Korean Journal of English Language and Linguistics, **Vol 25, November 2025, pp. 1585-1602**
DOI: 10.15738/kjell.25..202511.1585

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

Exploring Raters' Criteria for Assessing Interactional Competence in Paired Discussion Tasks

Sanghyo Goh · Youngsoon So (Seoul National University)

Received: October 15, 2025
Revised: November 15, 2025
Accepted: November 19, 2025

Goh, Sanghyo (First author)
Ph.D candidate, Department of English Language Education
Seoul National University
1 Gwanak-ro, Gwanak-gu, Seoul, Republic of Korea
Email: gkdlel1994@snu.ac.kr

So, Youngsoon (Corresponding author)
Professor, Department of English Language Education, Learning Sciences Research Institute
Seoul National University
Email: youngsoon_so@snu.ac.kr

ABSTRACT

Goh, Sanghyo and Youngsoon So. 2025. Exploring raters' criteria for assessing interactional competence in paired discussion tasks. *Korean Journal of English Language and Linguistics* 25, 1585-1602.

This study explores raters' perceptions of successful interaction in paired discussion tasks within a second language (L2) context, aiming to operationalize interactional competence (IC) in assessment. A qualitative analysis of raters' notes collected during the evaluation process identified three key categories of successful interaction in paired discussion tasks: interactional management, interactive listening, and co-constructive interaction. Raters highlighted the importance of clear and well-reasoned opinion expression, eliciting partners' responses, and providing appropriate and genuine responses to previous utterances as important strategies for managing interaction in paired discussion tasks. Interactive listening reflects the importance of the listener's role in interaction, focusing on how attentively participants listen to each other's utterances. This category encompasses behaviors such as non-verbal cues (e.g., eye contact, nodding) and backchanneling (e.g., "okay", "mm"). Co-constructive interaction was recognized as essential, with raters valuing supportive behaviors such as assisting a struggling partner. While collaborative patterns were generally preferred, dominance was viewed positively when it compensated for a passive partner's lack of engagement. However, dominance driven by excessive self-expression was viewed negatively, highlighting the value of mutual engagement. These findings provide insights into developing rating criteria that capture the interplay of these features, enabling a more comprehensive assessment of IC in paired discussion tasks.

KEYWORDS

interactional competence, rating criteria, raters' perception, paired discussion tasks

1. Introduction

Interactional competence (IC) has gained increasing importance in second language (L2) education and assessment. Unlike communicative competence, which focuses on an individual's linguistic abilities (Bachman 1990, Canale 1983, Hymes, 1972), IC shifts the emphasis toward the ability to co-construct interaction. Defined as the ability to establish and maintain a shared understanding in communication (Kramsch 1986), IC emphasizes the reciprocal and co-constructed nature of interaction, where speakers coordinate their actions to achieve mutual understanding (Hall and Pekarek Doehler 2011, Pekarek Doehler 2019).

Recent research on IC has sought to identify interactional features to understand and operate IC in assessment. Developmental IC studies, which have examined how L2 speakers' IC evolves, have offered a wide range of sequence organizations such as post-expansion (Greer 2016), repair organization (Hellermann 2009, 2011), and preference structure (Al-Gahtani and Roever 2012, 2018, Pekarek Doehler and Pochon-Berger 2011). Similarly, IC assessment studies have identified interactional features like turn organization, back-channeling, response tokens, and adjacency pairs (Ikeda 2017, Youn 2013). While these IC studies have enhanced our understanding of how interactions unfold and established a foundation for developing rating criteria, significant gaps persist in operationalizing IC in language assessments, particularly in identifying the criteria that raters prioritize when evaluating interaction.

Given that raters' evaluations directly impact the assessment of L2 learners' abilities, understanding their perspectives is essential to bridging the gap between theoretical constructs and practical assessment. Previous studies, such as Ducasse and Brown (2009) and May (2011), have explored raters' viewpoints and the criteria they use to assess L2 IC. Ducasse and Brown (2009), for example, examined raters' perspectives using informal interaction tasks, such as discussing personal topics like family. Given the context-dependent nature of IC where different tasks can elicit different interactional features (Vo 2024), it is necessary to investigate raters' views on interaction across a broader range of task types, including those commonly found in academic settings. To address this need, May (2011) investigated raters' perspectives on assessing paired discussion tasks involving students learning English for Academic Purposes (EAP) using a predefined rating scale that focused on three aspects of interactional competence (i.e., understanding the interlocutor's message, responding appropriately, and effectively using communicative strategies). While this study provides valuable insights into the features that raters consider when assessing L2 IC, the use of a predefined rating scale may have influenced raters' decision-making processes, potentially limiting the range of constructs they were able to assess.

Building on this foundation, the present study explores raters' perspectives on what constitutes successful interaction in paired discussion tasks, focusing on intermediate-level Korean English language learners. To avoid constraining evaluations within predefined constructs of IC and encouraging raters to focus on a wide range of factors that contribute to successful interaction, this study uses the term 'interaction quality' to elicit raters' perceptions without relying on predefined rating scales. This open-ended approach enables a more comprehensive exploration of the features raters prioritize, offering deeper insights into the aspects of interaction they consider essential for effective communication. Raters were encouraged to note any features they found salient when evaluating interaction quality,and thematic analysis was conducted to identify features raters prioritize in assessing interaction quality. By identifying these features, this study seeks to contribute to the development of more empirically grounded and accurate rating criteria for IC assessment.

This study addresses the following research question: What features do raters pay attention to when assessing the effectiveness of interactions in paired discussion tasks?

2. Literature Review

2.1 Interactional Competence in Second Language Assessment

Interactional competence (IC) emerged as an extension of traditional definitions of communicative competence, addressing the interactive nature of communication between interlocutors. While communicative competence is typically defined as an individual's language abilities encompassing grammatical, sociolinguistic, and discourse competencies (Bachman 1990, Canale 1983, Hymes 1972), IC focuses on a mutual understanding of communication, introducing the concept of 'intersubjectivity,' where participants co-construct meaning (Kramsch 1986). Hall and Pekarek Doehler (2011) define IC as "the ability to accomplish meaningful social actions, to respond to co-participants' previous actions, and to make recognizable for others what our actions are and how these relate to their own actions" (p.1). This definition underscores not only an individual's communicative competence but also the reciprocal and context-dependent nature of communication. Thus, this reconfiguration of communicative competence among multiple participants focuses on the collaborative use of linguistic resources to achieve communicative goals (Pekarek Doehler 2019).

Building on this conceptual framework, IC research has sought to identify interactional features by examining the development of L2 speakers' IC over time and the strategies they employ in interactions. Developmental studies have shown how features such as turn-taking (Cekaite 2007, Galaczi 2014), repair organization (Hellermann 2009, 2011), post-expansion (Greer 2016), and preference organization (Al-Gahtani and Roever 2012, 2018, Pekarek Doehler and Pochon-Berger 2011) evolve as their L2 proficiency increases. For instance, Galaczi (2014) demonstrated that learners with higher L2 proficiency manage turns more effectively and use backchanneling not simply as fillers but as tools for maintaining interaction flow. Specifically, the frequency of backchanneling increased with proficiency until reaching the B2 level in the Common European Framework of Reference for Languages (CEFR), after which it decreased. More proficient participants, particularly those at levels C1 and C2, often used confirmations of comprehension, such as "absolutely" and "exactly," rather than backchanneling. These findings suggest that backchanneling may serve as a strategy to mask comprehension difficulties at lower proficiency levels, while more advanced learners employ substantive confirmations to facilitate interaction.

Furthermore, Al-Gahtani and Roever (2018) investigated how learners' strategies for managing refusals evolved with increased proficiency. They found that as proficiency increased, learners expanded their range of interactional methods to delay and soften refusals effectively. Low-proficiency learners tended to use direct and unmitigated refusals such as "no." In contrast, intermediate-level learners adopted more nuanced strategies, including "yes, but" constructions and other refusal formats. These strategies often involved pre-accounts, such as "you know what the problem is," to provide explanations, as well as insert expansions like "Do you have a problem with your car?" to elicit additional information. These developments illustrate learners' growing ability to delay and soften refusals in interaction.

Similarly, IC assessment research has identified features such as adjacency pairs, appropriate pauses, backchanneling, and relevance to the next turn (Ikeda 2017, Youn 2013) as foundational for developing rating rubrics to assess IC. Adjacency pairs (e.g., question-answer sequences), a significant indicator in assessing IC, demonstrate participants' ability to produce contextually appropriate responses. Furthermore, Youn (2020) quantified interactional features and examined their influence on raters' judgments of interaction quality, showing how these features correlated with higher interaction quality. The study highlighted the impact of specific interactional features, such as turn length, number of turns, and sequential organization, on the interaction quality.

Longer turns, for example, were associated with higher-quality interactions, reflecting a speaker's ability to provide detailed and coherent responses. Learners who demonstrated higher-quality interactions were also more effective in turn management, using extended responses and appropriate discourse markers, particularly when handling dispreferred actions like refusals or disagreements. These findings underscore the importance of specific interactional features in defining and assessing L2 IC and provide valuable insights into how such features might be evaluated during assessments.

While these studies identify interactional features through the analysis of spoken discourse, little is known about what raters prioritize when evaluating interaction quality during the assessment process. This raises important questions about whether the identified features are consistently observable, relevant, and meaningful in raters' perceptions. Since raters' evaluations directly influence decisions about test-takers' abilities, exploring their perspectives is essential. Such insights can inform the development of more accurate and comprehensive rating rubrics that capture the complexity of IC in L2 settings, ultimately enhancing the validity and reliability of IC assessments.

2.2 Raters' Perspectives on Interaction

Raters' perspectives are crucial as their assessments significantly influence decisions regarding language learners' competencies. Given the critical role of raters in assessment, understanding the criteria they employ is essential for enhancing assessment practices. Moreover, the multifaceted nature of the interaction, which encompasses linguistic, pragmatic, and sociocultural dimensions, underscores the need for further research to deepen our understanding of how raters assess interaction.

Previous studies have explored raters' perspectives on what constitutes successful interaction. Ducasse and Brown (2009), for example, focused on beginner-level Spanish learners participating in peer-to-peer interaction tasks, such as talking about their families. They identified three key interaction parameters: non-verbal interpersonal communication, interactive listening, and interactional management. Non-verbal interpersonal communication involves the use of gaze and body language, such as gestures. Interactive listening entailed assessing whether candidates demonstrated attention and engagement while listening, for example, by helping an interaction partner find words, filling silences, or providing verbal affirmations to support the conversational flow. Interactional management reflected raters' views on how candidates handled the conversation, including aspects of turn-taking and topic handling. However, as these findings were based on casual, informal conversations, further research is needed to determine their applicability to other types of discourse, such as paired discussions commonly used in academic contexts.

May (2009) investigated raters' perspectives on interaction, particularly focusing on the concept of 'co-construction.' The study found that raters assessed mutual achievements as key features of interaction, suggesting that shared scores could acknowledge the co-constructed nature of interaction in assessments. Building on this framework, May (2011) investigated interaction among 12 EAP students in China through paired discussion tasks. This study utilized a rating scale that focused on three aspects of IC: understanding the interlocutor's message, responding appropriately, and using communicative strategies effectively. The rating scale included descriptors such as: "Unable to understand the interlocutor's message. Often cannot respond. Inadequate use of communicative strategies" (p.132). While May (2011) provided valuable insights into raters' perspectives on interaction in paired discussion tasks, the use of a predefined rating scale may have influenced their decision-making processes, potentially restricting the range of constructs they considered. This limitation underscores the need for open-ended approaches to explore raters' perspectives more fully.

Dai (2022) addressed such limitation by adopting an open-ended approach that allowed raters to articulate their own criteria without predetermined frameworks. The study elicited rating criteria from domain experts, who are knowledgeable about interactional norms in the target language use (TLU) domain though not necessarily linguistically trained, when assessing role-play tasks. This approach enabled the identification of indigenous rating criteria that might not have emerged through predefined scales. However, given the context-dependent nature of IC where different task formats elicit different rating criteria (Vo, 2024), further research is needed to examine how raters' perspectives vary across different task types. Such research could enable the identification of both task-specific and generalizable assessment criteria, supporting the development of empirically grounded rating criteria that better reflect the complexity of interaction.

3. Method

3.1 Participants

3.1.1 English language learners

Thirty Korean English language learners participated as test-takers in the paired discussion tasks. The group included 24 males and six females, all of whom were undergraduate or graduate students with Korean as their first language. To control for variability in language proficiency, participants were selected based on their TEPS scores ranging from 268 to 386. This range corresponds to an intermediate level of communicative competence as defined by TEPS[1] and was chosen to ensure that participants could perform tasks in general fields without significant difficulty. By limiting proficiency variability, the study aimed to provide a more focused examination of IC that was independent of participants' overall language skills.

3.1.2 Raters

Five raters, all non-native English speakers with advanced English proficiency, were recruited to explore their perceptions of what constitutes successful interaction. Despite being non-native speakers, their advanced English proficiency, demonstrated by TEPS scores above 453 (indicative of near-native or native-like communicative competence), ensured their ability to comprehend English conversations and evaluate the quality of interaction.

All raters either held or were pursuing a master's degree in English language education and had experience teaching English to Korean English language learners. Their academic background and teaching experience qualified them to evaluate interaction and provide valuable insights into what constitutes successful interaction.

3.2 Data Collection

3.2.1 Paired discussion task

Paired discussion tasks were employed to examine English language learners' IC in an academic setting. These tasks were selected to reflect the interactional demands commonly encountered in English as a Foreign Language

[1] https://www.teps.or.kr/Info/Teps#

(EFL) university classrooms, where paired discussions are frequently used to promote student interaction.

The discussion tasks involved two topics: (a) "Living in a dormitory is preferable to living off-campus," and (b) "University lectures should be conducted in English to enhance students' global competence." These topics were chosen for their relevance to the participants' experiences as students and were designed to provoke meaningful discussions.

Participants provided informed consent and were randomly paired to minimize interlocutor effects. Each session took place in a quiet room, where participants were given five minutes to review the two discussion prompts and ask questions about the tasks (see Appendix). Each discussion lasted ten minutes, with the two discussions conducted consecutively. The entire session, including preparation time, lasted approximately 15–20 minutes. All performances were video-recorded for subsequent analysis.

3.2.2 Raters' notes

The speech performances collected from 30 test-takers were used as stimuli to elicit raters' perspectives on interactional competence. Raters were instructed to rate speech performances based on their intuitive understanding of interactional success and freely write down the rationale behind their intuition for the rating. Raters considered the two discussion tasks completed by each pair of test-takers collectively, providing a single observation note per pair. As a result, a total of 75 observation notes were collected from five raters across 15 pairs of test-takers.

To encourage raters to consciously reflect on salient interactional features during their evaluations, they were provided with the following two guided questions:

1. What key observations did you make when assessing the interaction quality?
2. For performances you considered to have high interaction quality, were there any characteristics that stood out? Conversely, for performances you considered to have low interaction quality, were there any distinguishing characteristics?

These questions were intended to capture a comprehensive understanding of the features that raters focus on, ensuring a thorough exploration of the nuances that differentiate high-quality from low-quality interactions. By using the term 'interaction quality,' this study sought to avoid leading raters to predefined theoretical constructs of IC that might influence their observations. Instead, it encouraged raters to focus on a broader range of factors contributing to successful interaction, potentially uncovering aspects not traditionally associated with IC in existing literature.

3.3 Data Analysis

After collecting raters' notes, the researcher conducted a thematic analysis (Dai 2022, Pill 2016) to identify the rating criteria that raters employed when assessing paired discussion tasks. The first author of this paper conducted multiple thorough reviews of the raters' notes to identify interactional features that raters attended to during their evaluation. The notes were then segmented into meaningful units, with each unit representing a distinct observation or comment about interaction. When a single sentence in the notes contained multiple distinct features, each feature was separated and coded individually.

The data were organized and analyzed using MAXQDA software (VERBI Software 2021). To ensure reliability,

30 percent of the data was co-coded by an independent third party, resulting in an agreement rate of 94%, indicating a high level of reliability. Any discrepancies between the researcher and the co-coder were discussed, and revisions to the coding scheme were made as necessary. This coding process resulted in the following coding scheme.

Table 1. Coding Scheme for Interactional Features

Feature	Description
Backchanneling	Using verbal acknowledgments (e.g., "aha," "I see," "mm," etc.)
Collaboration	Engaging in interaction that emphasizes working together effectively
Constructive dominance	Taking control of the discussion to compensate for a passive partner
Disruptive dominance	Taking control of the discussion by prioritizing self-expression over collaborative engagement.
Eliciting the partner's response	Eliciting the partner's responses (e.g., inviting partner's opinions, using questions to develop topics)
Helping a partner	Assisting the partner during moments of difficulty or helping to maintain interaction
Non-verbal cues	Employing non-verbal cues such as nodding and eye contact
Opinion expression	Expressing ideas and opinions with supporting rationale
Responding to the previous utterance	Reacting appropriately to the partner's previous utterance

After identifying all interactional features noted by the raters as summarized in Table 1, these features were grouped into categories based on their functional similarities and their roles in interaction. Table 2 presents the categorization of interactional features. This categorization abstracts related features and indexes different aspects of IC as perceived by raters.

Table 2. Categorization of Interactional Features

Category	Feature
Interactional management	Eliciting the partner's response
	Opinion expression
	Responses to the previous utterance
Interactive listening	Backchanneling
	Non-verbal cues
Co-constructive interaction	Collaboration
	Constructive dominance
	Disruptive dominance
	Helping a partner

In the following section, the results of the study are discussed within the framework of these categories, with raters' comments used to illustrate key interactional features. Although all raters' comments were thoroughly analyzed, the level of detail in their comments varied. This variation influenced the selection of representative samples for each interactional feature, with more detailed comments prioritized to best illustrate the findings in this paper.

4. Results and Discussion

This study examines raters' perspectives on what constitutes successful interaction in paired discussion tasks to identify the criteria they employed during evaluations. To achieve this, raters' notes were analyzed to extract key features, which were then grouped into broader categories based on shared themes. Three categories emerged: 'Interactional management,' 'Interactive listening,' and 'Co-constructive interaction.' These categories represent distinct aspects of interaction that raters emphasized in their evaluations, providing the foundation of potential IC rating criteria in paired discussion tasks. The following sections discuss each category in detail, with excerpts to demonstrate how these features were evaluated by raters.

4.1 Raters' Criteria and their Categorization

4.1.1 Interactional management

This category focuses on how test-takers manage interactions, including expressing their opinions, eliciting their partners' responses, and responding to their partners' utterances during discussions. Since the tasks are designed to assess test-takers' ability to sustain interaction in discussion-based settings, opinion expression plays a central role in fulfilling the task goals. Raters consistently attended to features in test-takers' performance that reflected their ability to articulate opinions clearly and support them with relevant reasoning. The following excerpts illustrate how clarity and reasoning in opinion expression directly influence interactional quality.

> Excerpt 1
> "Fluently **expressed their opinions with appropriate reasoning**, minimizing pauses between turns." (Rater 3, Pair 3[2])

> Excerpt 2
> "It seemed that Participant B **was unable to convey their opinions clearly**, resulting in **asymmetrical participation** in the conversation." (Rater 2, Pair 8)

As shown in Excerpt 1, expressing opinions fluently and logically contributes to maintaining interactional flow. In contrast, Excerpt 2 shows that unclear or incomplete opinion expression can disrupt interaction by leading to asymmetrical participation. Such difficulties were perceived negatively by raters not only because they indicated limited ability to express opinions but also because they undermined the collaborative nature of the conversation.

Raters' comments further suggest that opinion expression is intertwined with broader interactional patterns rather than functioning as an isolated skill. While articulating one's own opinions is essential for completing the task, an overemphasis on self-expression without attending to the partner's ideas can hinder interactional balance. Participants who dominated the discussion by focusing solely on their own viewpoints were perceived as less effective in sustaining mutual engagement and conversational flow, as discussed under the category of 'Co-constructive interaction' (see 4.1.3.).

In addition to expressing their own opinions, raters valued participants' ability to draw their partners into the

[2] The rater number and pair number, shown in parentheses, indicate which rater evaluated the discussion of which pair of participants.

discussion. Participants who elicited their partners' responses by explicitly inviting opinions or asking questions were viewed as facilitating mutual involvement in the interaction. Such strategies were salient in raters' evaluations because they not only helped maintain conversational flow but also expanded interaction by prompting partners to elaborate and co-construct ideas.

> Excerpt 3
> "Leading the conversation by **asking for opinions** like 'Why do you think so?'" (Rater 1, Pair 13)

> Excerpt 4
> "**Participant A draws B's opinion by asking**, 'What do you think about that?'" (Rater 3, Pair 15)

Beyond inviting their partners' viewpoints, participants who used probing or personalized questions were also viewed favorably by raters for deepening engagement.

> Excerpt 5
> "**Asking probing questions** such as 'How about money?' elicited more in-depth responses and was a key factor in my evaluation." (Rater 1, Pair 14)

> Excerpt 6
> "Participants fostered deeper engagement by posing **personal questions relevant to the ongoing topic.**" (Rater 3, Pair 4)

Such questions not only sustain conversational flow but also expand the scope of the discussion. For example, a question like "How about money?" invites a deeper exploration of the topic, prompting the conversation partner to elaborate. Similarly, questions that connect to the partner's experiences or interests encourage more meaningful and personalized discussion. Such strategies enable participants to develop sophisticated and engaging discussions from simple conversational threads, a feature that raters consistently associated with successful interaction.

The ability to develop sophisticated and engaging discussions from simple conversational threads contributes to raters' perceptions of successful interaction. Skillfully developing topics is crucial for maintaining a dynamic and engaging dialogue, as noted by Ducasse and Brown (2009) in their study of interaction quality in informal tasks. This holds true not only in informal contexts, as found in Ducasse and Brown (2009), but also in formal, academic tasks, as demonstrated in the present study. The consistent emphasis on these interactional management strategies across diverse studies underscores their importance in varied communicative settings.

Finally, raters emphasized the importance of responding appropriately to partners' previous utterances as another critical component of interactional management. Responding to previous utterances reflects both the listener's role—attending to and understanding their partner's contribution—and the speaker's role—providing contextually relevant replies that sustain and develop the conversation. The following excerpts illustrate that effective interactions require participants to maintain conversational flow through appropriate responses to their partners' previous statements.

Excerpt 7
"Participant A sustained the conversation flow by **referencing prior statements**, contributing to meaningful interaction" (Rater 4, Pair 10)

Excerpt 8
"Interlocutors exhibited **appropriate responses to each other's comments**, facilitating a smooth conversational flow." (Rater 5, Pair 5)

Raters further emphasized that it was not merely the presence of a response that mattered, but its depth and authenticity. Responses that lacked genuine engagement or felt mechanical were perceived as less effective and indicative of lower interaction quality. Excerpts 9 and 10 illustrate raters' critiques of such responses.

Excerpt 9
"Although the task of expressing opinions was met, responsiveness to each other's remarks was lacking. **Responses seemed mechanical and disengaged**." (Rater 5, Pair 5)

Excerpt 10
"I think Korean English learners **tend to say 'I agree' a bit habitually**. It seems to be a customary way to finish the other person's sentence and to be polite. They don't delve deeper into the other person's opinion." (Rater 1, Pair 9)

As illustrated in Excerpts 9 and 10, mechanical or habitual responses detracted from the perceived authenticity of the interaction. Such responses were seen as signaling superficial engagement or limited processing of the partner's ideas, which negatively affected interaction quality. In contrast, genuine and thoughtful responses that acknowledged and elaborated on the interlocutor's ideas were viewed as indicators of strong interactional competence.

These findings extend the observations made by May (2011), who emphasized the role of appropriate responses in assessing IC in paired discussion tasks. While May (2011) focused primarily on turn-taking and relevance, the current study reveals that raters also attend to the authenticity and meaningfulness of engagement with a partner's contributions. This broader perspective highlights that interactional quality is shaped not only by the presence of responsive behavior but also by its depth—how genuinely participants understand, acknowledge, and build upon their partners' ideas.

Overall, the findings suggest that rating scales for IC should capture a range of interactional management abilities that collectively sustain discussion. First, they should include criteria assessing the clarity and reasoning of opinion expression, especially in paired discussion tasks where participants are required to express, negotiate, and develop ideas to achieve communicative goals. Second, they should recognize the ability to elicit the partner's responses which encourages mutual involvement and topic development. Third, scales should evaluate not merely the presence of responses but their quality and authenticity, reflecting the extent to which test-takers engage meaningfully with their interlocutors. By incorporating them, rating rubrics can more accurately capture the interactive skills necessary for successful interaction in paired discussion tasks.

4.1.2 Interactive listening

This category reflects how attentively participants listen to each other's utterances, emphasizing their role as listeners. It encompasses behaviors such as non-verbal cues (e.g., eye contact, nodding, and facial expressions) and backchanneling (e.g., "okay," "mm-hmm") to signal attentiveness and active engagement with a partner's utterances. Raters evaluated interactions positively when participants displayed clear signs of attentive listening and comprehension, which fostered mutual engagement.

The following excerpt demonstrates how non-verbal cues were perceived positively by raters in evaluating the quality of the interaction.

> Excerpt 11
> "Participant B's **nonverbal cues conveyed active engagement**, enhancing the overall conversational flow." (Rater 3, Pair 14)

When these cues are lacking or inadequate, the consequences can be detrimental to interaction. Insufficient non-verbal cues are interpreted as a lack of interest or involvement, thereby undermining the perceived effectiveness of the interaction, as noted below:

> Excerpt 12
> "The content of the conversation is not problematic, but from the observer's perspective, it prompts me to think about the importance of voice tone and **non-verbal signals (such as eye contact, nodding, and responses)** in interaction." (Rater 3, Pair 15)

> Excerpt 13
> "Participant A's **lack of eye contact** suggested **disengagement** from the conversation." (Rater 2, Pair 3)

The importance of non-verbal cues aligns with the findings of Ducasse and Brown (2009), who emphasized the role of gaze and other non-verbal signals in maintaining interactional flow. However, there is a notable distinction between the two studies. Ducasse and Brown (2009) identified the negative impact when speakers over-rely on non-verbal cues, particularly as a means of compensating for limited linguistic abilities, which could hinder communication. In contrast, raters in the present study focused on the negative impact of insufficient non-verbal engagement, with no observations reported regarding the negative effects of excessive use of non-verbal cues.

This difference may be attributed to the varying proficiency levels of participants in the two studies. The beginner-level learners in Ducasse and Brown (2009) relied heavily on non-verbal cues, possibly due to limited linguistic resources. In contrast, the participants in this study, presumably of higher proficiency, were expected to engage more in verbal interaction rather than rely on non-verbal cues to convey meaning. Consequently, excessive reliance on non-verbal cues to compensate for limited linguistic resources was not perceived by raters in this study. Instead, non-verbal cues were viewed as a strategy to show attentiveness, leading raters to consider insufficient non-verbal cues negatively.

This different purpose in the use of non-verbal cues can also be seen in other interactional features. Galaczi (2014), for instance, observed that lower-level learners use backchanneling to mask comprehension difficulties, while more advanced learners employ backchanneling not simply as fillers but as tools to facilitate interaction. Both non-verbal cues and backchanneling serve different roles at varying proficiency levels, highlighting their

changing functions from compensation to enhancement as proficiency increases. This transition underscores the need to consider how the purpose of using these interactional features may vary across proficiency levels.

In addition to non-verbal cues, backchanneling was mentioned in raters' comments as an important indicator of interactional engagement. These verbal cues signal that listeners are following the conversation and encourage their partner to continue speaking, thereby maintaining interactional flow.

> Excerpt 14
> "Participant A's **responsive cues, such as nodding and verbal affirmations (e.g., ah, um, hm)**, encouraged the speaker to continue sharing." (Rater 3, Pair 10)

Excerpt 14 illustrates that both non-verbal cues and backchanneling are essential for fostering effective interaction. Raters did not limit their focus to just one type of cue; instead, they recognized and valued the combined effect of these features in signaling attentive listening. Verbal and non-verbal cues work together to convey engagement, facilitating smoother, more meaningful conversations. While they differ in form—one being verbal and the other non-verbal—they share the common purpose of supporting the speaker and ensuring smooth communication, even though they do not contribute substantive content.

These two features collectively form the foundation of 'Interactive listening,' reflecting a participant's ability to sustain interaction through attentiveness and to actively engage with their partner's utterances. This underscores the need for rating criteria that capture these features.

4.1.3 Co-constructive interaction

'Co-constructive interaction' emphasizes mutual achievement and supportive behaviors which participants make efforts to engage cooperatively in the task, focusing on the overall interaction pattern rather than specific strategic behaviors to maintain and expand interaction. This category includes the features of *collaboration*, *constructive dominance*, *disruptive dominance* and *helping a partner*, all of which relate to balanced participation and mutual engagement.

Collaboration reflects the participants' ability to work together effectively, emphasizing shared responsibility in achieving the task's goals. This feature aligns closely with Kramsch's (1986) concept of 'intersubjectivity,' where participants co-construct meaning through their interactions. Raters frequently used terms such as "collaboration," "cooperative," "asymmetric," or "dominant," to describe the dynamics they observed. Among these, collaboration was particularly valued as an indicator of high-quality interaction, as demonstrated in the following excerpts:

> Excerpt 15
> "I observed whether the speakers interact **collaboratively**." (Rater 1, Pair 3)

> Excerpt 16
> "Open to each other's ideas, working **cooperatively** on the task. **Equal participation** with no one **dominating**." (Rater 3, Pair 11)

These excerpts reflect raters' emphasis on cooperative engagement and shared responsibility in the conversation, marking these as signs of successful interaction. This emphasis on mutual achievement aligns with previous

findings, underscoring the importance of co-constructed dialogues where participants fluidly alternate between listener and speaker roles (Galaczi 2008, May 2009).

A key aspect of collaborative interaction is the interaction in which participants assist each other when communication challenges arise. This support, which can take various forms such as helping with word searches or confirming understanding when a partner struggles, exemplifies the co-construction of meaning. Raters valued these behaviors for their role in maintaining cooperative interaction. The following excerpts illustrate how raters appreciated participants' efforts to understand and assist each other in overcoming communication challenges.

> Excerpt 17
> "Rated it quite high because the participants made significant efforts to understand each other. For example, they **provided answers to each other's word searches**." (Rater 5, Pair 11)

> Excerpt 18
> "Actively **helping others when their partner was stuck** was perceived positively by me." (Rater 1, Pair 8)

These instances of supportive behaviors underscore the importance of collaboration in IC assessment, highlighting how effective interaction relies not just on individual linguistic abilities, but on the participants' efforts to maintain interaction. These findings reinforce that helping a partner, particularly during moments of difficulty, is a valuable strategy for sustaining collaborative interaction and ensuring that both participants remain engaged in the conversation. These observations suggest that IC assessment frameworks should explicitly include criteria for evaluating how test-takers work together to achieve communicative goals.

While collaborative talk is often recognized as the preferred interaction pattern, dominance does not necessarily detract from interaction quality. Raters acknowledged that in some instances, one participant assuming a dominant role could facilitate interaction—particularly when the dominance emerged from a passive partner. This type of dominance, referred to as *constructive dominance,* was viewed as an essential strategy for maintaining interaction when one participant was less active. Excerpt 19 shows that raters appreciated efforts to sustain collaboration, even when one participant took a leading role.

> Excerpt 19
> "The girl **dominated the interaction** as her partner didn't perform well, but **it didn't lower my assessment of the interaction quality.**" (Rater 1, Pair 12)

This example demonstrates how dominance can be viewed positively when one participant takes responsibility for sustaining the conversation, such as by guiding the discussion or posing questions to ensure their partner stays engaged. Raters' positive view of this type of dominance likely stems from recognizing its practical benefits as it prevents communication breakdowns, encourages participation from less confident speakers, and helps maintain the flow of interaction.

In EFL classrooms where students may vary in participation due to factors such as lack of confidence or limited linguistic resources, constructive dominance becomes particularly valuable for ensuring continued interaction. Given raters' background as English teachers in EFL contexts, they may be especially attuned to this feature in assessing interaction. This perspective underscores the complex nature of effective communication, highlighting the importance of adapting interaction patterns to accommodate a partner's passivity.

However, it's crucial to distinguish constructive dominance from *disruptive dominance*, stemming from

excessive focus on one's own opinions without engaging the partner, detracts from interaction quality. Galaczi (2008) observed two types of dominance in spoken discourse: one marked by aggressive interruptions and overlaps, and another stemming from a partner's passivity. Raters in this study demonstrated sensitivity to these distinctions, highlighting their relevance in assessing interaction quality. This alignment between observable features in performances and raters' judgments affirms its value as an essential aspect to consider in IC assessment. While raters view dominance that emerged in response to a passive partner more favorably, as illustrated in Excerpt 19, dominance that focused on expressing personal opinions at the expense of mutual engagement was critiqued more negatively. Excerpts 20 and 21 exemplify raters' perspectives on such interactions.

> Excerpt 20
> "Interlocutors appeared **more focused on expressing their own opinions** rather than engaging with each other." (Rater 1, Pair 5)

> Excerpt 21
> "Participant A seemed **more focused on presenting his own opinion** with reading the prompt, rather than engaging in interaction." (Rater 3, Pair 5)

Although raters acknowledged the importance of participants expressing their opinions, they often critique cases where participants focused excessively on their viewpoints. Such interactions often devolve into monologues rather than dialogues, with participants speaking *at* each other rather than *with* each other. This behavior detracted from interaction quality, indicating a lack of consideration for their interlocutor. These observations suggest that raters' views on dominance varied depending on their source. While dominance was sometimes viewed as necessary for maintaining conversation particularly when compensating for a partner's passivity, raters criticize cases where dominance was rooted in excessive opinion expression, which hindered collaborative engagement. These findings highlight the need for assessment descriptors that differentiate between positive dominance where a participant takes the lead to compensate for a passive partner while still fostering collaboration and negative dominance where a participant monopolizes the conversation without engaging their partner.

Overall, the findings show that successful interaction may require not only collaborative efforts but also the ability to constructively lead the interaction when necessary. To capture these dynamics effectively, assessment descriptors should emphasize collaborative behaviors and distinguish the source of dominance, recognizing that dominance can serve as a helpful strategy for maintaining interaction when it compensates for a passive partner.

4.2 Rating Criteria for Paired Discussion Tasks

The categories identified in this study provide a foundation for IC rating criteria for paired discussion tasks. By grouping individual features into broader categories, this framework aligns with raters' perceptions and highlights distinct aspects of IC that are critical for evaluation. Table 3 summarizes the findings, showing these three categories along with their associated features as identified through the analysis.

Table 3. Summary of Rating Categories and Description

Category	Description
Interactional management	Reflects the ability to manage interaction in paired discussion tasks. Includes expressing well-reasoned opinions, eliciting a partner's responses to sustain and expand discussion, and responding genuinely and contextually relevant to the partner's utterances.
Interactive listening	Reflects the listener's ability to sustain interaction through attentive and responsive behaviors to show engagement and encourage the partner to continue. Includes non-verbal cues (e.g., gaze, nodding) and backchanneling (e.g., "okay," "mm-hmm")
Co-constructive interaction	Demonstrates cooperative engagement through shared responsibility, mutual support, and balanced participation. Includes assisting a partner during difficulties and showing constructive dominance when needed to sustain interaction, while avoiding disruptive dominance that hinders partner involvement.

Guided by raters' perspectives, these three rating categories have the potential to be applied to L2 IC assessment settings and task types that are similar to the ones in the IC test in this study. Incorporating these categories into a rating scale enables a more accurate and comprehensive assessment of IC.

5. Conclusion

This study examined the operationalization of L2 IC in assessment by analyzing raters' perspectives on evaluating interaction quality in paired discussion tasks. Instead of relying on predefined rating criteria, raters assessed interaction quality based on their observations and elaborated on the factors influencing their evaluations. This approach minimized the influence of preconceived notions about IC, enabling a closer examination of its characteristics. The insights gained contribute to the development of rating criteria for L2 IC assessment.

The analysis identified three categories of successful interaction in paired discussion tasks: 'Interactional management,' 'Interactive listening,' and 'Co-constructive interaction'. Raters emphasized the importance of managing interaction through expressing well-reasoned opinions, eliciting a partner's responses, and providing appropriate and genuine responses. 'Interactive listening' encompasses non-verbal cues and backchanneling which reflect listener's ability to sustain interaction. 'Co-constructive interaction' demonstrates the efforts to build mutual achievement. While collaborative interactions where participants equally contributed to constructing interaction were generally preferred, raters acknowledged efforts to maintain interaction even when dominance was present, especially when it resulted from a partner's passivity. Dominance arising from excessive self-expression was frequently criticized for its negative impact on interaction quality. These findings highlight the need to develop a rating rubric that includes these categories as rating criteria to capture the complexity of interaction, such as the value of genuine responses, strategies to maintain interaction, and the dual aspects of dominance depending on its source. Incorporating these features into rating rubrics could lead to more accurate and comprehensive evaluations of IC in paired discussion tasks.

While this study provides valuable insights into operationalizing IC assessment in L2 settings, further study is required to develop valid and reliable rating rubrics. Firstly, a differentiation of varying interaction qualities (e.g., successful, average, and not so good) is required to establish the different levels needed to distinguish interaction quality accurately. In addition, the rating criteria need to be empirically validated through CA inspection of test-taker's discourse. These studies can help identify levels required to differentiate varying performances and develop descriptors for each level and criteria that articulate the differences between higher and lower levels of performance.

Second, all raters in this study were non-native English speakers with EFL teaching experience, which may have influenced their evaluative criteria in ways that differ from those of native speakers. Future research should include native English-speaking raters to examine whether rating criteria vary across different rater backgrounds and to identify potential cultural or linguistic biases in interaction assessment. Lastly, comparative analyses between raters' perspectives and interlocutors' perceptions of interaction quality could shed light on potential discrepancies or agreements in evaluation criteria, leading to a further understanding of interaction. By addressing these areas, future research can further refine rating rubrics to assess IC in paired discussion tasks, ensuring they more accurately reflect the complexities of interaction.

References

Al-Gahtani, S. and C. Roever. 2012. Proficiency and sequential organization of L2 requests. *Applied Linguistics* 33(1), 42-65.

Al-Gahtani, S. and C. Roever. 2018. Proficiency and preference organization in second language refusals. *Journal of Pragmatics* 129, 140-153.

Bachman, L. F. 1990. *Fundamental Considerations in Language Testing*. Oxford University Press.

Canale, M. 1983. From communicative competence to communicative language pedagogy. *Language and communication* 1(1), 1-47.

Cekaite, A. 2007. A child's development of interactional competence in a Swedish L2 classroom. *The Modern Language Journal* 91(1), 45-62.

Dai, D.W. 2022. *Design and Validation of An L2-Chinese Interactional Competence Test.* Doctoral dissertation, University of Melbourne.

Ducasse, A. M. and A. Brown. 2009. Assessing paired orals: Raters' orientation to interaction. *Language Testing* 26(3), 423-443.

Galaczi, E. D. 2008. Peer–peer interaction in a speaking test: The case of the First Certificate in English examination. *Language Assessment Quarterly* 5(2), 89-119.

Galaczi, E. D. 2014. Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics* 35(5), 553-574.

Greer, T. 2016. Learner initiative in action: Post-expansion sequences in a novice ESL survey interview task. *Linguistics and Education* 35, 78-87.

Hall, J. K. and S. Pekarek Doehler. 2011. L2 interactional competence and development, In J. K. Hall, J. Hellermann and S. Pekarak Doehler, eds., *L2 Interactional Competence and Development*, 1-15. Multilingual Matters.

Hellermann, J. 2009. Looking for evidence of language learning in practices for repair: A case study of self-initiated self-repair by an adult learner of English. *Scandinavian Journal of Educational Research* 53(2), 113-132.

Hellermann, J. 2011. Members' methods, members' competencies: Looking for evidence of language learning in longitudinal investigations of other-initiated repair. In J. K. Hall, J. Hellermann and S. Pekarak Doehler, eds., *L2 Interactional Competence and Development*, 147-172. Multilingual Matters.

Hymes, D. 1972. On communicative competence. In J. B. Pride and J. Holmes, eds., *Sociolinguistics*, 269-293. Penguin.

Ikeda, N. 2017. *Measuring L2 Oral Pragmatic Abilities for Use in Social Contexts: Development and Validation*

of An Assessment Instrument for L2 Pragmatics Performance in University Settings. Doctoral dissertation, University of Melbourne.

Kramsch, C. 1986. From language proficiency to interactional competence. *The Modern Language Journal* 70(4), 366-372.

May, L. 2009. Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing* 26(3), 397-421.

May, L. 2011. Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly* 8(2), 127-145.

Pekarek Doehler, S. and E. Pochon-Berger. 2011. Developing 'methods' for interaction: disagreement sequences in French L2. In J. K. Hall, J. Hellermann and S. P. Doehler, eds., *L2 Interactional Competence and Development*, 206-243. Multilingual Matters.

Pekarek Doehler, S. 2019. On the nature and the development of L2 interactional competence: State of the art and implications for praxis. In M. R. Salaberry and S. Kunitz, eds., *Teaching and Testing L2 Interactional Competence: Bridging Theory and Practice*, 25–59. Routledge.

Pill, J. 2016 . Drawing on indigenous criteria for more authentic assessment in a specific-purpose language test: Health professionals interacting with patients. *Language Testing* 33(2), 175-193.

TEPS. 2024. TEPS score band system. Available online at https://www.teps.or.kr/Info/Teps#

VERBI Software. 2021. MAXQDA 2022 [Computer software]. Available online at maxqda.com.

Vo, S. 2024. Exploring the construct of interactional competence in different types of oral communication assessment. *Interaction Studies* 25(1), 1-35.

Youn, S. J. 2013. *Validating Task-based Assessment of L2 Pragmatics in Interaction Using Mixed Methods.* Doctoral dissertation, University of Hawaii at Manoa.

Youn, S. J. 2020. Interactional features of L2 pragmatic interaction in role-play speaking assessment. *TESOL Quarterly* 54(1), 201-233.

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

Appendix

Task Sheets Given to Test-takers

Instruction

Please read the following task description carefully. If you have any questions, ask for clarification. After you understand the task, you will have 10 minutes to discuss the following topics. You can begin the discussion at any time when you and your partner feel ready. Once you begin speaking, you will have 10 minutes to discuss the following two topics, and it will be recorded with a camera. There is no set time for each topic, and you are free to discuss it for the full 10 minutes.

Your task

Discuss the following topics. Exchange opinions with your partner and provide specific reasons and examples supporting your position.

a. "Living in a dormitory is better than living off-campus (living by oneself near college)."
b. "Lectures in university should be given in English to help students improve their global. competence."