiety and self-collaborativeness). These ratings were similarly derived per speaker at each rating episode

and expressed numerically. Audio recordings of interaction were transcribed to determine each speaker's lexical output during interaction so a content measure could be used as a covariate.

To control various potential influences on ratings of comprehensibility, anxiety, and collaborativeness over time, several covariates were retained from the original dataset. The first covariates were each speaker's IELTS speaking and listening scores, included on the assumption that the speakers' ratings might have reflected their own or their partners' L2 skill level. The second covariate was a measure of type frequency, derived through lexical profiling, from each speaker's output in each segment preceding the rating episode (i.e., before Time 1, between Time 1 and 2, and so on). This covariate was included to account for the speakers' lexical contribution, assuming that the ratings might reflect the amount of content produced before each assessment. The final covariate was a time deviation variable, which captured each pair's deviation from the intended rating time. Although all pairs engaged in each task for comparable amounts of time and performed repeated assessments at similar intervals (see Trofimovich et al., 2020), individual variations (ratings occurring earlier or later than intended) may have impacted them.

STATISTICAL MODELING

We used mixed-effects models to estimate relationships between anxiety and collaborativeness, the primary predictor variables of interest in this study, and comprehensibility. Mixed-effects models are especially appropriate for analyzing longitudinal data because they are robust in the face of missing data and make simpler statistical assumptions than other analyses such as ANOVA (for an overview, see Cunnings & Finlayson, 2015; Linck & Cunnings, 2015). Mixed-effects models are also well-suited to hierarchical data structures, where one unit is nested within a higher-order unit, such as students within classes or, in the current study, speakers within pairs. Mixed-effects models allow researchers to account for hierarchical data (i.e., to account for the fact that the students in one class or the speakers in one pair are more likely to be similar to one another than to the students in another class or the speakers in another pair) through random effects. Most importantly for our purposes, mixed-effects modeling is a more flexible statistical option that is conducive to time-varying independent variables, or independent variables that take on unique values at each point in time, such as the partner- and self-ratings of anxiety and collaborativeness that participants provided at each of the seven rating episodes.

We fit models using the *lme4* package (Bates et al., 2015) in R Version 4.0.2 (R Core Team, 2020). In our previous work, we focused on the effect of time on comprehensibility ratings to determine the extent to which the speakers' ratings of one another changed and potentially converged over time. Our final (piecewise) model included separate predictors involving a quadratic trend for time over Tasks 1–2 and a linear trend over Task 3, covariates to control for the speakers' speaking and listening proficiency, lexical output (type frequency), and variability in the timing of repeated assessments (time deviation), as well as random intercepts for speakers and pairs. We adopted this model as the baseline model here, integrating anxiety and collaborativeness as predictors of comprehensibility while controlling for the effect of time and other covariates. To streamline the analyses, as in our earlier work, we split the data into two comparable datasets—one for Tasks 1–2 (four rating episodes) and another for Task 3 (three rating episodes)—that enabled us to

explore whether the effect of anxiety and collaborativeness on comprehensibility varied across tasks, providing insight into task-induced variation between speakers' affective state and engagement in relation to comprehensibility.

To limit model complexity, we adopted a conservative approach, integrating each of the partner- and self-ratings of anxiety and collaborativeness as predictors into the baseline model. Using the Akaike Information Criterion (AIC) in single-effect models, where a lower AIC indicates better fit, we then ranked the anxiety and collaborativeness predictors according to their informativeness. This ranking dictated order of entry as we evaluated more complex models. At each step, we compared model fit through likelihood ratio tests, retaining the corresponding anxiety or collaborativeness predictor only if it significantly improved fit. We opted for a simple random-effects structure consisting of random intercepts for speakers and pairs, with all fixed effects standardized.

For each final model, we computed variance inflation factors to check for multicollinearity among the predictors and plotted model residuals to confirm that they were distributed normally. All inflation values were below 2, indicating that multicollinearity was not a concern, but the residual plots showed that both models had a heavy lower tail. To correct for this excursion, we screened the datasets for residuals larger than 2.5 and refit the models to the pruned data. Following this procedure, we removed 5 of 160 observations for Tasks 1–2 and 6 of 120 observations for Task 3, which brought the distribution of residuals closer to normality (though some deviation was still observed at the tails). The final models were thus fit to 155 observations for Tasks 1–2 and 114 observations for Task 3.

RESULTS

GENERAL PATTERNS

As a first step, we plotted self and partner anxiety and collaborativeness ratings to examine change over time (with descriptive statistics summarized in online Supplementary Materials). As shown in [Figure 1](#), anxiety and collaborativeness were near mirror images of each other; the more collaborative the speakers considered themselves and their partner to be, the less anxiety they perceived. The figure also underscores the importance of task characteristics, where dotted lines indicate a shift to the next task. During Task 2, where the speakers worked with separate images to narrate a story together, they gave low ratings for collaborativeness while indicating that they were relatively anxious. In contrast, Task 3, where the speakers discussed potential solutions to common challenges faced by international students, showed the opposite pattern, with high collaborativeness coinciding with low anxiety.

To gain insight into the relationship between these measures, we computed global correlation coefficients between the anxiety and collaborativeness ratings, pooling over data points. As shown in [Table 1](#), anxiety and collaborativeness were negatively linked, insofar as higher collaborativeness was associated with lower anxiety. Self-self and partner-partner ratings showed moderate to large correlations, such that the speakers' self-perceptions of anxiety and collaborativeness were strongly linked ($r = -.70$) as were the speakers' judgments of their partner's anxiety and collaborativeness ($r = -.56$). Although the relationships across self and partner ratings were predictably weaker, they

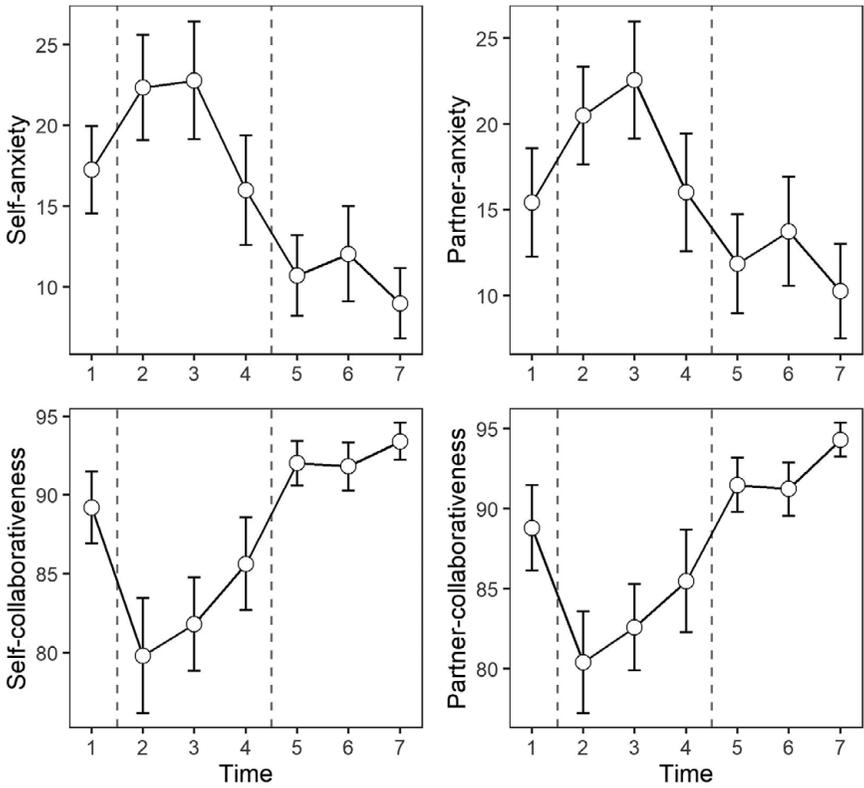


FIGURE 1. Self- and partner-specific anxiety and collaborativeness ratings over seven rating episodes. The vertical dashed lines indicate a change in task (Task 1 = 1, Task 2 = 2–4, Task 3 = 5–7). Dots represent the group mean, and error bars enclose 95% confidence intervals.

TABLE 1. Correlations between anxiety and collaborativeness ratings.

Rating	1	2	3
1 Self-anxiety			
2 Self-collaborativeness	-.70***		
3 Partner-anxiety	.05	-.24	
4 Partner-collaborativeness	-.38*	.50**	-.56***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

revealed links between anxiety and collaborativeness that were codependent on the two partners.

ANXIETY AND COLLABORATIVENESS IN TASKS 1–2

For Tasks 1–2, comparing AICs for single-predictor models revealed the following informativeness ranking for the target predictors: partner-collaborativeness, partner-

TABLE 2. Summary of final mixed-effects model for comprehensibility in Tasks 1–2.

Fixed effects	Estimate	SE	<i>t</i>	95% CI	<i>p</i>
Intercept	83.72	1.35	62.06	[80.95, 86.52]	<0.001
Partner-collaborativeness	7.19	1.13	6.36	[4.96, 9.43]	<0.001
Partner-anxiety	-5.96	1.17	-5.12	[-8.28, -3.60]	<0.001
Self-collaborativeness	2.85	0.98	2.91	[0.88, 4.82]	0.004
Covariates					
Time (linear)	-1.41	0.81	-1.75	[-3.00, 0.18]	0.08
Time (quadratic)	1.18	0.85	1.39	[-0.50, 2.85]	0.17
IELTS speaking	-.21	1.35	-0.16	[-3.05, 2.65]	0.88
IELTS listening	1.28	1.27	1.00	[-1.32, 3.99]	0.32
Type frequency	-.33	1.11	-0.30	[-2.54, 1.88]	0.77
Time deviation	-3.51	2.46	-1.43	[-8.67, 1.44]	0.17
Random intercepts		SD			
Pair		2.92			
Speaker		5.44			

Note: The poly function was used to fit orthogonal polynomials for time. The lmerTest package was used to estimate *p* values. All predictors were standardized using the scale function.

Abbreviation: CI, confidence interval.

anxiety, self-collaborativeness, and self-anxiety. The first three predictors improved model fit, whereas self-anxiety did not. Therefore, model building concentrated on integrating the significant factors in stepwise order of informativeness. The addition of these effects significantly improved model fit, resulting in the best-fitting model reported in Table 2 (for a summary of models fit and model comparisons, see online Supplementary Materials). The marginal R^2 (.590) for this model showed that the fixed effects alone explained 59% of the variance in comprehensibility ratings over the first two tasks. The complete model, including random effects, accounted for nearly 72% of the variance in comprehensibility (conditional $R^2 = .717$). According to Plonsky and Ghanbar's (2018) benchmarks for interpreting R^2 values for multiple regression, the explanatory power of this model would be considered moderate to large.

As shown in Table 2, the speakers' perception of their partner's collaborativeness emerged as the strongest predictor of comprehensibility, with the largest coefficient. The more collaborative the speakers perceived their partner to be, the higher they rated their partner's comprehensibility. The speakers' perception of their partner's anxiety was also negatively linked to comprehensibility, although with a smaller coefficient indicative of a slightly weaker relationship. The more anxious the speakers perceived their partner to be, the lower they rated their partner's comprehensibility. Finally, although the speakers' self-rating of collaborativeness was significantly related to their partner's comprehensibility, its contribution was much weaker, yet far from trivial, such that the speakers' own degree of collaboration positively predicted how they viewed their partner's comprehensibility. None of the covariates emerged as significant predictors of comprehensibility over the first two tasks.

TABLE 3. Summary of final mixed-effects model for comprehensibility in Task 3.

Fixed effects	Estimate	SE	<i>t</i>	95% CI	<i>p</i>
Intercept	90.46	0.68	132.20	[89.04, 91.89]	<0.001
Partner-anxiety	-5.51	0.74	-7.48	[-7.17, -3.77]	<0.001
Partner-collaborativeness	2.12	0.71	2.99	[0.71, 3.56]	0.004
Covariates					
Time (linear)	0.42	0.54	0.77	[-0.67, 1.51]	0.44
IELTS speaking	-0.60	0.69	-0.86	[-1.97, 0.88]	0.40
IELTS listening	0.71	0.63	1.13	[-0.57, 1.99]	0.26
Type frequency	1.42	0.62	2.28	[0.16, 2.66]	0.02
Time deviation	-1.70	1.18	-1.45	[-4.22, 0.69]	0.17
Random intercepts		SD			
Pair	1.47				
Speaker	1.30				

Note: All predictors were standardized using the scale function.

Abbreviation: CI, confidence interval.

ANXIETY AND COLLABORATIVENESS IN TASK 3

We followed the same procedure to model comprehensibility as a function of anxiety and collaborativeness in Task 3. However, for this task, single-predictor models showed that only partner-anxiety and partner-collaborativeness significantly improved the baseline model and that partner-anxiety, unlike partner-collaborativeness in Tasks 1–2, emerged as the more informative predictor. Thus, partner-anxiety was integrated into the baseline model first, followed by partner-collaborativeness. With each step, model fit improved, resulting in the best-fitting model shown in Table 3. In this model, the fixed effects explained 60% (marginal $R^2 = .603$) of the variance in comprehensibility and the full model with random effects approximately 65% (conditional $R^2 = .647$). This model would also be considered to have moderate to large explanatory power (Plonsky & Ghanbar, 2018).

As shown in Table 3, the speakers' perception of their partner's anxiety and collaborativeness was associated with that partner's comprehensibility, such that the less anxious and more collaborative the speakers perceived their partner to be, the higher they rated their partner's comprehensibility. Compared to the model for Tasks 1–2, the effect of partner-anxiety remained similar (i.e., coefficients were comparable), but the effect of partner-collaborativeness decreased substantially, from 7.19 (Tasks 1–2) to 2.12 (Task 3). Most covariates remained nonsignificant, save type frequency, which was positively associated with comprehensibility, with greater type frequency in the partner's speech in the segment immediately preceding the rating linked to a higher comprehensibility rating for that partner.

In summary, modeling demonstrated that speakers' perception of their partner's comprehensibility was associated with their perception of their partner's collaborativeness and anxiety. In the first two tasks, collaborativeness was a stronger predictor than anxiety, whereas in the third task, anxiety was a stronger predictor than collaborativeness. Additionally, speakers' perception of their partner's comprehensibility was tied to their perception of their own collaborativeness, albeit to a lesser extent and only during the first two tasks.

DISCUSSION

As a metric of a person's subjective experience of the ease or difficulty with which information is processed (Reber & Greifeneder, 2017), comprehensibility likely captures various influences that enhance or impair listener experience with speech. Some influences might derive from the linguistic attributes of speech, such as its lexical sophistication, grammatical complexity, or segmental and suprasegmental accuracy (Saito et al., 2017; Trofimovich & Isaacs, 2012). Other contributors to comprehensibility might stem from the clarity or coherence of the speech content, as the speaker creates discourse (Nagle et al., 2019). Yet other influences on comprehensibility—some of which were explored here—might be related to interpersonal fluency (Ackerman & Bargh, 2010), or people's experience of effortlessness arising through social coordination. This coordination can involve behavior, such as people appropriating one another's gestures and speech patterns (Paxton et al., 2016), and affect, such as people becoming sensitive to one another's emotional and affective states (Parkinson, 2011).

Set against this backdrop, it is hardly surprising that collaborativeness and anxiety predicted comprehensibility. Conceptualized within the broader construct of engagement (Philp & Duchesne, 2016), collaborativeness ratings likely reflected various behavioral dimensions of social coordination. For instance, collaborativeness may have encompassed attention to task instructions, orientation toward task completion, quality of task-relevant talk, and reciprocity of participation, in the sense that partners needed to work together to attain the task goal without surrendering or seizing full control of the interaction. Although unpacking the distinct facets of the collaborative behavior relevant to comprehensibility was not feasible in the present study, the collaborativeness-comprehensibility link is revealing, in that L2 speakers' general perception of their partners' task involvement has a bearing on the ease or difficulty with which they understand those partners. It is also worth noting that the role of collaborativeness in promoting comprehensibility was evident even after controlling for speakers' lexical contribution to the conversation through the type frequency covariate, reinforcing the view that at least some aspects of collaborative behavior are distinct from and/or transcend linguistic output.

The association between anxiety and comprehensibility is a novel finding, linking comprehensibility to a socioaffective dimension of interaction. Anxiety ratings likely captured visual signs of anxious L2 speakers, such as restrained facial expressions, decreased eye contact, rigid postures, and hand movements focused on manipulating objects (e.g., clicking a pen) rather than on enhancing the meaning of speech (Gregersen, 2005). Anxiety ratings may have also reflected linguistic and interactional behaviors shown by anxious speakers, including generic rather than detailed utterances, avoidance in claiming or volunteering a turn, and frequent single-syllable backchannels with nonverbal encouragement (e.g., nodding) for the interlocutor to continue talking (Ely, 1986; Steinberg & Horwitz, 1986). These cues, individually or combined, may have made processing the L2 speakers' message more effortful for the interlocutor, leading to lower comprehensibility ratings.

It is important to acknowledge that the linguistic and behavioral cues of collaborativeness and anxiety that informed participants' holistic ratings of each partner-oriented dimension likely overlap. For example, an absence of task-relevant content detail, general

state of uneasiness, avoidance in claiming a turn, or lack of interest may be signs of both reduced collaborativeness and increased anxiety. It is little surprise, therefore, that the dynamic curves of anxiety and collaborativeness were mirror images of each other and that the two ratings shared up to 49% of their variance. Nonetheless, the two ratings remained sufficiently distinct, in that they predicted comprehensibility differently depending on the task. Most speakers felt that the picture narrative (Task 2) was the most difficult of the three tasks (Trofimovich et al., 2020). For the picture narrative task, speakers had to reconstruct a coherent, shared narrative from 14 scrambled images, which required close collaboration from both partners. Because collaboration was task-essential, it makes sense that collaborativeness was a stronger predictor of comprehensibility than anxiety. In contrast, for the discussion task focusing on the shared, lived experiences of international students adjusting to life in a new environment (Task 3), collaboration was less critical, insofar as every speaker had ample input to contribute, which could explain why anxiety emerged as a stronger predictor than collaborativeness. Additionally, in Task 3, speakers often discussed personally relevant emotional themes (e.g., culture shock), which likely heightened the task's socioaffective load, resulting in stronger links between comprehensibility and anxiety than between comprehensibility and collaborativeness. At the same time, these task-related findings should be interpreted with caution given that all pairs completed the tasks in a fixed order, which means that we could not separate the effects of task and time in the current analysis. To arrive at a full understanding of how socioaffective variables change depending on the task and the amount of time spent with a particular interlocutor, it would be necessary to counterbalance the order of tasks across pairs.

Finally, L2 speakers' judgments of their partner's comprehensibility were predicted by the speakers' own behavior, namely, their collaborativeness. Although this relationship emerged only for Tasks 1–2, it nonetheless implies that a speaker's comprehensibility—as assessed in dialogue—is coconstructed by both interacting partners. Put differently, a speaker's comprehensibility may reflect not only that speaker's linguistic and nonlinguistic behaviors but may also encompass the interlocutor's contributions to the dialogue. This relationship might reflect the halo effect, whereby speakers project a positive image of themselves on their partner, whose comprehensibility they are assessing. Alternatively, it might arise because people often misattribute their assessment of ease or difficulty to an irrelevant source (Greifeneder et al., 2011), which, in this case, amounts to speakers upgrading their partner's comprehensibility based on their own participation in dialogue. Regardless of its source, this self-oriented influence on partner comprehensibility in interaction represents a novel contribution to existing work, which to date has chiefly targeted individual differences in raters' cognitive and experiential profiles (e.g., Saito et al., 2019).

CONCLUSION

To conclude, this exploratory study revealed links between interlocutor-rated comprehensibility and affective (anxiety) and behavioral (collaborativeness) dimensions of interaction. For anxiety, this study extended prior work, where anxiety is typically rated retrospectively while speakers view their recorded performances in monologic tasks (Gregersen et al., 2014), into an interactive domain, with both interlocutors evaluating

their own and their partner's anxiety. For collaborativeness, the study allowed for tracking speaker participation on a minute-by-minute timescale, which complements previous longitudinal work (Oga-Baldwin & Nakata, 2017). Despite their promise, these findings must be revisited in future work. First, because the residuals in this study's mixed-effects models deviated from normality even after outlier cases had been removed, it would be important to replicate the present findings before making broader generalizations about the role of anxiety and collaborativeness in L2 comprehensibility. Similarly, in follow-up work, researchers could also target speakers of different proficiency levels engaged in other tasks (whose order must be rotated across speaker dyads) and employ other measures of collaborativeness (e.g., turn-taking frequency) and anxiety (e.g., galvanic skin response). Lastly, linking comprehensibility to various facets of task engagement and anxiety requires an understanding of whether speakers notice and use the cues that signal their partner's collaborativeness and anxiety. Such insight can be gained through stimulated recall, (video) observation, or eye-tracking. Online teaching and research environments may be particularly conducive to designs similar to ours, where partners are asked to periodically evaluate one another as they work through a set of communicative tasks. Above all, researchers should intensify work exploring links between speech assessments and various social, affective, and behavioral measures to clarify the multi-dimensional nature of L2 communication in both face-to-face and virtual environments. This work should prioritize interactive approaches to L2 communication, given that the relationships between linguistic features and speech ratings that have been documented using monologic speaking tasks may not hold during interaction, when a wider range of time-varying affective and behavioral influences are at play.

REFERENCES

- Ackerman, J. M., & Bargh, J. A. (2010). Two to tango: Automatic social coordination and the role of felt effort. In B. Bruya (Ed.), *Effortless attention: A new perspective in the cognitive science of attention and action* (pp. 335–371). MIT Press.
- Baralt, M., Gurzynski-Weiss, L., & Kim, Y. (2016). Engagement with the language: How examining learners' affective and social engagement explains successful learner-generated attention to form. In M. Sato & S. Ballinger (Eds.), *Peer interaction and second language learning: Pedagogical potential and research agenda* (pp. 209–239). John Benjamins.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *The Modern Language Journal*, 99, 80–95. <https://doi.org/10.1111/modl.12185>
- Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 159–181). Routledge.
- Dao, P., & McDonough, K. (2018). Effect of proficiency on Vietnamese EFL learners' engagement in peer interaction. *International Journal of Educational Research*, 88, 60–72. <https://doi.org/10.1016/j.ijer.2018.01.008>
- de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10, 7–21. <https://doi.org/10.1017/S1366728906002732>
- Dewaele, J.-M. (2010). Multilingualism and affordances: Variation in self-perceived communicative competence and communicative anxiety in French L1, L2, L3 and L4. *International Review of Applied Linguistics*, 48, 105–129. <https://doi.org/10.1515/iral.2010.006>

- Educational Testing Service. (2017). TOEFL iBT® and IELTS® academic module scores: Score comparison tool. <http://www.ets.org/toefl/institutions/scores/compare>
- Ely, C. M. (1986). An analysis of discomfort, risktaking, sociability, and motivation in the L2 classroom. *Language Learning*, 36, 1–25. <https://doi.org/10.1111/j.1467-1770.1986.tb00366.x>
- Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, 42, 120–133. <https://doi.org/10.1002/ejsp.862>
- Gardner, R. C., & MacIntyre, P. D. (1993). On the measurement of affective variables in second language learning. *Language Learning*, 43, 157–194. <https://doi.org/10.1111/j.1467-1770.1992.tb00714.x>
- Gregersen, T. S. (2005). Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign Language Annals*, 38, 388–400. <https://doi.org/10.1111/j.1944-9720.2005.tb02225.x>
- Gregersen, T., MacIntyre, P. D., & Meza, M. D. (2014). The motion of emotion: Idiodynamic case studies of learners' foreign language anxiety. *The Modern Language Journal*, 98, 574–588. <https://doi.org/10.1111/modl.12084>
- Greifeneder, R., Bless, H., & Pham, M. T. (2011). When do people rely on affective and cognitive feelings in judgment? A review. *Personality and Social Psychology Review*, 15, 107–141. <https://doi.org/10.1177/1088868310367640>
- Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, 10, 135–159. <https://doi.org/10.1080/15434303.2013.769545>
- Isaacs, T., Trofimovich, P., Yu, G., & Chereau, B. M. (2015). Examining the linguistic aspects of speech that most efficiently discriminate between upper levels of the revised IELTS Pronunciation scale. *IELTS Research Report Series*, 4, 1–48.
- Lambert, C., Philp, J., & Nakamura, S. (2017). Learner-generated content and engagement in second language task performance. *Language Teaching Research*, 21, 665–680. <https://doi.org/10.1177/1362168816683559>
- Linck, J. A., & Cummings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, 65, 185–207. <https://doi.org/10.1111/lang.12117>
- MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, 44, 283–305. <https://doi.org/10.1111/j.1467-1770.1994.tb01103.x>
- Munro, M. J. (2018). Dimensions of pronunciation. In O. Kang, R. I. Thomson, & J. M. Murphy (Eds.), *The Routledge handbook of contemporary English pronunciation* (pp. 413–431). Routledge.
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- Nagle, C. L., & Huensch, A. (2020). Expanding the scope of L2 intelligibility research: Intelligibility, comprehensibility, and accentedness in L2 Spanish. *Journal of Second Language Pronunciation*, 6, 329–351. <https://doi.org/10.1075/jslp.20009.nag>
- Nagle, C., Trofimovich, P., & Bergeron, A. (2019). Toward a dynamic view of second language comprehensibility. *Studies in Second Language Acquisition*, 41, 647–672. <https://doi.org/10.1017/s0272263119000044>
- Oga-Baldwin, W. L. Q., & Nakata, Y. (2017). Engagement, gender, and motivation: A predictive model for Japanese young language learners. *System*, 65, 151–163. <https://doi.org/10.1016/j.system.2017.01.011>
- Parkinson, B. (2011). Interpersonal emotion transfer: Contagion and social appraisal. *Personality and Social Psychology Compass*, 5, 428–439. <https://doi.org/10.1111/j.1751-9004.2011.00365.x>
- Paxton, A., Dale, R., & Richardson, D. C. (2016). Social coordination of verbal and nonverbal behaviors. In P. Passos, K. Davids, & J. Y. Chow (Eds.), *Interpersonal coordination and performance in social systems* (pp. 259–274). Routledge.
- Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, 36, 50–72. <https://doi.org/10.1017/S0267190515000094>
- Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R² values. *The Modern Language Journal*, 102, 713–731. <https://doi.org/10.1111/modl.12509>
- Qiu, X., & Lo, Y. Y. (2017). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research*, 21, 681–698. <https://doi.org/10.1177/1362168816684368>
- R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org>

- Reber, R., & Greifeneder, R. (2017). Processing fluency in education: How metacognitive feelings shape learning, belief formation, and affect. *Educational Psychologist, 52*, 84–103. <https://doi.org/10.1080/00461520.2016.1258173>
- Saito, K., Tran, M., Suzukida, Y., Sun, H., Magne, V., & Ilkan, M. (2019). How do L2 listeners perceive the comprehensibility of foreign-accented speech? Roles of L1 profiles, L2 proficiency, age, experience, familiarity and metacognition. *Studies in Second Language Acquisition, 41*, 1133–1149. <https://doi.org/10.1017/S0272263119000226>
- Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics, 38*, 439–462. <https://doi.org/10.1093/applin/amv047>
- Steinberg, F. S., & Horwitz, E. K. (1986). The effect of induced anxiety on the denotative and interpretive content of second language speech. *TESOL Quarterly, 20*, 131–136. <https://doi.org/10.2307/3586395>
- Taylor, L. (2004). IELTS, Cambridge ESOL examinations and the Common European Framework. *Cambridge ESOL Research Notes, 18*, 2–3.
- Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition, 41*, 363–387. <https://doi.org/10.1017/S0272263118000311>
- The Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *The Modern Language Journal, 100*, 19–47. <https://doi.org/10.1111/modl.12301>
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition, 15*, 905–916. <https://doi.org/10.1017/S1366728912000168>
- 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*, 430–457. <https://doi.org/10.1075/jslp.20003.tro>