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*. Published online 27 July 2021. https://doi.org/10.1075/jslp.21006.tro

Task engagement and comprehensibility in interaction: Moving from what second language speakers say to what they do

Pavel Trofimovich, Oguzhan Tekin, and Kim McDonough

Concordia University, Montreal, Canada

This exploratory study examined the relationship between second language (L2) English speakers' comprehensibility and their interactional behaviors as they engaged in a conversation with fellow L2 speakers. Thirty-six pairs of L2 English university students completed a 10-minute academic discussion task and subsequently rated each other's comprehensibility. Transcripts of their conversation were coded for eight measures of task engagement, including cognitive/behavioral engagement (idea units, language-related episodes), social engagement (encouragement, responsiveness, task and time management, backchanneling, nodding), and emotional engagement (positive affect). Speakers who showed more encouragement and nodding were perceived as easier to understand, whereas those who produced more frequent language-focused episodes and demonstrated more responsiveness were rated as harder to understand. These findings provide initial evidence for an association between L2 speakers' interactional behaviors and peer-ratings of comprehensibility, highlighting L2 comprehensibility as a multifaceted and interaction-driven construct.

Keywords comprehensibility, interaction, interactional behavior, task engagement, English as lingua franca, second language

1. Introduction

In a classic study published nearly 40 years ago, Varonis and Gass (1982) concluded that listeners' reactions to second language (L2) speech depend on comprehensibility, which they described as "how easy it is [for a listener] to interpret the [speaker's] message" (p. 125). They argued that comprehensibility is shaped by multiple influences, with pronunciation, grammar, fluency, familiarity, first language, conversational topic, and other contextual and social variables all affecting comprehensibility and influencing listener experience with L2 speech. In the subsequent decades, comprehensibility, defined as "judgments on a rating scale of how difficult or easy an utterance is to understand" (Derwing & Munro, 1997, p. 2), has indeed been linked to multiple linguistic dimensions in L2 speech, such as fluency, lexis, grammar, and pronunciation (Kang et al., 2010; O'Brien, 2014; Saito et al., 2016). Comprehensibility also seems to decrease or increase as speakers engage in more versus less demanding tasks (Crowther et al., 2015) and to vary as a function of listeners' exposure to L2 speech and their familiarity with the speaker's first language (Foote & Trofimovich, 2018; Munro et al., 2006). Comprehensibility is likewise influenced by social factors, such as listener attitudes or biases toward L2 speakers (Taylor Reid et al., 2020; Vaughn & Whitty, 2020).

However, conspicuously absent from this work is an understanding of how comprehensibility might be shaped in interaction through interlocutor behaviors. For example, as interlocutors work together to achieve understanding in dialogue, their mutual comprehensibility (i.e., how easy it is for them to understand each other) might be influenced not only by various linguistic dimensions of their speech or listener- and context-related characteristics but also by their behavior, such as providing backchannels (e.g., *uh huh*, *yeah*, *right*), displaying visual signals of understanding (e.g., nodding), or demonstrating reciprocity and mutuality in conversation (e.g., by encouraging the partner to speak or elaborating the partner's idea). Thus, to describe comprehensibility as a multifaceted, interaction-driven construct and to fully address the question articulated by Varonis and Gass (1982)—namely, why listeners react to L2 speakers in particular ways—it is important to investigate potential links between comprehensibility and interlocutor behavior.

A construct which may shed light on the relationship between interlocutor behavior and comprehensibility is engagement. In their model of task engagement, Philp and Duchesne (2016) define it as "a state of heightened attention and involvement" (p. 51) that can be expressed along behavioral, cognitive, social, and affective dimensions. The behavioral dimension of engagement captures a speaker's amount of task-relevant talk or the time invested in speaking. The cognitive dimension encompasses a speaker's sustained mental effort to maintain communication through elaboration or negotiation for meaning along with attention to language through language-related episodes (LREs), where a speaker might discuss language forms or engage in self- or peer-correction (Storch, 2008). The social dimension manifests as collaboration and mutuality where interlocutors support each other through various verbal and nonverbal cues, such as backchanneling and completing each other's utterances. Lastly, the emotional dimension corresponds to a speaker's expression of affect, such as frustration, boredom, or enjoyment.

To date, these dimensions of engagement have been linked to contextual variables in L2 communication. For example, L2 speakers contribute more content and display greater enthusiasm when they discuss familiar rather than unfamiliar topics (Qiu & Lo, 2017) or when they engage in tasks with fewer built-in constraints

(Nakamura et al., 2020). L2 speakers produce more task-relevant talk, elaborate their ideas more often, and provide interlocutors with more help and encouragement when they discuss personal experiences rather than externally imposed topics (Lambert et al., 2017). L2 speakers also supply more task-relevant talk, provide more frequent LREs, and demonstrate more tokens of encouragement, help, and responsiveness toward their interaction partner when they communicate with L2 speakers of higher proficiency than lower proficiency (Dao & McDonough, 2018). Finally, compared to computer-mediated tasks, face-to-face interaction appears to encourage speakers to pay greater attention to language, engage in more collaborative talk, and show more emotion, particularly during complex tasks (Baralt et al., 2016). Considering that L2 speakers show a variety of engagement behaviors, it is important to understand whether any of these behaviors shape their perception of each other's comprehensibility, especially because understanding is co-constructed in dialogue (Brennan et al., 2018).

Although no prior task engagement research has examined the potential link between interlocutor behavior and L2 speakers' comprehensibility, previous work on speaking, especially perceived fluency, offers insight into how interlocutor behavior may relate to comprehensibility. Defined as "a judgment made about speakers based on impressions drawn from their speech" (Segalowitz, 2010, p. 48), perceived fluency is similar to comprehensibility in that both reflect a listener's impressions about a speaker. When evaluating fluency, listeners primarily rely on temporal characteristics of speech, such as pausing and articulation speed (Bosker et al., 2013); however, when listeners assess a speaker's comprehensibility, they mostly orient to linguistic features, such as phonology, lexis, and grammar (Kang et al., 2010; O'Brien, 2014; Saito et al., 2016). Despite drawing on somewhat different speech properties, fluency and comprehensibility ratings are often correlated (Derwing et al., 2004), suggesting an overlap in their underlying dimensions.

To explore listeners' orientation to interlocutor behavior when rating fluency, Sato (2014) asked raters to describe the aspects of fluency that stood out to them as they listened to L2 speakers' paired interactions. The raters reported such behaviors as backchanneling, scaffolding, and turn-taking, with the idea that more fluent speakers readily contributed a speaking turn and encouraged their partner to claim one while acknowledging their partner's ideas and actively responding to them. Examining question–answer sequences, van Os et al. (2020) compared raters' judgments of fluency when there was a long pause between the turns or an overlap between the question and the answer. The raters downgraded L2 fluency when speakers had long pauses but not when there was overlap, suggesting that overlapping turn-taking was interpreted as a collaborative behavior with no detriment to perceived fluency. Given a strong link between perceived fluency and comprehensibility (Derwing et al., 2004; Derwing & Munro, 2015), it is possible that the various interlocutor behaviors captured through engagement measures may be linked to comprehensibility in the same way they are associated with perceived fluency during conversation (Sato, 2014; van Os et al., 2020), particularly if comprehensibility is evaluated by interactants themselves rather than by external raters.

2. The Present Study

The goal of this exploratory study is to extend prior work about the factors that influence comprehensibility, which to date has focused on various linguistic properties in L2 speakers' speech or targeted listener profiles (e.g., Saito et al., 2016, 2019), to the interactive domain, examining whether L2 speakers' task engagement

behaviors are related to their perceptions of peer comprehensibility. Given the paucity of prior research, our predictions as to how interlocutors' behaviors might contribute to the ease or difficulty with which they understand each other are unavoidably tentative. For behavioral and cognitive engagement (number of idea units, LREs), it is possible that speakers who produce more task content will be viewed as more comprehensible, given previously-reported associations between comprehensibility and story breadth in speaker narratives (Trofimovich & Isaacs, 2012). In contrast, a stronger focus on language in the form of frequent LREs, particularly when used excessively, may lead to lower comprehensibility because listeners consider language problems distracting from the speaker's message (Derwing et al., 2002; Fayer & Krasinski, 1987). As social engagement promotes collaboration toward a common task goal (through encouragement, responsiveness, task management) and facilitates mutual understanding (through backchanneling, nodding), it may alleviate some processing burden for the interlocutor (Sato, 2014), thereby increasing comprehensibility. For emotional engagement, speakers who show greater positive affect either verbally (e.g., *that's hilarious*) or nonverbally (e.g., smiling) might be perceived more favorably by the interlocutor, resulting in more generous comprehensibility ratings.

Because conversation is a social process where understanding is co-constructed between interlocutors (Brennan et al., 2018), it is possible that a speaker's behavior may also influence her perception of a peer's comprehensibility. For instance, speakers who produce more content, engage in collaborative talk by providing verbal or nonverbal feedback, or show greater range of positive affect might project a favourable view onto their partners, rating them as more comprehensible (Nagle et al., 2021). Lastly, given that 50–85% of variance in listener judgments of comprehensibility in monologic tasks is explained through linguistic properties in a speaker's speech, such as pronunciation, fluency, lexis, and grammar (Isaacs & Trofimovich, 2012; Kang et al., 2010; Saito et al., 2016), potential associations between interlocutor behaviors and comprehensibility were expected to be modest at best. In light of these assumptions, we addressed the following exploratory question: Do L2 speakers' task engagement interactional behaviors predict how they perceive each other's comprehensibility?

3. Method

3.1. Paired Oral Interactions

Paired interactions were sampled from the Corpus of English as a Lingua Franca Interaction (CELFI), where a diverse sample of L2 English students from Canadian English-medium universities in Montreal carried out three 10-minute communicative tasks in pairs (McDonough & Trofimovich, 2019). All students in the corpus had met the minimum English proficiency required for admission to university degree programs (minimum TOEFL iBT score of 75 or equivalent), which corresponded to the B2 to C1 levels in the Common European Framework of Reference. Students were randomly assigned to carry out the tasks (exchanging personal narratives, participating in an academic discussion, and proposing solutions to problems faced by newcomers to Canada) with a partner from a different language background, and all interactions were video- and audio-recorded. The present analyses focus on the academic discussion task because speaking tasks are central to university coursework (Zhu & Flaitz, 2005) but they often pose challenges to students (Cheng et al., 2004). For the discussion task, each interlocutor first read a different brief research report on the same topic and

then summarized the report and exchanged opinions with their partner. There were four topics (pros and cons of advertising, motivation for language learning, nature vs. nurture, and medical ethics), and each pair selected which topic they preferred to discuss.

The target sample included 36 paired academic task performances selected from the corpus according to three criteria: (a) mixed gender composition across dyads (12 male–male pairs, 12 female–female pairs, 12 male–female pairs), (b) similar distribution across the two most-frequently selected topics (17 for advertising, 19 for motivation), and (c) different order in the set of three tasks (five task orders attested across 36 dyads). The dyads were composed of 72 L2 English speakers with a mean age of 24.21 years (*SD* = 4.08, *range* = 18–37) who were pursuing various undergraduate (31) and graduate (41) degrees. The speakers had studied English for a mean of 13.70 years (*SD* = 5.80, *range* = 1–30) and resided in Canada for approximately 1.70 years (*SD* = 2.10, *range* = 0.10–14.0). They came from 20 different language backgrounds, the largest being Mandarin (14), French (8), Arabic (7), and Hindi (7). Appendix A provides a summary of the speakers' home languages, genders, and discussion topics.

After the discussion task, both speakers completed peer-ratings of comprehensibility, in addition to other assessments not analyzed here. The scale, which was a 100-millimeter line printed on paper, included no numerical values except anchor descriptors: *difficult to understand* on the left (equivalent to 0) and *easy to understand* on the right (equivalent to 100). Although comprehensibility is often measured via Likert scales (Munro & Derwing, 1995), continuous measurements are also common, including lines anchored by endpoint descriptors (Isaacs et al., 2015) and moving sliders (Saito et al., 2016). As there are few differences in comprehensibility ratings obtained through different scale types of various lengths and resolutions (Isaacs & Thomson, 2013; Munro, 2018), the scale choice was unlikely to have impacted rating validity in this study. The scale was accompanied by a definition, which described comprehensibility as how easy it was for each speaker to understand their partner. After reading the definition, each speaker put a checkmark or a cross on the line to estimate their partner's comprehensibility. The speakers' rating of their partners' comprehensibility was defined as the distance, measured with a ruler to the nearest millimeter, between the left endpoint and the intersection of the cross or angle point of the checkmark on the 100-millimeter scale.

3.2. Data Coding

All audio- and video-recorded paired interactions were first transcribed and verified by trained research assistants, then coded by the second researcher for eight variables across the domains of behavioral, cognitive, social, and emotional engagement (see Table 1 for examples). These variables reflect the categories in Philp and Duchesne's (2016) task engagement model and findings from recent engagement research. First, because various subcomponents of behavioral and cognitive engagement have been difficult to tease apart in prior work (Lambert et al., 2017; Oga-Baldwin & Nakata, 2017), we created a joint category of behavioral/cognitive engagement. Second, we operationalized social engagement in line with Dao and McDonough (2018), who focused on Philp and Duchesne's categories of collaboration, mutual support, and reciprocity in conversation, but added backchanneling (Bjørge, 2010; Lambert et al., 2017) and nodding (Nash, 2007; Whitehead, 2011) as additional cues indicative of speakers' social engagement. Finally, we characterized emotional engagement as Dao and McDonough (2018), who along with Philp and Duchesne

defined it as speakers' expression of positive affect, such as verbal expression of positive emotion, smiling, and laughter.

The category of behavioral/cognitive engagement included two measures reflecting each speaker's content contributions to the dialogue, based on the idea that a speaker's task engagement requires both the effort of providing discussion content and the attention to language use (Lambert et al., 2017; Philp & Duchesne, 2016; Shin et al., 2016; Storch, 2008).

1. Idea units, a measure compatible with McCarthy's (1991) theme–rheme concept, was defined as a clause providing a topic or a comment that is syntactically and intentionally distinct from others (Ellis & Barkhuizen, 2005). When speakers repeated or paraphrased the same idea or when their utterance did not convey a cohesive idea, an idea unit was not counted.
2. LREs were defined as episodes where a speaker either initiated or responded to a language issue (Swain & Lapkin, 1998), such as searching for an appropriate word to express an idea or repairing a language form. LREs were counted per speaker unless an episode involved both interlocutors, in which case that episode was counted for both.

Social engagement included five measures encompassing instances where speakers demonstrated collaboration, mutual support, and reciprocity in conversation behaviors, on the assumption that a speaker's task engagement requires not only an effort to maintain a language and content focus but also a social investment to scaffold and sustain the interaction (Baralt et al., 2016; Lambert et al., 2017; Storch, 2001).

1. Encouragement was defined as any instances where a speaker explicitly encouraged the interlocutor to take the floor, proceed speaking, or elaborate their argument. Instances where encouragement was expressed through a question (e.g., *What do you think? Do you agree?*) were also counted as they prompted the interlocutor to respond.
2. Responsiveness was defined as any instances where a speaker reacted to the interlocutor by repeating, completing, commenting, or elaborating on previously expressed ideas. Instances of responsiveness encompassed a speaker's mutuality and reciprocity, where speakers scaffolded the conversation, without necessarily providing new content (Storch, 2001). Unlike encouragement by which a speaker prompted the partner to react, responsiveness captured this reaction, expressed either spontaneously or in response to encouragement.
3. Management was defined as instances where a speaker commented on procedural aspects of the task, such as how to divide or organize turn-taking, manage task timing, or approach the discussion. For instance, all episodes in which a speaker read directly from their task handout, reminded their partner of the task if the partner had digressed, or mentioned how much time remained to finish the task were counted under this category.
4. Backchanneling, which signals a speaker's willingness to listen to the interlocutor (Lambert et al., 2017) and provides feedback that facilitates conversation (Bjørge, 2010), was defined as a speaker's response token or turn continuer acknowledging the speaker's utterance (e.g., *uh huh*, *yeah*, *sure*, *I see*, *okay*).

5. Nodding, a visual cue associated with a listener's acknowledgement of the speaker's utterance (Nash, 2007; Whitehead, 2011) and a gestural equivalent of backchanneling (Duncan, 1972), was defined as a visible up-down movement of the head and was counted by observing the interaction videos.

Emotional engagement included one measure which captured a speaker's affective response to the interaction, on the idea that a speaker can be invested in a conversation emotionally, for example, by expressing interest or excitement and showing signs of enjoyment, such as smiling or laughter (Imai, 2010; Philp & Duchesne, 2016).

1. Positive emotion was defined as instances where a speaker either commented on the emotion experienced during interaction (e.g., *that's so funny*) or nonverbally expressed an emotion via episodes of laughter or smiling in the videos (Glenn & Holt, 2013). Because the speakers did not express or show negative emotions, this category only included positive affect. This variable was coded by analyzing both the transcripts and the videos.

Table 1 Examples of Coded Engagement Variables

Variable	***Example***
Idea units	***A: Yeah yeah and then [some people they just love (smiles) the language] then [all their motivation I think is integrative] yeah I know [I know a professor in X] [he knows 26 languages] (both laugh)***
LRE	***B: Maybe some healthy food can be um the… I don't know how to say… the snacks the healthy snacks can be exposed to children***
	A: Yeah. So they don't think of the, what's the word I forgot…
	B: Money?
	A: The cost!
	A: What's the name of the first motivation you you get… it's uh inter-integ…
	B: Integra- integrative
	A: Integrative the second is instra-instra…

	B: Instrumental
Encouragement	***A: Mm hmmm uh a child is more likely to be influenced by it than say a teen or an adult… what do you think?***
	B: Like initially I was learning but then like it's not continued. You understand right? (laughs)
	A: Yeah you stopped (laughs)
Responsiveness	***A: I think…. I think it impacts life dissatisfaction… what do you think?***
	B: I tend to agree.
	A: Uh should advi… advertisers be be allowed to advertise to children?
	B: Uh personally I don't feel this is the right thing.
	A: That's a hundred percent because children will listen to anything you say because they think adults are always right and because it's like it's visual it can appear to them more
	B: Mmm yeah different colors and different sounds can be very attractive.
Management	***B: So what we have to do?***
	A: Uh I think we have to wait for 10 minutes.
	B: Yeah but I I just looked at the question it said undesirable advertising effects.
	A: Yes so… yeah I'm not sure.
Backchanneling	***A: Those lines are like a sort of advertisement***
	B: Exactly (both laugh) come experience a new life (laughs)
	A: So I feel it does life dissatisfaction for sure.

	B: Okay. Uh… I think it makes a nice transition to the next question which is… should advertisers be allowed to advertise to children.
Nodding	***Any instance of nodding (as seen in a video-recording)***
Positive emotion	***B: If you have your job, you earn a high salary then you can integrate*** ***A: Yeah inte—that's funny*** ***plus any instance of smiling or laughter (as seen in a video-recording)***

Note. A and B refer to Speaker A and B in each dyad. Except idea units, which are in brackets, other relevant examples are in boldface.

Following the initial coding, a trained research assistant coded 25% of the data (18 speakers). For frequency counts, coding reliability (Cronbach's α) was .97 for nods, .99 for backchanneling, and .99 for positive emotion. For the remaining categorically coded variables, reliability (Cohen's κ) was .71, indicating substantial agreement (Landis & Koch, 1977). Although all paired interaction were comparable in length (*M* = 11.25 minutes, *SD* = 1.64), the eight coded variables were normalized per speaker by dividing the relevant value by the total number of turns in that speaker's production, so as to account for variation in each speaker's contribution to the dialogue. Therefore, all engagement measures corresponded to the rate with which a given behavior occurred per speaking turn.

4. Results

As shown in Table 2, across the 36 dyads, one speaker in each pair (henceforth, Speaker A) produced approximately 49 speaking turns whereas the other speaker (Speaker B) produced an average of 48 turns, with Speaker A and B designations assigned randomly within each dyad. Speakers A and B also did not differ in terms of the mean comprehensibility ratings that they gave each other (73 vs. 79 on a 100-point scale). With respect to engagement behaviors, idea units were the most frequent category (about three distinct idea units produced every two speaking turns), followed by backchannelling (approximately one instance per turn) and cases of responsiveness and nodding (one instance in every two turns). Displays of positive emotion and management were less common (one instance in every five turns). The least frequently attested were cases of speakers providing each other with encouragement and LREs, with both occurring on average fewer than once in every 10 turns.

As shown in Table 2, across the 36 dyads, Speakers A and B did not differ in displays of engagement, with mean and range values being similar or overlapping between them. A comparison of 95% confidence intervals (CIs) for the mean values across Speakers A and B, which is considered an alternative (and often superior) practice to traditional null hypothesis significance testing (Cumming & Calin-Jageman, 2017), confirmed that all CIs for the mean values in Speaker A versus Speakers B performances were either fully

overlapping or fully subsumed within each other. Thus, insofar as the eight engagement measures were concerned, the two interacting partners showed similar behaviors across all dyads. However, as clearly shown by a broad range of individual values under each category, there was variability in individual speakers' engagement levels. We return to address this variability in a post hoc analysis of engagement behaviors across individual dyads.

Table 2 Speakers' Comprehensibility Ratings and Coded Engagement Variables

	Speaker A			*Speaker B*		
Variable	*M*	*95% CI*	*Range*	*M*	*95% CI*	*Range*
Comprehensibility	*73.33*	*[65.56, 81.10]*	*1–100*	*78.50*	*[73.02, 83.98]*	*39–100*
Idea units	*1.57*	*[1.23, 1.91]*	*0.45–6.27*	*1.78*	*[1.43, 2.12]*	*0.52–5.14*
LREs	*0.03*	*[0.01, 0.04]*	*0.00–0.20*	*0.03*	*[0.01, 0.05]*	*0.00–0.24*
Encouragement	*0.06*	*[0.04, 0.08]*	*0.00–0.26*	*0.08*	*[0.05, 0.11]*	*0.00–0.49*
Responsiveness	*0.46*	*[0.41, 0.50]*	*0.16–0.89*	*0.39*	*[0.33, 0.45]*	*0.05–0.93*
Management	*0.19*	*[0.15, 0.24]*	*0.00–0.42*	*0.19*	*[0.15, 0.23]*	*0.00–0.53*
Backchannelling	*0.95*	*[0.73, 1.16]*	*0.03–3.27*	*0.83*	*[0.67, 0.99]*	*0.03–1.94*
Nodding	*0.65*	*[0.40, 0.90]*	*0.00–3.21*	*0.52*	*[0.31, 0.73]*	*0.00–2.42*
Positive emotion	*0.21*	*[0.14, 0.27]*	*0.00–0.79*	*0.21*	*[0.15, 0.27]*	*0.00–0.67*
Total turns	*48.89*	*[43.41, 54.37]*	*19–85*	*47.78*	*[42.78, 52.78]*	*22–78*

Note. In each dyad, Speaker A's comprehensibility is rated by Speaker B; Speaker B's comprehensibility is rated by Speaker A.

To examine whether L2 speakers' interactional behaviors are associated with peer-ratings of comprehensibility, Spearman correlations (two-tailed) were carried out between each speaker's comprehensibility rating (as assessed by their partner) and engagement variables. These analyses were run

separately for Speakers A and B across the 36 conversational dyads (rather than for the entire sample of 72 speakers) because comprehensibility in interaction might be co-constructed in dialogue and co-dependent on each interlocutor's contribution (Nagle et al., 2021; Trofimovich et al., 2020). Therefore, it was important to determine whether L2 speakers' comprehensibility ratings were correlated with their own engagement measures and with those of their partners. Due to a relatively small sample (n = 36), correlation strength was assessed using the benchmark of .30 to indicate a non-trivial association, which roughly corresponded to 10% of shared variance (R^2 = .10) and the cut-off for a small effect (Plonsky & Ghanbar, 2018).

As summarized in Table 3, for Speaker A (i.e., 36 speakers randomly designated as Speaker A in each dyad), comprehensibility was positively associated with two of their own behaviors: encouragement (.34) and nodding (.34), such that the speakers who provided more encouragement and displayed more frequent nodding elicited higher comprehensibility ratings from their partners. Speaker A comprehensibility did not seem to be linked to any partner behaviors, with all associations falling below .30.

Table 3 Correlations Between Speaker A's Comprehensibility (As Rated by Speaker B) and the Coded Engagement Measures in Each Partner's Performance (n = 36)

	Speaker A's comprehensibility	
Variable	*Speaker A's engagement*	*Speaker B's engagement*
Idea units	*.23*	*–.06*
LREs	*–.07*	*–.03*
Encouragement	*.34*	*–.16*
Responsiveness	*–.03*	*–.01*
Management	*.21*	*–.19*
Backchannelling	*–.12*	*.23*
Nodding	*.34*	*.15*
Positive emotion	*.05*	*.12*
Total turns	*–.14*	*.01*

As shown in Table 4, for Speaker B (i.e., 36 speakers randomly designated as Speaker B in each dyad), comprehensibility was negatively associated with two speaker behaviors: LREs (–.31) and responsiveness (–.34), implying that the speakers who demonstrated more focus on language and who repeated, completed, or commented on their partners' ideas elicited lower comprehensibility ratings from their partners. In addition, Speaker B comprehensibility was also negatively associated with a partner behavior, such that the speakers whose partners produced more LREs were assessed as having lower comprehensibility (–.49), with a medium effect. Appendix B lists the associations between engagement variables within and across speakers.

Table 4 Correlations Between Speaker B's Comprehensibility (As Rated by Speaker A) and the Coded Engagement Measures in Each Partner's Performance (n = 36)

	Speaker B's comprehensibility	
Variable	*Speaker B's engagement*	*Speaker A's engagement*
Idea units	*.12*	*–.05*
LREs	*–.31*	*–.49*
Encouragement	*.26*	*.02*
Responsiveness	*–.34*	*–.04*
Management	*.04*	*.08*
Backchannelling	*–.11*	*–.17*
Nodding	*–.11*	*.03*
Positive emotion	*–.27*	*–.28*
Total turns	*.07*	*–.14*

To determine the relative weight of various engagement variables associated with comprehensibility, two regression analyses were carried out, separately for Speakers A and B. Comprehensibility was the criterion variable, and the engagement behaviors shown by Speakers A and B served as predictors. The predictors chosen for each model only included those that had the potential to influence the criterion variable in each model, based on their non-trivial associations with comprehensibility (> .30). For Speaker A comprehensibility, the predictors were Speaker A's encouragement and nodding (Table 3); for Speaker B comprehensibility, the predictors were Speaker B's LREs and responsiveness and Speaker A's LREs (Table 4). The sample size was deemed sufficiently large for up to three predictors (Field, 2005), and tests of multicollinearity revealed no tolerance values below 0.20 (0.96–1.00) and no VIF values above 10 (1.00–1.04). Analysis of the residuals indicated excellent model fit (only a single case per model had a standardized residual greater than ±2) and no cases with undue influence on the model (Cook's distance values < 1.00).

The regression model for Speaker A comprehensibility yielded a two-factor solution, accounting for 23% of variance (see Table 5). Comprehensibility was predicted positively by encouragement (10%), $F(1, 34) = 5.02$, $p = .032$, and nodding (13%), $F(2, 33) = 6.12$, $p = .005$, such that the speakers who produced more instances of encouragement and nodding elicited higher comprehensibility ratings from their partners. The regression model for Speaker B comprehensibility yielded a single-factor solution, accounting for 15% of variance (see Table 5). Comprehensibility was predicted negatively by their partner's focus on LREs, $F(1, 34) = 7.04$, $p = .012$, such that the speakers whose partners engaged in more LREs were assessed as having lower comprehensibility.

Table 5 Results of Multiple Regression Analyses Using Coded Engagement Measures as Predictors of Each Partner's Performance

Predictors	*R*	*ΔR2*	*B*	*SE B*	*β*	*t*	*p*
Speaker A's comprehensibility							
Constant			*56.94*	*5.79*		*9.84*	*.001*
Encouragement (Speaker A)	*.36*	*.10*	*143.87*	*50.34*	*.43*	*2.86*	*.007*
Nodding (Speaker A)	*.52*	*.13*	*12.07*	*4.77*	*.38*	*2.53*	*.016*
Speaker B's comprehensibility							
Constant			*82.28*	*2.87*		*28.64*	*.001*
LREs (Speaker A)	*.41*	*.15*	*–136.22*	*51.33*	*–.41*	*–2.65*	*.012*

Because the speakers demonstrated a range of frequencies with which they engaged in various interactional behaviors (Table 2) and because Speaker A versus Speaker B comprehensibility was associated with different engagement measures (cf. Tables 3 and 4), our final post hoc analysis targeted speaker engagement across individual dyads so as to capture at least some of this variability. Figure 1 illustrates the frequencies of four behaviors that showed non-trivial associations with comprehensibility—encouragement, nodding, responsiveness, and LREs—separately for each speaker across the 36 dyads. As shown in Figure 1, generally speaking, the strongest predictors of comprehensibility for each speaker were among the least frequently attested behaviors for that speaker (encouragement for Speaker A, LREs for Speaker B) or were those occurring at moderate rates (nodding for Speaker A, responsiveness for Speaker B). In fact, the speaker's behavior associated with their own comprehensibility was often less frequent for that speaker than for their

partner. For instance, Speaker A encouragement predicted Speaker A comprehensibility, yet in 25 of the 36 dyads, Speaker B provided more frequent encouragement than Speaker A. Similarly, Speaker B responsiveness was associated with Speaker B comprehensibility, yet their conversation partners displayed more instances of responsiveness in 22 of the 36 dyads. Thus, the interactive behaviors relevant for comprehensibility were relatively infrequent, yet sufficiently impactful to factor into interlocutor ratings.

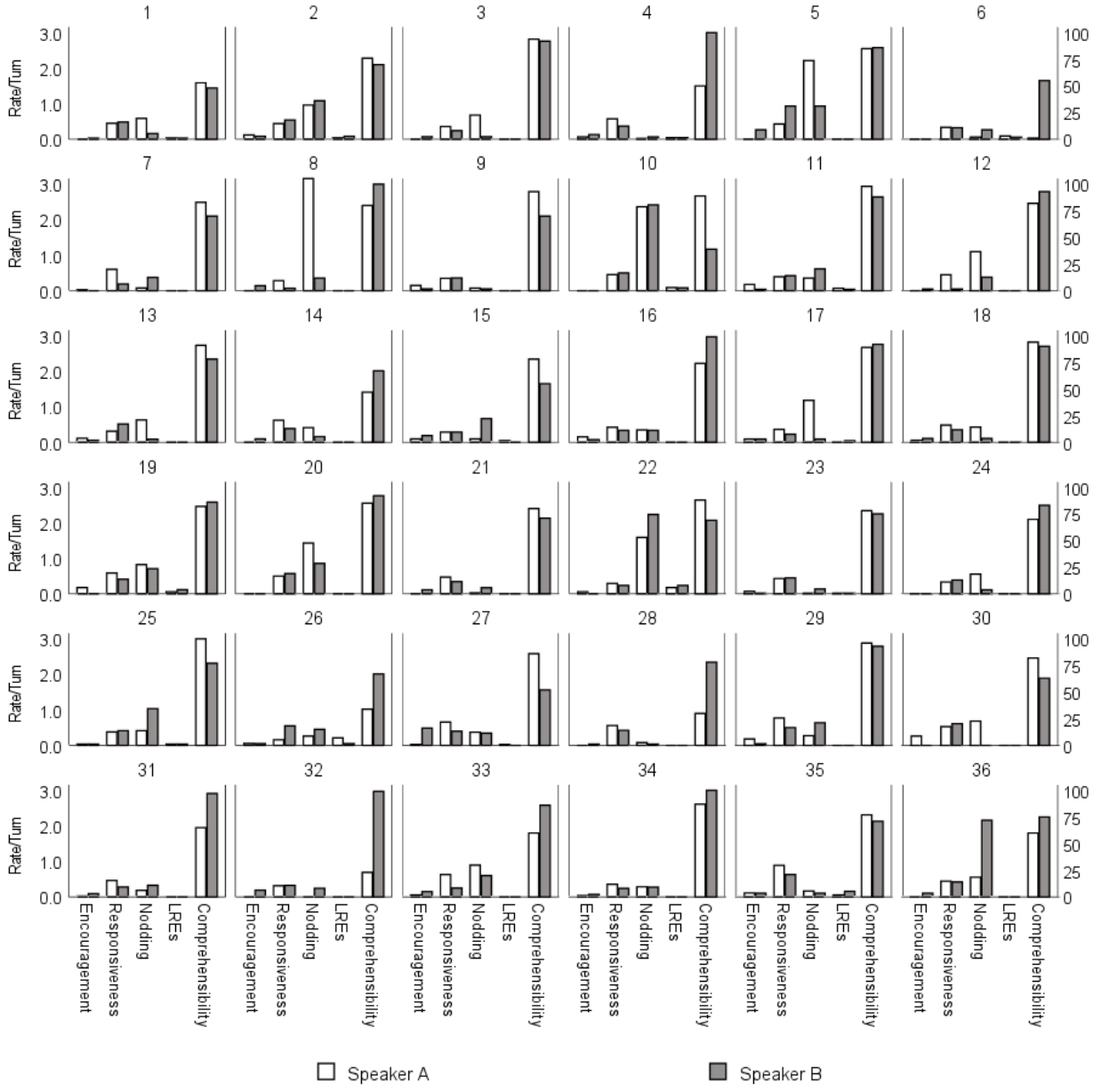

Figure 1. Frequency (rate per speaking turn) of encouragement, responsiveness, nodding, and LREs and comprehensibility rating (0–100 scale) for Speaker A versus Speaker B across the 36 individual dyads.

There was also limited evidence that the two interacting partners demonstrated convergent or otherwise highly interlinked interactive behaviors. As shown visually (in Figure 1) and through corelations between engagement variables (in Appendix B), the speakers demonstrated similar frequencies of LREs (.85) and positive emotions (.73), both with strong associations (Plonsky & Ghanbar, 2018), such that greater engagement by one speaker co-occurred with greater engagement by the other. However, other engagement measures, such as idea units (.31), time and task management (–.34), and nodding (.40), showed weak associations across the partners, and the frequencies of the remaining behaviors were unrelated. Thus, although the speakers were sensitive to each other's interactive behaviors in terms of LREs and displays of positive affect, they generally appeared to have acted independently, demonstrating their own, person-specific behaviors, which might explain why the two partners' comprehensibility was predicted through different engagement measures.

5. Discussion

This exploratory study examined the relationship between L2 speakers' interactional behaviors and their perception of each other's comprehensibility. Across eight engagement measures, four showed associations with comprehensibility which were greater than the benchmark of .30, although these associations differed for different dialogue partners. Speakers who showed more encouragement and nodding (social engagement) were perceived as easier to understand. In contrast, speakers who produced more frequent LREs (cognitive engagement) and demonstrated more responsiveness (social engagement) were rated as harder to understand, although only partner-produced LREs emerged as the significant predictor. Considering that a speaker's language, in terms of its pronunciation, lexis, and grammar, can explain up to 85% of variance in listener-rated comprehensibility, albeit in non-interactive tasks (Isaacs & Trofimovich, 2012; Kang et al., 2010; Saito et al., 2016), these relationships are noteworthy, with 15–23% of variance explained through interlocutor behaviors alone. To our knowledge, this is the first demonstration that L2 comprehensibility is shaped in conversation through speakers' interactional behaviors.

Encouragement was one of two social engagement measures associated positively with comprehensibility. Although instances of encouragement generally occurred infrequently, individual speakers varied widely in this behavior (0.00–0.49), with some expressing no encouragement while others producing one episode every other turn. Because there were no appreciable correlations between comprehensibility and either the frequency of idea units or the number of speaking turns (Tables 3–4), it is unlikely that the speakers increased their comprehensibility by relying on their partner to speak, thus providing less content or engaging in less talk themselves. More plausibly, the encouragement–comprehensibility link might be explained through turn-taking, which refers to how speakers earn and manage their speaking rights in dialogue (Sacks et al., 1974). As the speakers encouraged their partners to claim a turn or sought a response from them, they presumably engaged in partner-sensitive conversation management, where no interlocutor dominated the talk and the speaking and listening burden was equitably distributed (Ducasse & Brown, 2009). This pattern of turn-taking has been previously linked to greater negotiation for meaning that facilitates comprehension (Foster & Ohta, 2005; Pica et al., 1987), so a similar positive link with comprehensibility is not unexpected. Moreover, when the speakers yielded the floor to their partners and encouraged them to continue, they also likely scaffolded the discourse by breaking it into shorter, easier-to-understand chunks or contributed to the

perception of conversational flow (McCarthy, 2010; Sato, 2014), both of which may have enhanced speaker comprehensibility.

Besides encouragement, nodding was also positively associated with comprehensibility. As with encouragement, occurrences of nodding varied across speakers (0.00–3.21), such that some produced no nods while others as many as three per speaking turn. From a linguistic perspective, a speaker's nods may have corresponded to gestural movements aligned with prominence markers, such as when a head nod co-occurs with a stressed word (Pelachaud et al., 1996), which may aid listener comprehension. However, because there were moderate correlations (.45–.55) between the number of idea units produced by one partner and the frequency of nodding by the other partner (Appendix B), nodding was likely content- or idea-driven. Nodding may thus have functioned as a sign of acknowledgement (Nash, 2007; Whitehead, 2011), whereby interlocutors provided each other with supportive visual cues (Bavelas et al., 2002; Knapp et al., 2013) that also allowed them to track each other's comprehension (Aoki, 2011), resulting in a benefit to interlocutor-rated comprehensibility. Finally, from an affective perspective, more frequent head nodding has been attested in non-anxious rather than anxious L2 speakers (Gregersen, 2005). If interlocutors are sensitive to visual signs of anxiety, then a speaker's visual cues associated with relaxed, non-anxious states might reduce the processing burden for the listener, leading to increased speaker comprehensibility. Regardless of the mechanism by which nodding contributes to interlocutor perception, to the best of our knowledge, this is the first reported association between a speaker's visual cue and the speaker's comprehensibility in interaction.

Both LREs and speaker responsiveness showed negative associations with comprehensibility, although only LREs emerged as a significant predictor. In terms of LREs, a negative association was not surprising. Negative correlations between comprehensibility and various issues of grammar and lexis in L2 speech have been reported previously (Isaacs & Trofimovich, 2012; Munro & Derwing, 1995; Varonis & Gass, 1982), and L2 speakers are particularly sensitive to various language problems (Derwing et al., 2002), often finding them distracting or annoying (Fayer & Krasinski, 1987). Even though LREs occurred infrequently and mostly concerned lexical issues, such as a speaker searching for a word, a focus on language (away from the discussion content) was likely sufficiently disruptive to the speaker's message, with consequences for interlocutor-rated comprehensibility. A striking outcome of this study, however, was that the speakers' comprehensibility was predicted by a focus on language in their partners' speech, implying that the impact of LREs as a (negative) predictor is co-constructed by interlocutors. In fact, the frequencies of LREs were highly correlated between the two partners (.85), meaning that they shared language-focused episodes or amplified each other's attention to language. Thus, a tentative conclusion emerging from this dataset is that a focus on language by either conversation partner might have negative consequences for one or both speakers.

As for responsiveness, its negative association with comprehensibility implies that L2 speakers' social engagement in a conversation can help or hinder their comprehensibility. In cases of encouragement, a speaker prompted the interlocutor to take the floor, proceed speaking, or elaborate an argument. Therefore, encouragement likely deflected attention from the speaker, helped scaffold the conversation, promoted supportive listening, and improved conversational flow (Ducasse & Brown, 2009; Sato, 2014), leading to an upgrade in comprehensibility for that speaker. By contrast, in episodes of responsiveness, a speaker

repeated, completed, or commented on previously expressed ideas, without necessarily providing new content. In this case, responsiveness (especially when carried out excessively) implied that the speaker produced socially appropriate yet perhaps task-irrelevant or repetitive talk, with the consequence that the message was harder to understand. It is plausible, then, that contributions of partner-sensitive conversation management to comprehensibility might depend on the right balance between encouragement and responsiveness, in the sense that an interlocutor-sensitive conversational role (e.g., leading the discussion, keeping it relevant, and interjecting politely and purposefully) might be helpful whereas a domineering role (e.g., with frequent interruptions and redundant contributions) might be harmful (Galaczi, 2008; May, 2009, 2011).

In this dataset, all engagement predictors of comprehensibility were relatively infrequent behaviors. It may be that engagement–comprehensibility links are subject to an inverse preference effect, which describes a language processing phenomenon whereby an infrequent structure, such as English passives, tends to have a stronger effect on a speaker's comprehension and production performance when it is heard or seen in recent discourse, compared to a structure that is relatively more frequent, such as English actives (Bernolet & Hartsuiker, 2010; Jaeger & Snider, 2013). The inverse preference effect (or a conceptually similar phenomenon) would be generally consistent with the present findings, in that the relevant interactional behaviors are not ubiquitous but frequent enough to have a bearing on speaker comprehensibility. From this vantage point, the engagement dimensions of comprehensibility might play a qualitatively different role than the linguistic dimensions of comprehensibility, such as lexis, grammar, and pronunciation (Kang et al., 2010; O'Brien, 2014; Saito et al., 2016). Whereas various linguistic properties of L2 speech might directly influence the ease or difficulty with which a speaker is understood, engagement behaviors might shape comprehensibility indirectly, for instance, through framing in negative or positive ways the language produced for the interlocutor. In this sense, a speaker's focus on language through LREs, even if infrequent, would draw attention to a language-related processing difficulty for the interlocutor, while a speaker's nodding or encouraging behaviors, when used sparingly, might positively frame the speaker's utterance, thus lessening the interlocutor's linguistic processing burden. Overly frequent engagement behaviors would likely be less effective—in terms of framing a speaker's utterance—and would therefore be less likely to contribute to comprehensibility. Needless to say, this speculative explanation, which implies that the linguistic and behavioral dimensions of a speaker's interactive performance work in tandem in predicting comprehensibility, must be revisited in future work.

Lastly, the associations between interlocutors' behaviors and their comprehensibility differed for one partner (designated as Speaker A) versus the other partner (Speaker B) in each dyad (cf. Tables 3 and 4). This is not surprising and not unexpected, given that the conversation partners engaged with the same task and their interlocutor in different ways. As shown through the post hoc analysis of individual dyads, there was no extensive evidence, apart from the speakers' focus on LREs and their displays of positive affect, that they engaged in interlocutor-sensitive interactive performance, in the sense that they appropriated and reused each other's engagement behaviors. A most likely reason for inter-individual differences in interlocutor engagement stems from the nature of the speaking task, where the academic discussion—particularly as it occurred in a formal laboratory context in a university setting—promoted individual performances requiring the speakers to demonstrate their own understanding of the source texts and to exchange their personal

opinions. As a result, the speakers approached the task differently, drawing on their individual prior academic experience, background knowledge, and understanding of the task and its content, with the consequence that their interactive behaviors mostly reflected a speaker- rather than a listener-centered performance. Other task types, such as exchanges of personally-relevant information (Lambert et al., 2017; Qiu & Lo, 2017), might elicit more cross-partner alignment in engagement behaviors, with similar or highly comparable consequences for both interlocutors' comprehensibility.

6. Limitations, Implications, and Future Work

The present exploratory findings do not lend themselves to generalizable conclusions. Even though the speaker sample was balanced in terms of gender composition, topic distribution, and task order, the dataset was relatively small (36 dyads) and speaker engagement was analyzed only in one task. Future research should examine larger speaker samples, focusing on how interactional behaviors might contribute to comprehensibility as a function of speakers' gender and cultural backgrounds in different tasks (e.g., academic discussion vs. personal conversation) and across various interactional patterns (e.g., dominant vs. passive) and conversational roles (e.g., information holder vs. receiver). The current findings are also based solely on quantitative analyses, so adding qualitative insights, for example, through interviewing speakers or engaging them in retrospective recall or by having raters evaluate or comment on speaker performances, might provide a more nuanced understanding of the engagement–comprehensibility links. This future work might have implications for both language assessment and teaching practices where teachers or examiners evaluate L2 speakers' paired oral interactions for comprehensibility.

In terms of engagement specifically, future work should more closely examine the interactional behaviors that did not show associations with comprehensibility greater than the .30 benchmark. Idea units, backchanneling, positive emotion, and time and task management did not appear particularly relevant to comprehensibility in this dataset, yet this does not mean that they are inconsequential for interlocutor perceptions of each other. As the occurrence of various engagement behaviors depends on speaker and task characteristics (Baralt et al., 2016; Lambert et al., 2017; Qiu & Lo, 2017; Shin et al., 2016), engagement measures might vary in their relevance to comprehensibility for speakers interacting in different tasks. For instance, time and task management might contribute to comprehensibility in timed tasks or those with a game-playing element, while the breadth and depth of ideas expressed by a speaker might shape comprehensibility in information-exchange tasks. Similarly, displays of emotion might factor into comprehensibility in conversations about personal or emotion-laden topics, whereas backchannels might matter in exchanges where signals of interlocutor comprehension are expected, as in air-traffic communication. Needless to say, associations between comprehensibility and various engagement behaviors must be revisited in future work.

Lastly, it would be important to examine L2 speakers' engagement behaviors and their peer-ratings of comprehensibility from a perspective that embodies discourse as a shared, interdependent, and co-constructed phenomenon (Brennan et al., 2018). Doing so would require researchers to account for social coordination in dialogue (Ackerman & Bargh, 2010), exploring comprehensibility, for example, as interlocutors demonstrate reciprocity and mutuality in conversation (Galaczi, 2008; Storch, 2001), reuse each

other's facial and body movements and speech patterns (Paxton et al., 2016), or align (or diverge) in each other's emotional and affective states (Parkinson, 2011), all as a function of different task demands.

6. Conclusion

As predicted by Varonis and Gass (1982), comprehensibility is shaped by various influences, including pronunciation, grammar, and fluency of L2 speakers' speech. The present findings highlight engagement as yet another potential influence by showing that L2 comprehensibility in paired oral interaction was associated with episodes of encouragement, responsiveness, and nodding (social engagement) and instances of language focus (cognitive engagement). These findings imply that comprehensibility can be determined not only by what speakers say (e.g., in terms of their choice of vocabulary, pronunciation accuracy, or grammar use) but also by what they do (i.e., how they interact with their interlocutors).

From a broader perspective, these findings also substantiate, but not yet fully address, a key question posed by Varonis and Gass (1982): Why do listeners react to L2 speakers in particular ways? When the interlocutors observed by Varonis and Gass heard an L2 speaker ask for directions, they tended to repeat the request (often with a rising intonation), showing some reluctance to get involved in a conversation, often accompanied by a sigh or a filler like *oh boy*. Varonis and Gass argued that these reactions were determined by how comprehensible the L2 speaker sounded, in the sense that, even though the interlocutors understood the speaker's message, their experience was effortful and the likelihood of a future non-understanding was real. Whereas our findings reveal several links between interlocutor-rated comprehensibility and speaker behaviors, our goal was ultimately to predict comprehensibility. It therefore remains for future work to examine whether and to what extent L2 speakers' comprehensibility might determine how interlocutors act toward them.

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.

Aoki, H. (2011). Some functions of speaker head nods. In C. Goodwin, C. LeBaron, & J. Streeck (Eds.), *Embodied interaction: Language and body in the material world* (pp. 93–105). Cambridge University 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.

Bavelas, J., Coates, L., & Johnson, T. (2002). Listener responses as collaborative process: The role of gaze. *Journal of Communication*, *52*, 566–580. https://doi.org/10.1111/j.1460-2466.2002.tb02562.x

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming? *Cognition*, *114*, 455–461. https://doi.org/10.1016/j.cognition.2009.11.005

Bjørge, A. K. (2010). Conflict or cooperation: The use of backchannelling in ELF negotiations. *English for Specific Purposes*, *29*, 191–203. https://doi.org/10.1016/j.esp.2009.04.002

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & de Jong, N. H. (2013). What makes speech sound fluent? The contribution of pauses, speed and repairs. *Language Testing*, *30*, 159–175. https://doi.org/10.1177/0265532212455394

Brennan, S. E., Kuhlen, A. K., & Charoy, J. (2018). Discourse and dialogue. In S. L. Thompson-Schill (Ed.), *The Stevens' handbook of experimental psychology and cognitive neuroscience* (pp. 145–209). Wiley.

Cheng, L., Myles, J., & Curtis, A. (2004). Targeting language support for non-native English-speaking graduate students at a Canadian university. *TESL Canada Journal*, *21*, 50–71. https://doi.org/10.18806/tesl.v21i2.174

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

Cumming, G., & Calin-Jageman, R. (2017). *Introduction to the new statistics: Estimation, open science, and beyond.* 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

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*, 1–16. https://doi.org/10.1017/S0272263197001010

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals*. John Benjamins.

Derwing, T. M., Rossiter, M. J. & Ehrensberger-Dow, M. (2002). They speaked and wrote real good: Judgments of native and non-native grammar. *Language Awareness*, *11*, 84–99. https://doi.org/10.1080/09658410208667048

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, *54*, 655–679. https://doi.org/10.1111/j.1467-9922.2004.00282.x

Ducasse, A. M., & Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction. *Language Testing*, *26*, 423–443. https://doi.org/10.1177/0265532209104669

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, *23*, 283–292. https://doi.org/10.1037/h0033031

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford University Press.

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, *37*, 313–326. https://doi.org/10.1111/j.1467-1770.1987.tb00573.x

Field, A. (2005). *Discovering statistics using IBM SPSS Statistics*. Sage.

Foote, J., & Trofimovich, P. (2018). Is it because of my language background? A study of language background influence on comprehensibility judgments. *Canadian Modern Language Review*, *74*, 253–278. https://doi.org/10.3138/cmlr.2017-0011

Foster, P., & Ohta, A. S. (2005). Negotiation for meaning and peer assistance in second language classrooms. *Applied Linguistics*, *26*, 402–430. https://doi.org/10.1093/applin/ami014

Galaczi, E. D. (2008). Peer-peer interaction in a speaking test: The case of the First Certificate in English examination. *Language Assessment Quarterly*, *5*, 89–119. https://doi.org/10.1080/15434300801934702

Glenn, P., & Holt, E. (2013). *Studies of laughter in interaction*. Bloomsbury.

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

Imai, Y. (2010). Emotions in SLA: New insights from collaborative learning for an EFL classroom. *The Modern Language Journal*, *94*, 278–292. https://doi.org/10.1111/j.1540-4781.2010.01021.x

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. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, *34*, 475–505. https://doi.org/10.1017/S0272263112000150

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.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*, 57–83. https://doi.org/10.1016/j.cognition.2012.10.013

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, *94*, 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Wadsworth.

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

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–74. https://doi.org//10.2307/2529310

May, L. (2009). Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing*, *26*, 397–421. https://doi.org/10.1177/0265532209104668

May, L. (2011). Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly*, *8*, 127–145. https://doi.org/10.1080/15434303.2011.565845

McCarthy, M. (1991). *Discourse analysis for language teachers.* Cambridge University Press.

McCarthy, M. (2010). Spoken fluency revisited. *English Profile Journal*, *1*, 1–15. https://doi.org/10.1017/S2041536210000012

McDonough, K., & Trofimovich, P. (2019). *Corpus of English as a Lingua Franca Interaction (CELFI)*. Montreal, Canada: Concordia University.

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

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*, 111–131. https://10.1017/S0272263106060049

Nagle, C., Trofimovich, P., O'Brien, M. G., & Kennedy, S. (2021). Beyond linguistic features: Exploring the behavioral and affective correlates of comprehensible second language speech. *Studies in Second Language Acquisition*. Published online 23 March 2021. https://doi.org/10.1017/S0272263121000073

Nakamura, S., Phung, L., & Reinders, H. (2020). The effect of learner choice on L2 task engagement. *Studies in Second Language Acquisition*. Published online 13 October 2020. https://doi.org/10.1017/S027226312000042X

Nash, C. E. (2007). Gestural regulators in French, Japanese and American English dialogues. In M. Grein & E. Weigand (Eds.), *Dialogue and culture* (pp. 115–140). John Benjamins.

O'Brien, M. G. (2014). L2 learners' assessments of accentedness, fluency, and comprehensibility of native and nonnative German speech. *Language Learning*, *64*, 715–748. https://doi.org/10.1111/lang.12082

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.

Pelachaud, C., Badler, N. I., & Steedman, M. (1996). Generating facial expressions for speech. *Cognitive Science*, *20*, 1–46. https://doi.org/10.1016/S0364-0213(99)80001-9

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

Pica, T., Young, R., & Doughty, C. (1987). The impact of interaction on comprehension. *TESOL Quarterly*, *21*, 737–758. https://doi.org/10.2307/3586992

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 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

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.

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. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, *37*, 217–240. https://doi.org/10.1017/S0142716414000502

Sato, M. (2014). Exploring the construct of interactional oral fluency: Second language acquisition and language testing approaches. *System*, *45*, 79–91. https://doi.org/10.1016/j.system.2014.05.004

Segalowitz, N. (2010). *Cognitive bases of second language fluency.* Routledge.

Shin, S.-Y., Lidster, R., Sabraw, S., & Yeager, R. (2016). The effects of L2 proficiency differences in pairs on idea units in a collaborative text reconstruction task. *Language Teaching Research*, *20*, 366–386. https://doi.org/10.1177/1362168814567888

Storch, N. (2001). How collaborative is pair work? ESL tertiary students composing in pairs. *Language Teaching Research*, *5*, 29–53. https://doi.org/10.1177/136216880100500103

Storch, N. (2008). Metatalk in a pair work activity: Level of engagement and implications for language development. *Language Awareness*, *17*, 95–114. https://doi.org/10.1080/09658410802146644

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal*, *82*, 320–337. https://doi.org/10.1111/j.1540-4781.1998.tb01209.x

Taylor Reid, K., O'Brien, M., Trofimovich, P., & Bajt, A. (2020). Testing the malleability of teachers' judgments of second language speech. *Journal of Second Language Pronunciation*, *6*, 236–264. https://doi.org/10.1075/jslp.19015.tay

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

van Os, M., De Jong, N. H., & Bosker, H. R. (2020). Fluency in dialogue: Turn-taking behavior shapes perceived fluency in native and nonnative speech. *Language Learning*, *70*, 1183–1217. https://doi.org/10.1111/lang.12416

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

Vaughn, C., & Whitty, A. (2020). Investigating the relationship between comprehensibility and social evaluation. *Journal of Second Language Pronunciation*, *6*, 483–504. https://doi.org/10.1075/jslp.20022.vau

Whitehead, K. A. (2011). Some uses of head nods in "third position" in talk-in-interaction. *Gesture*, *11*, 103–122. https://doi.org/10.1075/gest.11.2.01whi

Zhu, W., & Flaitz, J. (2005). Using focus group methodology to understand international students' academic language needs: A comparison of perspectives. *TESL-EJ*, *8*, 1–11.

Appendix A Background information for speaker pairs

	Speaker A		*Speaker B*		
Pair	*Native language*	*Gender*	*Native language*	*Gender*	*Discussion topic*
1	*Mandarin*	*female*	*Turkish*	*female*	*advertising*
2	*Japanese*	*female*	*Mandarin*	*female*	*advertising*
3	*Farsi*	*female*	*Mandarin*	*female*	*motivation*
4	*Mandarin*	*female*	*Bengali*	*female*	*motivation*
5	*Arabic*	*female*	*Bengali*	*female*	*motivation*
6	*Tamil*	*female*	*Farsi*	*female*	*motivation*
7	*Malayalam*	*female*	*French*	*female*	*motivation*
8	*Russian*	*female*	*Mandarin*	*female*	*advertising*
9	*Arabic*	*female*	*French*	*female*	*motivation*
10	*German*	*female*	*Russian*	*female*	*motivation*
11	*French*	*female*	*German*	*female*	*motivation*
12	*Turkish*	*female*	*Mandarin*	*female*	*motivation*
13	*Mandarin*	*male*	*Turkish*	*male*	*motivation*
14	*Portuguese*	*male*	*Tamil*	*male*	*advertising*

15	*Tamil*	*male*	*Mandarin*	*male*	*advertising*
16	*Danish*	*male*	*Turkish*	*male*	*motivation*
17	*French/Fon*	*male*	*Tamil*	*male*	*motivation*
18	*Hindi*	*male*	*Marwadi*	*male*	*advertising*
19	*Arabic*	*male*	*Spanish*	*male*	*advertising*
20	*Hindi*	*male*	*French*	*male*	*advertising*
21	*Arabic*	*male*	*Tamil*	*male*	*advertising*
22	*Hindi*	*male*	*French*	*male*	*motivation*
23	*Russian*	*male*	*Dutch*	*male*	*motivation*
24	*Urdu*	*male*	*Arabic*	*male*	*advertising*
25	*Mandarin*	*female*	*Hindi*	*male*	*advertising*
26	*Mandarin*	*female*	*Kannada*	*male*	*motivation*
27	*Arabic*	*female*	*Mandarin*	*male*	*advertising*
28	*Tamil*	*female*	*Farsi*	*male*	*motivation*
29	*French*	*female*	*Hindi/Telugu*	*male*	*motivation*
30	*French*	*female*	*Urdu*	*male*	*advertising*
31	*Mandarin*	*female*	*Dutch*	*male*	*motivation*

32	*Arabic*	*female*	*Hindi*	*male*	*advertising*
33	*Yoruba*	*female*	*Mandarin*	*male*	*advertising*
34	*Mandarin*	*female*	*Hindi*	*male*	*advertising*
35	*Tulu/Hindi*	*female*	*Urdu*	*male*	*advertising*
36	*French*	*female*	*Hindi*	*male*	*motivation*

Appendix B Relationships between engagement variables

Within-Speaker Correlations for Engagement Variables (Upper Triangle for Speaker A, Lower Triangle for Speaker B)

Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
1. LREs	*—*	*.08*	*.23*	*–.14*	*–.05*	*.01*	*–.06*	*–.05*
2. Idea units	*–.01*	*—*	*.06*	*–.14*	*.28*	*.05*	*.52*	*.22*
3. Encouragement	*–.35*	*.10*	*—*	*.08*	*.28*	*–.23*	*–.11*	*.05*
4. Responsiveness	*.22*	*.17*	*–.17*	*—*	*–.07*	*–.03*	*–.05*	*.09*
5. Management	*.11*	*.29*	*.20*	*.12*	*—*	*–.04*	*.26*	*–.05*
6. Backchanneling	*.01*	*.16*	*–.01*	*.01*	*.21*	*—*	*.17*	*.25*
7. Nodding	*.24*	*.24*	*–.07*	*.07*	*.18*	*.25*	*—*	*–.01*
8. Positive emotion	*.06*	*.13*	*.16*	*.12*	*.26*	*.20*	*.10*	*—*

Between-Speaker Correlations for Engagement Variables

		Speaker A							
	Variable	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Speaker B	*1. LREs*	*.85*	*.06*	*.27*	*–.07*	*–.07*	*–.03*	*.09*	*–.02*
	2. Idea units	*–.09*	*.31*	*–.26*	*.00*	*.09*	*.60*	*.55*	*.02*
	3. Encouragement	*–.25*	*–.01*	*–.08*	*.03*	*–.13*	*.15*	*–.12*	*.13*
	4. Responsiveness	*.25*	*.16*	*.21*	*.24*	*.03*	*.05*	*.06*	*.15*
	5. Management	*.06*	*.42*	*–.45*	*.04*	*–.34*	*.14*	*.22*	*–.05*
	6. Backchanneling	*–.03*	*.72*	*.13*	*–.10*	*.36*	*.16*	*.46*	*.25*
	7. Nodding	*.35*	*.45*	*–.07*	*–.06*	*.32*	*.03*	*.40*	*–.13*
	8. Positive emotion	*.09*	*.32*	*–.05*	*.12*	*–.17*	*.22*	*.03*	*.73*

Address for correspondence

Pavel Trofimovich

Concordia University

1455 de Maisonneuve Blvd. W.

Montreal, QC H3G 1M8

Canada

pavel.trofimovich@concordia.ca

https://orcid.org/0000-0001-6696-2411

Co-author information

Oguzhan Tekin

Concordia University

oguzhan.tekin@concordia.ca

https://orcid.org/0000-0002-0490-853X

Kim McDonough

Concordia University

kim.mcdonough@concordia.ca

https://orcid.org/0000-0003-3165-9687