OPEN ACCESS

Digital Applied Linguistics
ISSN 2982-2300
https://www.castledown.com/journals/dal/

Digital Applied Linguistics, 4, 103204 (2026)
https://doi.org/10.29140/dal.2026.103204

Assessing L2 Japanese speaking in role-play tasks on an online meeting platform

CHIE OGAWA[a]
NANCY SHZH-CHEN LEE[b]
SHIMING XIA[c]

[a]Kyoto Sangyo University, Japan
[b]Osaka University, Japan
[c]Professional College of Arts and Tourism, Japan

Abstract

This study examines the assessment of L2 Japanese speaking in online role-play tasks from a sociolinguistic-interactional perspective. Twenty L2 learners from China engaged in online role-plays with a professor in various speech acts. Four raters evaluated the participants' video-recorded performances using five criteria adapted from Youn (2015, 2020). Multifaceted Rasch Measurement (MFRM) confirms the raters were able to assess the participants' oral performances effectively using the rubric. In addition to the quantitative analysis, qualitative transcript analysis highlights differences between higher- and lower-rated learners' oral performances. Compared with the high-scoring performances, the low-scoring performances showed difficulties such as inaccurate language use and long pauses both within and between turns. When communication breaks down, managing intersubjectivity becomes essential. Speakers with high interactional competence handled it subtly, while for low-proficiency learners, the negotiation of meaning, use of repairs, and visible struggles to maintain intersubjectivity were much more apparent. Findings provide insights into L2 learners' interactive performance in online settings compared to face-to-face contexts, contributing to a deeper understanding of assessing L2 speakers' sociolinguistic-interactional skills in online role-play.

Keywords: Online role-play; speaking assessment; interactional competence; L2 Japanese

Introduction

L2 speaking skill plays a vital role in communication since it enables L2 users to convey their thoughts, share ideas, and engage in meaningful interaction. This dynamic process is essential for achieving mutual comprehension and for building relationships in L2 settings. Previous speaking research focused mainly on L2 learners' linguistic competence by measuring complexity, accuracy, fluency (CAF), evaluating pronunciation, and lexical usage in a monologue speaking task. These quantifiable measures help researchers understand L2 speakers' ability from psycholinguistic-individualist perspective. However, researchers have pointed out lack of assessing L2 users' language usage in social interaction (e.g., Dai, 2024, 2025; Plough et al., 2018; Roever & Kasper, 2018; Taguchi & Kim, 2018; Youn & Burch, 2020).

According to Roever and Kasper (2018), foregrounding a sociolinguistic-interactional perspective is essential for L2 speaking assessment, as it highlights the importance of designing tasks that elicit purposeful, interactive language use. Such tasks allow test takers to demonstrate their ability to adjust their speech to the interlocutor and context and to engage actively in interaction. However, because tests developed from psycholinguistic-individualist perspective do not fully capture L2 learners' interactional skills, the test taker might get a low score on tests even if they are highly competent in real-world communication (e.g., resolving misunderstandings, working well with others).

Interactional competence has attracted growing attention in second language acquisition (SLA) research. While interactional competence, not limited to L2 contexts, reflects a universal communicative ability, the emergence of L2 interactional competence as a research area stemmed from critiques of proficiency-centered models (Kramsch, 1986). Hall and Doehler (2011) characterize interactional competence as "context-specific constellations of expectations and dispositions about our social worlds that we draw on to navigate our way through our interactions with others" (p. 1), implying the complex ability to mutually coordinate actions. Dai (2025) also explains interactional competence is a speaker's ability to interact with others in a range of real-world social, cultural, and interpersonal contexts. Building on these conceptualizations, this study adopts definition of L2 interactional competence tailored to the demands of speaking assessments that involve turn-taking and interaction as an expected social role in a given context. Specifically, L2 interactional competence is characterized as learners' ability to participate in turn-taking in ways that maintain the structure of interaction (Kasper & Ross, 2013). This perspective emphasizes the procedural and collaborative nature of spoken interaction, which is particularly salient in performance-based speaking tasks. Moreover, this definition aligns with a growing body of work that views interactional competence as co-constructed in interaction rather than residing solely in individual ability (Kasper & Ross, 2013; Roever & Kasper, 2018; Ross, 2017).

To capture the L2 learners' interactional competence, speaking assessment should better ensure that L2 speakers are able to demonstrate their interactional behavior in the real-life scenarios such as responding to the interlocutor, negotiation of meaning, or co-construction of meaning. In this regard, role-play tasks can provide a situation similar to a natural conversation setting and it functions well to understand how L2 speakers achieve a task in interaction (Okada, 2010). During role-play, L2 learners usually attempt to figure out how they can achieve a task goal in a given scenario. Learners need to constantly consider how to address the person, the problem, and use appropriate utterances in a role play. Dai (2024) states that role-specific interactional abilities are always required in any form of interpersonal interaction. This study aims to investigate the L2 Japanese speakers' interactional competence as an expected social role in a specific speech act context.

Literature Review

L2 Japanese and Computer-Mediated Communication

In today's digital era, the use of synchronous online platform such as Zoom, Skype, or Microsoft Teams has become increasingly common. Such communication tools are now widely recognized as effective means for delivering language instruction including L2 Japanese. One advantage of using these online meeting tools is that they provide increased opportunities for learners to communicate with distant interlocutors. Especially for L2 Japanese learners who are not in Japan, these tools offer a convenient and efficient way to interact with native speakers of the target language. For instance, Kato et al. (2023) investigated the effectiveness of video-based synchronous learning in a Japanese program at an American university. In their study, L2 Japanese learners engaged in project-based learning to collaboratively develop a homepage with students in a Japanese university through Skype. The participants met their Japanese partners 15 times and communicated both English and Japanese for at least 15 minutes each throughout the semester. The result of analyzing pre-and post-speaking tests indicated that video-based synchronous learning was effective to enhance their speaking fluency to some extent, and the questionnaire result showed that it also increases their motivation.

Sato et al. (2017) implemented both asynchronous (e.g., Google docs) and synchronous communication tools (e.g., Google Hangouts) for a task-based Japanese course at an American university. A simulated Oral Proficiency Interview (OPI) was conducted for the online students, and their oral performances were compared with that of the face-to-face cohort in a previous semester. The quantitative result shows that online students outperformed the face-to-face group, particularly in their use of communication strategies such as negotiating meaning or solving communication breakdowns. These online tools also provided students with greater opportunities to meaningfully interact with peers, Teaching Assistants (TAs), and the instructor in the target language, which in turn appeared to motivate them to strengthen their functional communicative skills throughout the online course. Sato further acknowledged that online video-conferencing tools can be effectively employed for the assessment purposes as performance tasks can be recorded and carefully evaluated afterward. These previous studies highlight the potential of synchronous video-based learning for developing L2 Japanese speaking skills. Yet, further studies are warranted to examine alternative assessment modes using the online meeting platform given the current and future needs of L2 language assessment.

Assessing L2 Interaction in Role-Plays

In role-play, test takers are required to act out given roles in various scenario. One of the advantages of employing role-play is that it goes well beyond the criteria of fluency and accuracy (e.g., Grabowski, 2013; Kasper & Youn, 2018; Okada, 2010; Ross, 2017; Ross & O'Connell, 2013; Youn, 2015; 2020). This is because a role-play task has a distinctive structure, as a candidate draws on interactional competencies to "co-construct it as a coherent interaction" (Ross, 2017, p. 99). For example, role-play tasks can assess L2 learners' pragmatic competence in spoken interaction in different speech acts such as request or refusal (e.g., Okada, 2010; Youn, 2015, 2019). This suggests that role-plays can function as a good way to measure L2 leaners' sociolinguistic-interactional ability.

Several studies have demonstrated the construct validity of role-play. Okada (2010) examined what interactional competencies a test taker draws on to deliver a role-play task in the OPI. His carefully examined 71 role-play tasks of Japanese learners of English using Conversation

Analysis (CA). The findings revealed that L2 speakers were able to respond to interlocutors' explicit and implicit requirements of the next desirable responses, which is similar to how it is done in ordinary conversation. Okada's study suggests that role-play tasks have potential to evaluate and monitor L2 learners' pragmatic ability that can be also applicable in the real-world scenarios.

Nakatsuhara et al. (2021) investigated the comparability of face-to-face and online (Zoom-based) delivery modes in the IELTS Speaking Test, focusing on both test scores and language functions. A total of 99 test-takers were assessed by 10 trained IELTS examiners, who rated performance in both modes using the operational IELTS 1–9 scale across the four analytic rating criteria: Fluency and Coherence, Lexical Resource, Grammatical Range and Accuracy, and Pronunciation. Analysis using the Many-Facet Rasch Measurement (MFRM) model revealed that, although the video-conferencing mode resulted in slightly lower scores, the differences were not statistically significant. However, the online mode elicited more frequent clarification requests from test-takers. The authors suggest that this may be a characteristic of the video-conferencing context, where audio is transmitted digitally. These findings indicate that while the two modes (online vs face-to-face) yield comparable performance outcomes, video-conferencing may influence specific interactional features of spoken language use.

When assessing interactional competence, it is essential to understand how interlocutors manage communication problems during interaction. When communication breakdowns occur in interaction, effective handling of these moments is key to successful interaction. A central concept in this process is intersubjectivity, which refers to the cooperative action between two or more people (Goodwin, 2018). Speakers build mutual understanding by responding to each other's actions and words, drawing on shared experiences and contextual knowledge (Burch & Kley, 2020). Because interactional competence is not solely an individual ability but is also co-constructed through discursive practices among participants (Kasper & Ross, 2013; Young, 2019), managing intersubjectivity becomes a critical component of communication. Successful interaction requires recognition and negotiation of meaning (Lam, 2018). In this process, interactive listening, such as backchanneling, shows awareness of being a listener during a conversation; either through verbal or nonverbal cues (Lam et al., 2023).

Rating Rubrics for Assessing

Building on these theoretical perspectives, researchers have examined how speaking in interaction can be operationalized and assessed in language testing contexts. For example, Youn (2015) included interactional components in the role-play assessment to validate assessing L2 English learners' pragmatics ability. The participants were 102 English as a Second Language (ESL) learners in the U.S. The participants had a role-play with a peer and with a professor in different speech acts such as request or refusal. The role-plays were carefully designed as open role-plays, which allowed test takers to deliver the role-play conversation without fixed outcomes. In that way, it would be more appropriate to elicit the role-play data to examine test takers' interactional competence. The raters scored each role-play using the five rating criteria with three-point Likert-scales. The rating criteria include 1) content delivery; 2) language use; 3) sensitivity to situation; 4) engaging with interaction; and 5) turn organization. Among the five, 4) engaging with interaction and 5) turn organization are considered to include descriptions of interactional features. Youn carefully examined the test takers' role-play performances of different English proficiency levels by employing CA to see the linguistic choices and interactional features that explain constructs of the rating criteria. After ensuring the degree of authenticity and standardization of role-play performances with the designed rating criteria,

MFRM was used to examine the test takers' performances, raters' severity, role-plays' task difficulties, and rating criteria difficulties. MFRM results indicated that the role-plays stably differentiated between the varying degrees of the test takers' interactional competence. The rating rubrics work appropriately, suggesting the validity of the rating criteria on the role-play tasks.

Another study conducted by Youn (2020) examined how L2 learners managed their interactional sequences using a CA approach. The study focused on role-play scenarios where examinees acted as classmates collaborating on a project. The task required them to negotiate scheduling details, including the time and mode of their meetings (face-to-face or online). The 38 extracts of role-play interactions from Youn (2015) were analyzed in this study. The study's analytical focus was on how examinees navigated the interactional demands of proposal sequences, such as initiating proposals, signaling transitions, and responding to proposals (Schegloff, 2007). CA revealed that higher-level examinees effectively employed coherent sequential organization throughout their role-plays. Conversely, lower-level examinees often struggled to establish shared understanding prior to proposing actions, resulting in less effective interactions. The findings show that proposal sequences represent a measurable aspect of L2 interactional competence. Therefore, this dimension should be explicitly incorporated into assessment criteria for evaluating interactional competence.

Building on previous studies by Youn (2015, 2020), there is a clear need for mixed-methods research in assessing L2 speaking in interaction. One approach is utilizing MFRM to evaluate the validity of rubrics and macro-level interactional competence. At the same time, incorporating qualitative methods provides valuable insights into how speakers engage in social practices during role-plays (Burch & Kasper, 2021; Kasper & Youn, 2018; Lam et al., 2023; Ross, 2017; Youn & Burch, 2020). Furthermore, future research should emphasize the context-specific and task-dependent nature of interactional features. Expanding on this, studies could explore role-plays conducted in online settings, which present both unique challenges and opportunities for assessing interactional competence.

Research Purposes

Previous studies related to assessing interactional competence have shown that the role-play tasks can work well to assess pragmatic competence and interactional competence (e.g., Okada, 2010; Ross, 2017; Youn, 2015, 2020). It also has proven that the construct validity of the role-play assessment was empirically demonstrated. Nevertheless, these previous studies were mainly for face-to-face situations. Although more studies are looking into usage of technology in L2 pragmatics learning (e.g., Klipp & Reinders, 2025; Taguchi, 2024; Tang & Taguchi, 2021) or human-machine role-plays (Dombi et al., 2022; Timpe-Laughlin et al., 2022, 2023), very few studies examined the assessment using an online conference tool. L2 Japanese online role-plays are sparse.

In this study, we employ the assessment criteria which was used to evaluate ESL learners' speaking in face-to-face role-play by Youn's studies (2015, 2020). Her finding has demonstrated that the criteria were valid to assess face-to-face interaction in English. As online interactions using video-conference tools has become a regular part of daily life, it is essential to assess the applicability of the similar rubrics across different settings. Furthermore, by using video-recorded assessments, we can explore how the L2 speakers' non-verbal features (e.g., nodding) can be evaluated, which was not included in Youn's studies. The purpose of this study is to investigate the applicability of replicated rating criteria for assessing L2 Japanese learners' interactional competence in an online context, and to explore

the characteristics of online interactions. Specifically, the research questions addressed in this study are as follows:

RQ1: Do the replicated rating criteria reflect the construct definition and stably assess L2 Japanese interactional competence in an online setting?
RQ2: How do interactional features appear in online role play interactions conducted in Japanese?

Method

Participants

Twenty L2 Japanese learners, five males and 15 females, participated in this study. The participants were 2nd to 4th year university students, majoring in Japanese. All of the participants were native speakers of Chinese. They were recruited from two universities located in China. One university was a leading foreign language institution in northern China, well-known for its Japanese program. The other was a comprehensive university in southern China with a foreign language faculty. The participants had been studying Japanese since entering university, and their proficiency was estimated to be between CEFR A2 and B1. Because many of them had not taken the Japanese Language Proficiency Test (JLPT) before data collection, proficiency levels were determined based on self-reports and their teachers' observations. All participants were given consent forms before participating in the study. Ethical approval for this research was obtained from the first author's university ethical committee.

Role-Play Tasks

Three versions of role-plays with a professor were adapted from Youn (2015, 2020). Table 1 represents role-play situational scenarios.

Table 1. Descriptions of role-plays with a Professor

Task	Scenarios	Speech act
Task 1	Requesting a recommendation letter for a graduate school application that is due in a week.	Request
Task 2	Requesting additional advising time to discuss a term paper topic.	Request/Negotiation
Task 3	Refusing a professor's request to change a presentation schedule.	Refusal

Role-Play Task Procedures

Participants were instructed to use their own personal computers and stable internet connections, and they were informed that the whole session was video-recorded. Each participant entered an online meeting platform called VooV Meeting by Tencent at an allotted scheduled time. One of the researchers, who is a Japanese native speaker, acted as an interviewer. The role-play was conducted only in Japanese. The participants met the interviewer for the first time, therefore, before starting the role-play tasks, five-minutes of free conversation (comprising of things such as self-introduction, discussion about hometowns, weekend plans, etc.)

was held for the purpose of a "warm-up". Then, instructions for the role-play were given in Japanese such as the situation in which the conversation occurs and the relationship between interlocutors. Each time before the role-play, an electronic role-play card which describes the situational scenario in Japanese was sent in a chat message and read aloud by the interviewer. After the card was explained, the interviewer made sure the participant understood the scenario and what to do in their role. Each of the three role-play situation cards including the task goal was sent each time before each role-play administration: "Please ask a professor for a recommendation letter" for Task 1, "Please make a meeting appointment" for Task 2, and "Please refuse the professors' request" for Task 3 (see Appendix A). The role-play card remained visible on the screen; however, the participants were instructed to *try not to read* directly from the card or use any other materials, including a dictionary or search engines, during the interaction. No specific time limit was set for completing each role-play. After the three role-play tasks, the interviewer asked the participant about their thoughts and reflections on their own performances in Japanese. In total, the interview took approximately 30 minutes. Table 2 shows procedures of how the role-play tasks were conducted online.

Table 2. Procedures of the role-play tasks

Timeline	Content	Time
Stage 1: Warm-up	Explanation of the interview	5 minutes
	Self-introduction and free conversation	
Stage 2: Main tasks	Role-play task 1	15–20 minutes
	Role-play task 2	
	Role-play task 3	
Stage 3: Reflection	Retrospective interview	5 minutes

Rating Criteria

Ratings were adapted from Youn (2015). A three-point scale was used: 1 = *unsuccessful*; 2 = *moderate*; 3 = *successful*. More detailed description is explained below and shown in Appendix B. The Japanese terminologies in the parentheses were created by the researchers to make it more comprehensible to the raters to assess L2 Japanese. Importantly, because the present study used video-mediated interactions, visible nonverbal behaviors (e.g., nodding, gaze) were included in the *Engaging with Interaction* criterion as an extension of the original rubric. Youn (2015) validated a rubric in face-to-face, audio-recorded role-plays where nonverbal behaviors were not considered. In contrast, the video-mediated format of the present study allowed raters to observe nonverbal behaviors (e.g., nodding, gaze, visible engagement). As such, the current rubric should be understood as a conceptual transfer from Youn's audio-based descriptors.

Content Delivery (naiyo no ryucho-sa)

This part assesses a speaker's ability to deliver a speech act smoothly and without disfluencies. When a speaker sounds choppy and disfluent in their own turn, a point is deducted. This includes factors such as the presence of repetitions or long pauses, as well as the overall flow of the speaker's turn. The differences in pauses between *content delivery* and *turn organization* is

whether the pauses appeared within the turn or between the turn. If the speaker makes a long pause within their turn, it influences content delivery.

Language Use (gengo shiyo)

This category looks at to what extent a speaker uses morpho-syntactically accurate utterances and appropriate pragmalinguistic. For example, asking a professor for a recommendation letter, saying "—*kaite itadake masenka* (would you please write….)?" is considered pragmatically more appropriate than a direct request such as "*kaite hoshii* (I want you to write)." Additionally, directly reading aloud the role-play card results in a point deduction because it is not an authentic production of the utterance.

Sensitivity to the Situation (jyoukyo ni tsuite no ninshiki)

This category examines to what extent a speaker can perform based on the knowledge of sociopragmatics (Leech, 1983). For example, it would be essential to explain the situation with sufficient reason when refusing a professor's request than just saying "*dekimasen* (I cannot do it)." Another example is that it is a crucial part to explain the situation and acknowledge the short notice when asking for a recommendation letter. In Youn's (2015) study, intermediate ESL learners did not bring up the letter due day until the professor explicitly asked when it should be sent. Not mentioning the deadline or only providing important information after being asked by the professor will result in a point deduction.

Engaging with Interaction (yaritori deno kakawari)

This part measures to what degree the speaker engages in interaction and establishes a shared understanding with the interlocutor. Acknowledgment tokens such as backchanneling or *aizuchi* (e.g., hai, ee, un as yes) are used to measure a speaker's engagement. Nonverbal features, such as nodding, are also included in this study because the role-play performances were video-recorded. In contrast, Youn's (2015) analysis was based on audio recordings and did not account for nonverbal features. Failure to respond to the interlocutor's utterance can lead to unclear communication and a deduction of points.

Turn Organization (ta-n)

This part measures the speakers' interactional competence. There are two components in this criterion. First, duration of pauses between turns is considered. A very long pause between turns sounds awkward. On the other hand, when engaging in dispreferred action such as refusing the request, no-pause immediately sounds inappropriate or rude. The second one is an adjacency pair. The speaker should provide a relevant pair part to the interlocutor's previous turn, such as greeting and greeting, asking a question and answering, or accepting a request and thanking (Schegloff, 2007).

Raters and Rater Training

After the rubric was developed in a Japanese version, additional raters were recruited while the first author and the third author acted as raters. Among four raters, two raters were Japanese native speakers while other two were Chinese native speakers, who were proficient in Japanese. At the time of data rating, two raters were doctoral candidates in Japanese linguistics, the third rater had a doctoral degree in applied linguistics, and the fourth rater had a doctoral degree in social science.

A 90-minute online rater training session was conducted to help the raters understand the rating criteria for each component. First, the first author explained the role-play tasks and the

rating rubric to the recruited raters. The raters then watched the three sample performances on the video and assessed using the Japanese rubric (see Appendix B). Next, the raters discussed their ratings and the reasons for their scores. After the rater training, they rated all the role-play performances at their own pace at home. During the take-home assignment, the raters reached the researcher when questions arose. In total, all the raters scored all 60 performances (20 participants × 3 role-play tasks) which took approximately one week to complete.

Quantitative Analysis

Regarding RQ1, a Rasch analysis was conducted using FACETS 3.85.1 (Linacre, 2023) to evaluate the interactional competence of the participants' role-play performances. The raters' raw scores were statistically analyzed; A total of 60 distinct performance codes were considered for MFRM analysis. The facets in this study were person ability (60 performances from 20 learners), rater severity (4 raters), and rating criterion difficulty (5 criteria). During the initial examination of MFRM, a subset problem arose, which refers to a lack of identifiability of the estimates. This problem impacts the relationship of the measures in the subsets, meaning that it is not possible to compare the measures of the subsets on the logit scale (McNamara et al., 2019). Following recommended practice (McNamara et al., 2019), the task facet was anchored, with all three tasks fixed at 0 logits. Anchoring resolved the identifiability issue and reflects the analytical decision to treat the tasks as equivalent in difficulty for this analysis. Person ability was then estimated while taking the effects of the other facets into account. The person facet represented individual task performances rather than individual learners. Although multiple performances were produced by the same learners, treating each performance as an independent unit allowed for a more fine-grained examination of rating task interaction in a specific situation. The logit person ability measures were produced from the FACETS analysis, which shows a single combined measure of the scores from the five rating criteria while taking the effects of the other facets into account such as rater strictness and scale difficulty.

Qualitative Analysis

Based on the findings from the quantitative analysis, specific interactions were selected for further examination. To address RQ2, the participants' interactions were qualitatively analyzed to identify significant features of their performance in the online setting. We adopted Youn's (2015) approach to analyzing highly evaluated and poorly evaluated performances from the MFRM analysis, which is explained in more detail in the qualitative findings. The video-recorded role-play performances were transcribed with abbreviations used in gloss translations (see Appendix C), following the Atkinson and Heritage (1984) transcription style (see Appendix D). The researchers cross-checked the data as follows: first, the first author transcribed the spoken data; then, the co-authors checked the transcripts. The first author was responsible for finding emergent interactional features, which was carefully reviewed by all authors. When disagreement occurred, the authors discussed in order to reach an agreement.

Results

Quantitative Findings

MFRM shows estimates of test-takers' abilities in a true interval scale called *logits* (log-odds units). The FACETS results showed that the infit MNSQ values for raters ranged from 0.81 to 1.16 (see Table 3), suggesting that they were well within the acceptable 0.5 to 1.5 range (Fisher, 2007). This means that the raters were not erratic or overly restrictive in the use of the scale. The FACETS analysis indicated that the mean Rasch difficulty estimates (measure) for the four raters ranged from –0.89 to 1.29; Rater 2 had the highest severity estimate

Table 3. Rater severity measurement report for interactional competence

Rater	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
2	1.29	.13	0.91	–1.1	1.17	.90	.79
3	0.30	.13	1.11	1.4	1.11	.8	.78
4	–0.50	.13	0.81	–2.4	.70	–2.30	.86
1	–0.89	.14	1.16	1.7	1.11	.7	.80

followed by Rater 3, Rater 4, and Rater 1. This means that Rater 2 was the most severe rater and Rater 1 was the most lenient.

In addition to these rater-level findings, the overall model fit indices supported the quality of measurement. Rater reliability from Rasch analysis was high (.97) and rater separation was 6.12, which indicates greater variability in rater severity. Person reliability was .96 (person separation = 4.69), indicating that the rating scale clearly distinguished among the speakers' performances. A bias interaction analysis confirmed that these severity differences did not systematically disadvantage any particular speakers or rating criteria.

The mean Rasch item difficulty estimates for each rating component ranged from –.74 to .66 (see Table 4). Language use had the highest difficulty estimate followed by content delivery, turn organization, engaging with interaction, and sensitivity to situation; thus, language use was the most difficult criterion, and sensitivity to situation was the easiest criterion on which to get a high score. Those five rating items met the infit MNSQ criterion of .50–1.50, and their standardized statistics did not exceed the ± 2.00 criterion. The part-measure correlation shows the extent to which each component correlated with the total score. The correlation coefficients for the given items were between .77 and .85, which shows the extent to which they correlated with the total score. The Rasch item reliability estimate of the five components was .93.

Table 4. Rating criteria difficulty measurement

Rater	Measure	SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD	Pt-measure correlation
Language use	.66	.15	1.07	.8	1.06	0.3	.77
Content Delivery	.46	.15	.84	–1.9	0.76	–1.4	.84
Turn Organization	–.22	.15	.91	–0.9	0.79	–1.3	.85
Engaging	–.26	.15	1.17	1.8	1.18	1.0	.80
Situation	–.74	.15	.98	–0.8	1.33	1.7	.81

Note: Engaging = Engaging with interaction; Situation = Sensitivity to the situation.

Table 5 shows the Rasch rating category statistics for interaction competence. All categories functioned well according to the rating scale diagnostic criteria: category frequency, average measures, threshold estimates, category fit, and probability curves. Across all ratings pooled

Table 5. Ratio of rating category

Rating category	Count (%)	Average measure	Outfit MNSQ	Rasch-Andrich threshold	SE
1. Unsuccessful	342 (29.0%)	–3.30	1.0	—	—
2. Moderate	499 (42.3%)	–.07	1.1	–2.10	.10
3. Successful	339 (28.7%)	3.27	0.9	2.10	.10

Note: Counts reflect individual ratings assigned by raters across five criteria for all task performances.

from performances, rating criteria, and raters, 29.0% of the ratings (342 counts) were classified as unsuccessful, 42.3% (499 counts) as moderate, and 28.7% (339 counts) as successful. The results met one of Linacre's (2002) guidelines for evaluating rating scale category functioning.

The FACETS map in Figure 1 represents an overview of the rating results. All facets were measured in *logits*, which are displayed on the left side of the map in the measure column (–6 to +6 logits). The second column shows the participants' Rasch ability estimates. More competent participants' performances are placed toward the top and less competent participants' performances toward the bottom. The third column shows the rater severity estimates. More severe raters appear toward the top, and more lenient raters appear toward the bottom. The fourth column shows the difficulty of the five rating criteria. Language use was the most difficult and sensitivity to situation was the easiest. The last column, Scale, shows raw scores.

Qualitative Findings

Based on the FACETS results, the logit values were used to determine the participants' performance levels. We then qualitatively analyzed the two highest and two lowest evaluated performances to gain insights into how the participants interacted with a professor in different speech acts. The reason we chose two performances from each level was to set a more modest goal, allowing us to focus on fewer excerpts and examine them in greater detail to answer RQ2. Excerpts from highly evaluated performances (Speaker 1 and 2 with logit = 5.51) and poorly evaluated performances of Speaker 3 and 4 (–5.37, –6.64) were selected for qualitative analysis (see Table 6). Through repeated listening and transcribing of their interactions, distinctive characteristics in participants' interactional features were identified.

Highly Scored Performances

Some notable features of high-scoring performances included smooth content delivery and appropriate language use. Excerpt 1 shows that Speaker 1 successfully negotiated a meeting time with the professor, demonstrating effective communication (P = Professor; S1 = Speaker 1). S1's smooth delivery stemmed from speaking in longer, uninterrupted utterances (lines 1, 4, 6), reducing choppiness and pauses. A key feature of S1's performance was politeness in requests, enhancing appropriateness in language use. For instance, in line 2, she asked, "*soudan shitemo yoroshii desu ka*?" ("Could I discuss it with you?"), using the polite form *yoroshii desu ka* to maintain respect. This showed awareness of student-professor interaction norms. S1 also displayed sensitivity to situation. In lines 4, she explains that she wants to arrange the schedule since it might take longer time and she provided detailed justifications for her availability, facilitating negotiation (line 10–12). The negotiation proceeded without miscommunication, indicating her effective interactional competence.

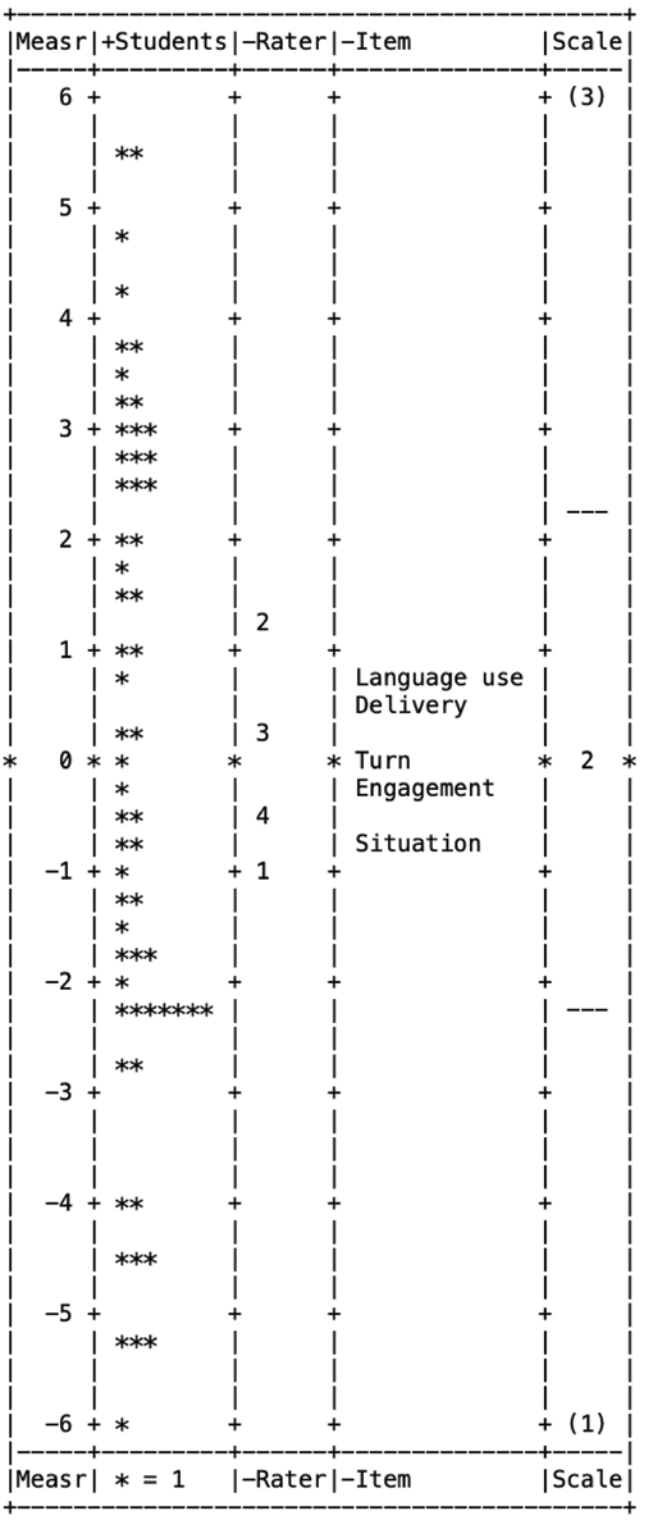

Figure 1. *FACETS summary for the role-play task*

Table 6. Selected performances for qualitative analysis

Selected performance	Task	Speech act	Logit
Speaker 1	2	Request a meeting	5.51
Speaker 2	3	Refusing	5.51
Speaker 3	3	Refusing	–5.37
Speaker 4	2	Request a letter	–6.64

Excerpt (1) Speaker 1: Requesting to make an appointment

1. S1: sensei dewa ano jyugyou no kadai ronbun nitsuite mada topikku ga
   teacher then well class N assignment paper about yet topic S

2. Kimatte- kimeteinai node soudan shitemo yoroshii desu ka?
   decided decide-NEG because discuss do may I CP Q
   "teacher, I haven't decided a class assignment paper topic yet, so could I discuss it with you?"

3. P: hai ii desuyo ima sodan ni nori masuyo. do sare mashita ka?
   yes ok CP now advice O give CP how do CP Q
   "Yes, you may. I can have time to discuss now. How can I help?"

4. S1: e chotto jikanga kakari sou nanode (.) [ano
   oh little time take seem because well
   "It might take a little time, so"

5. P: [hai
   "Uh huh"

6. S1: hokano jikan demo yoroshii desho ka
   other time O possible CP Q
   "Could we have it at another time?"

7. P: ah. hai iidesu yo↑.
   oh yes okay CP
   "yes, it's okay"

8. ano watashino ofisu awā ga kayobi no 12 ji kara desu
   well my office hour S Tuesday 12 time from CP
   "Well my office hour is from 12 on Tuesday."

9. kayobi no 12ji wa dodesuka(.)online kaigi settei shimasho ka?
   Tuesday of 12pm TP how about online meeting set let's Q
   "How about 12 on Tuesday. Shall we have an online meeting?"

10. S1: (Tilting head)
kayobi wa chotto jyu- sono jikan ga jyugyo ga aru node
Tuesday TP a little this time S class S have because

11. gogo wa ano suiyobi kara kinyobi made gogo ga zutto aiteru node
PM TP well wednesday from Friday till PM S all free so

12. sono jikantai dewa yoroshii desho ka?
this time slot S okay CP Q
"I have a class on Tuesday. I'm free in the afternoon from Wednesday to Friday, so is one of those time-slots, okay?"

For the refusal speech acts, Excerpt 2 demonstrates another highly scored performance. Speaker 2 (S2) refused a professors' request effectively. Her engagement in interaction and turn-taking was effective, as she consistently understood the previous turn and responded accordingly. Before she explained about her refusal, she paused for 3.6 seconds (line 9), which might indicate hesitation. She had several self-corrections and reforms (e.g., line 10: *"watashi wa ano watashi no"*). She used fliers (e.g., *"ah" "ano"*) when explaining a reason for declining the request, which functions subtly but effectively to softening refusal (line 10–12). Another striking feature was she suggested a possible alternative presenter (line 18–21) after she refused a professor's request. This was not included in the role-play instruction, but she created an imaginary classmate, Mr. Chaw. Not only did she demonstrate a high level of accurate language use with smooth turn-taking, she also demonstrated some sociopragmatic awareness by proposing an alternative person rather than simply refusing.

Excerpt (2) Speaker 2: Refusing a professor's request

1. P: Wang san konnichiwa.
Ms. Wang hello

2. S2: konnichiwa sensei
Hello teacher

3. P: Ano konshu no jyugyo nan des kedo::
Well this week N class N CP bu

4. Li-kun ga byoki nan desu yo:: sorede::
Mr Li S sick N CP and

5. Li-kun ga jugyo korarenai node Li-kun no kawarini
Mr. Li S class come cannot so Mr. Li N instead

6. purezentetion shite kuremasen ka?
presentation do give-NEG Q?
"Because Mr. Li cannot attend the class, could you give a presentation instead?"

7. S2: [ah
"Aw"

8. P: [jyugyo wa mikka go no jyugyo nandesu
class TP three days later N class CP
"Class is in three days"

9. (3.6)(S looking down)

10. S2: ah watashi wa:: ano: watashino prezen no ban wa
CP I TP well my presentation N order TP

11. Li-san no Li-san no ato ishuu nanode
Mr. Li N Mr. Li N later one week because

12. mada jyunbi hajimeteinai desu
yet prepare start-NEG CP
"Ah, a, my presentation is supposed to be in a week after Mr Li, Mr Li's, so I haven't started preparing yet."

13. P: [so:nan desu yo:: hai
"I see. Yes"

14. S2:[soshite ano mikka go wa ano chodo okina kimatsu tesuto
and well three days later TP well little big final test

15. ga ari(.) watashi wa ima issho:kenmei jyunbi shiteimasu
S have I TP now hard prepare do
"and three days from now, I have a big final exam and I am now preparing for it very hard."

16. P: aa sonan desune(.) ja kimatsu tesuto ga aruto un
Ah so CP then final test S have SF

17. konshu no presentation wa muzukashi desu ka?
this week presentation S difficult CP Q?
"I see. okay, you have a final test, so is it hard to give a presentation this week?"

18. S2: a hai: soshite ano a cho:san wa zutto mae kara u:n jyunbi
yes and well Mr Chaw TP far before from SF prepare

19. shiteimashita node
do-ing-PAST so

20. mou jyunbi dekita to omoi masu node cho:san ni tanon dara
already prepare finish-PAST think CP so Mr Chow O ask if

21. [do desu ka? [to
how about Q
"Oh, yes. Mr. Chow has been preparing for a long time, I I think he's ready now, so how about asking Mr. Chow?"

22. [(laugh) [a wakarimashita(.) jaa Chow-san ni
SF understand-PAST then Mr. Chow O

tanonde mimasu ne
ask try CP
"I understood. I will try to ask Mr. Chow then"

23. S2: hai sumimsen sensei
yes sorry teacher

24. P: tondemo nai desu
big deal-NEG CP
"Not a big deal"

Lower-Scored Performance

Compared to the highly scored performances, low-scored participants struggled to deliver the speech act smoothly. Excerpt 3 illustrates Speaker 3 (S3)'s performance. Notably, she kept noticeable awkward silences when it was her turn to speak (lines 7, 9, 12, and 15). In particular, when the professor made a request, she remained silent for almost six seconds (line 7). The professor then rephrased and repeated the request (line 8). In response, S3 asked for a repetition by saying, "*sumimasen ga, moichido shite kudasai* (Sorry, but could you *do* it again?)". Professor then sought clarification, asking if she meant "*iu* (say)" (line 11). S3's engagement in the interaction was limited, as she remained silent instead of responding to the previous turn. After a four-second pause (line 12), S3 attempted to refuse the professor's request while leaning in toward the screen (line 13). S3's speech sounded disfluent with frequent pauses and repetitions (lines 16, 18). S3 then leaned in again as if she was checking her screen (line 18), while proceeding with her refusal. She was able to explain the situation (line 16–18), which functions as an indirect refusal.

Excerpt (3) Speaker 3: Refusing a professor's request

1. P: Kyo san konnichiwa:: chotto ii desu ka?
Ms. Kyo hello little okay CP Q
"Hello Ms. Kyo, Can I have a little moment?"

2. Kyo: hai
yes

3. P: hai(.) konshu no jyugyo nan desu kedo:: Li-kun ga byoki ni
   yes this week N class N CP but Mr Li S sick O
4. Natteshimatte(.) Li kun no purezenteshon ga oyasumi ninarun desu ne(.)
   has become Mr. Li N presentation S absent become CP
5. kawarini kyo-san purezen yatte kuremasen ka?
   instead Ms. Kyo presentation do give-NEG Q
   "Since Mr. Li has become sick his presentation this week is canceled. Instead, Ms. Kyo, can you do a presentation?"

6. mikkago nandesuyo. Mmkka go no jyugyo desu
   three days CP three days-after N class CP
   "It is in three days, class in three days"

7. (6.6)

8. P: do desu ka ne(.) Li-kun byokini nacchatte oyasumi surun datte
   how CP Q CP Mr. Li. sick become-PAST absent do seem
   "What do you say?"

9. (6.6)

10. S3 (leaning in toward the screen)
    un (3.3)sumimasen ga mouichido shite °kudasai°(2.4).
    SF sorry but once more do please
    "Sorry but once more"

11. P: moichido iu?
    once more say
    "do you want me to say one more time?"

12. (4.2)

13. S3:(leaning in toward the screen)
    u: u:m sumimasen ga watashi wa
    SF sorry but I TP

14. P: hai
    yes

15. (4.6)

16. S3: mikka ato wa(1.8) um um um shumatsu shiken ga arimasu.
    three days after TP SF weekend exam S have
    "Sorry but I have a final exam in three days"

17. P: u::n
SF

18. S3: [ari- arimasu kara:: (gazing at the screen)
have have because
"I have, so"

19. P: [sokka
I see

20. hai (2.0) isogashii desu ka ne? (2.0)
yes busy CP Q CP
"okay, you are busy, aren't you"

21. S3: hai um: (5.1) sore wa- sore wa muzukashikutte(1.5)
Yes SF this TP this TP difficult-and

22. benkyo (2.6) muzukashi- muzukashii no ga- ka- desu
study difficul- difficult N CP
"It is diffi—, difficult ---."

Excerpt 4 illustrates Speaker 4 (S4)'s requesting of a recommendation letter. First, S4 was not able to respond to the greeting due to it overlapping with her request (line 3). Second, her content delivery appeared somewhat abrupt since her request was direct, lacking a pre-request sequence (e.g., Youn, 2020). She began with "*mazu daiichi* (first of all)" when requesting for a letter. Although S4 attempted to speak politely, she failed to use appropriate pragmalinguistic knowledge accurately. For example, her use of honorific prefix o- as a sign of politeness in "*okiki kudasai* (Please listen, line 9)" led to an error, as she used the wrong verb, saying "*kiki* (listen)" instead of "*kaki* (write)". This misuse of the verb prompted the professor to ask for clarification in line 11, asking "*okiki kudasai?* (listen to it?)" Then, S4 attempted to rephrase the request in a different polite form as *"go kiki itadake masen ka*" (line 11) with some errors of "*go-kiki*". Lastly, related to the situational sensitivity, when the professor agreed to write a letter (lines 15–16), S4 just nodded without expressing gratitude (line 17). S4 was unable to explain the due date until the professor asked (line 18). S4 initially responded directly with "next week" and then rephrased it as "this week." Overall, S4's delivery lacked explanation in the given scenario and a proper pre-request sequence.

Excerpt (4) Speaker 4: Requesting a recommendation letter

1. S4: um sensei um
SF teacher SF
"um teacher um"

2. P: hai [konnichiwa
yes hello

3. S4: [daigaku mazu daiichi (1.8) daigakuin no
university at first graduate-school N

4. shingaku no tameni um
applying N for SF
"To apply for a graduate school"

5. P: hai
yes

6. S4: suisenjo ga hitsuyo [dakara (1.5) mm
recommendation letter S necessary because SF
"I need a recommendation letter"

7. (3.8)

8. P: [a hai
yes

9. S4: n go- sensei sensei ni sensei ga wata- watashi ni (2.2)
SF teacher teacher O teacher S I- me O

10. sensei kara (1.6) suisenjo wo okiki kudasai
teacher from recommendation letter O listen please
"teacher please listen to a reference letter"

11. P: kiki kudasai?
listen please?

12. S4: go- goki- go-kiki itadake-masen ka
listen get-NEG Q
"teacher could you kindly listen to a reference letter"

13. P: a suisenjo wo kakun desu-ne
SP recommendation letter O write CP
"Ah, do you want me to write a recommendation letter, don't you"

14. S4: a hai(nodding)
yes

15. P: un hai jaa watashiga etto suisenjo wo kakimasu yo.
CP yes then I CP recommendation letter O write CP

16. un daijobu desu
no problem CP

17. S4: (smiling nodding)(1.0)

18. P: itsu made desu ka?
When by CP Q
"by when?"

19. S4: mm raishu m konshu
SF next week SF this week

Discussion

The results of the MFRM analysis confirmed assessment criteria functioned well, suggesting the applicability of face-to-face rating criteria created by Youn (2015) can be transferable in online role-play assessments. The raters were able to distinguish the rating criteria appropriately and consistently in online settings. However, given that the test-takers' abilities ranged widely (from –6 to +6 logits), the current three-level rating scale for each category may be insufficient. Therefore, additional rating levels may be necessary to more accurately distinguish between different levels of test-taker performance.

The MFRM results also showed that the rating criteria had a similar difficulty level with the face-to-face setting (Youn, 2015). Language use was the most difficult criterion in this study, followed by content delivery. The qualitative analysis with four speech samples revealed further explanations. In terms of language use, low-proficiency participants commonly failed to properly use the correct pragmalinguistic knowledge. In the follow-up interviews, the low-proficiency learners were aware of honorifics but failed to use them correctly. Content delivery was considered the second highest difficult criterion. For L2 learners, how they approach their speech act could be challenging. For example, in the qualitative analysis, the lack of a pre-request sequence made their delivery of speech abrupt. In addition, low-level learners sounded more disfluent delivering their speech act whereas high level learners were able to maintain a longer turn without awkward pauses. Turn organization was more challenging than engaging in interaction in an online setting. Long pauses between turns demonstrated awkward moments between interlocutors on the screen. Another notable feature was the overlapping of the two speakers' voices. At times, due to the nature of the online setting, adjacency pairs—such as greeting exchanges—were disrupted, with the interviewer's greeting overlapping and preventing a response from Speaker 4 (Excerpt 4). This might be related to the Nakatsuhara et al. (2021)'s finding, which shows that L2 test takers in online settings often had clarification requests due to the audio sound in an online meeting platform.

When facing a communication breakdown, managing intersubjectivity becomes a key solution. Highly evaluated performances displayed intersubjectivity in a very subtle way, whereas, the negotiation of meaning, repairs, and struggles in order to manage intersubjectivity are more noticeable for low-proficiency learners. For example, when there was a communication breakdown, Speaker 3 frequently gazed at the screen and referred to her screen without engaging in interactive listening, remaining mostly quiet. When such lack of interactive listening occurs, online communication might seem broken down because the interlocutor (the interviewer) does not know what is going on over the screen.

Sensitivity to the situation differed from face-to-face L2 English assessments by Youn (2015). In the current study, sensitivity to the situation was the easiest criterion on which to achieve a high score. Demonstrating awareness of contextual factors reflects one's sociopragmatic competence in academic settings—particularly in situations like explaining the situations

and reasons. One possible reason might be that the participants were able to refer to the role-play card on the screen while they engaged in their interaction, even though they could not directly read the card. Referring to the role-play card may have functioned as a form of prompting for L2 learners' performances. For example, Burch and Kasper (2021) found that role-play cards can serve as both material resources and embodied actions. Previous research on role-play instruction in OPIs has provided initial evidence that role-play prompts act as constitutive textual objects, with their delivery appearing to follow a standardized pattern (Okada, 2010; Ross, 2017). Since speakers in this study performed the role-play in front of a computer screen, it was difficult to determine exactly what was displayed on their screens during the interaction. Further, more detailed multimodal analysis is needed to better understand L2 learners' performances such as video-recording test takers' actions with the screen.

Several limitations may have influenced the results of this study, necessitating a cautious interpretation of certain findings. First, this study only examined L2 Japanese speaking competence of Japanese learners in professor-student interactions. To gain a more comprehensive understanding of L2 Japanese learners' interactional skills, future research could incorporate interactions between two learners, as explored in previous studies (Youn, 2015, 2020). The speaking competence of L2 learners when communicating with a teacher might be different from learners' speaking competence when communicating with other learners. Another limitation is concerned with collecting data for this study online. Although task procedures for the participants (e.g., warm-up, instructions, and topic card delivery in chatbox) were standardized, several technical factors were not adequately controlled. Participants used their own devices and internet connections, hence variations in audio quality, bandwidth, or latency could have influenced the timing of turns, overlaps, and backchannels. Because such factors are inherent to online communication environments, some interactional patterns observed in this study may partly reflect online platform-related constraints rather than participants' underlying interactional competence. Lastly, proficiency levels were based on self-report and teacher judgment in this study, and this may limit the generalizability of the CEFR labels used.

Despite these limitations, the study offers several key implications. First, the study confirms that interactional tasks can be assessed via online meeting platforms. With the increasing integration of digital communication technologies, online interactions have become more common. There is an immediate need for new rating criteria for assessing speaking online. It was found in this study that the rating criteria for assessing face-to-face interactional competence appear to be transferable to online assessment contexts. Second, this study was an innovative attempt to successfully apply rating rubrics for examining the speaking ability of L2 Chinese learners of Japanese. The applicability of rating rubrics across different target languages suggests that certain features of interactional competence could be universal, aligning with rating criteria (Schegloff, 2007). Finally, the mixed-methods analysis applied in this study provides another perspective for examining L2 learners' interactive competence in the online setting. An increasing number of researchers are incorporating more speaking assessment into the CA-SLA approach for assessing language proficiency (e.g., Dai, 2024; Lam, 2018; Lam et al., 2023; Seedhouse & Nakatsuhara, 2018; Youn, 2015, 2020), enabling a deeper understanding of the interactive features that inform rating constructs. This study contributes empirical evidence that offers a practical foundation for employing speaking assessment tools for digitally mediated interaction.

Conclusion

This study re-examined the assessment criteria originally developed for face-to-face interaction in previous study (Youn, 2015, 2020) into an online meeting platform. The MFRM confirms

the applicability of the rating criteria for assessing L2 Japanese learners' interaction competence in online contexts. Although additional rating levels may be required to differentiate the learners' proficiency levels more precisely, the findings suggest that the rating criteria can be transferable to online role-play assessments in a different L2. Furthermore, the mixed-methods analysis revealed features that are specific to online interaction, particularly when the low-proficient learners struggled to maintain intersubjectivity. In such cases, challenges related to turn-taking and sustained engagement in interaction became especially salient. Additional multimodal analysis of L2 speakers' performances and behaviors should be conducted to fully capture what occurs in front of their screen for future study. In doing so, researchers can better understand how technological mediation shapes interactional competence, thereby informing both assessment design and pedagogical practice. Ultimately, a more refined understanding of online interaction will contribute to the development of more valid frameworks for assessing L2 learners across diverse communicative contexts.

References

Atkinson, J. M., & Heritage, J. (1984). *Structures of social action: Studies in conversation analysis*. Cambridge University Press.

Burch, A., & Kasper, G. (2021). 4 task instruction in OPI roleplays. In M. Salaberry & A. Burch (Eds.), *Assessing speaking in context: Expanding the construct and its applications* (pp. 73–106). Multilingual Matters. https://doi.org/10.21832/9781788923828-005

Burch, A. R., & Kley, K. (2020). Assessing interactional competence: The role of intersubjectivity in a paired-speaking assessment task. *Papers in Language Testing and Assessment*, *9*(1), 25–63.

Dai, D. W. (2025) Assessment of interactional competence. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–7). Wiley.

Dai, D. W. (2024). *Assessing interactional competence: Principles, test development and validation through an L2 Chinese IC test* (Vol. 48). Peter Lang. https://doi.org/10.3726/b21295

Dombi, J., Sydorenko, T., & Timpe-Laughlin, V. (2022). Common ground, cooperation, and recipient design in human-computer interactions. *Journal of Pragmatics*, *193*, 4–20. https://doi.org/10.1016/j.pragma.2022.01.017

Fisher, W. P., Jr. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions, 21*(1), 1095. https://www.rasch.org/rmt/rmt211m.htm

Goodwin, C. (2018). *Co-operative action*. Cambridge University Press.

Grabowski, K. (2013). Investigating the construct validity of a role-play test designed to measure grammatical and pragmatic knowledge at multiple proficiency levels. In S. J. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp. 149–171). Palgrave MacMillan.

Hall, J. K., & Doehler, S. P. (2011). L2 interactional competence and development. In J. K. Hall, J. Hellermann, & S. P. Doehler (Eds.), *L2 interactional competence and development* (pp. 1–15). Multilingual Matters.

Kasper, G., & Youn, S. J. (2018). Transforming instruction to activity: Roleplay in language assessment. *Applied Linguistics Review*, *9*(4), 589–616. https://doi.org/10.1515/applirev-2017-0020

Kasper, G., & Ross, S. J. (2013). Assessing second language pragmatics: An overview and introductions. In S. J. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp. 1–40). Palgrave Macmillan.

Kato, F., Spring, R., & Mori, C. (2023). Incorporating project-based language learning into distance learning: Creating a homepage during computer-mediated learning sessions. *Language Teaching Research, 27*(3), 621–641. https://doi.org/10.1177/1362168820954454

Klipp, J., & Reinders, H. (2025). The effects of different types of computer-assisted corrective feedback on L2 pragmatics learning. *Digital Applied Linguistics, 3*, Article 102515. https://doi.org/10.29140/dal.v3.102515

Lam, D. M. (2018). What counts as "responding"? Contingency on previous speaker contribution as a feature of interactional competence. *Language Testing*, *35*(3), 377–401. https://doi.org/10.1177/0265532218758126

Lam, D., Galaczi, E., Nakatsuhara, F., & May, L. (2023). Assessing interactional competence: Exploring ratability challenges. *Applied Pragmatics, 5*(2), 208–233. https://doi.org/10.1075/ap.00014.lam

Leech, G. (1983). *Principles of pragmatics*. Longman.
Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878.
Linacre, J. M. (2023). FACETS computer program for many-facet Rasch measurement (Version 3.85.1) [Computer software]. Winsteps.com.
McNamara, T., Knoch, U., & Fan, J. (2019). *Fairness, justice & language assessment*. Oxford University Press.
Okada, Y. (2010). Role-play in oral proficiency interviews: Interactive footing and interactional competencies. *Journal of Pragmatics*, *42*(6), 1647–1668. https://doi.org/10.1016/j.pragma.2009.11.002
Roever, C., & Kasper, G. (2018). Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Language Testing*, *35*(3), 331–355. https://doi.org/10.1177/0265532218758128
Ross, S. (2017). *Interviewing for language proficiency.* Palgrave Macmillan.
Ross, S. J., & O'Connell, S. P. (2013). The situation with complication as a site for strategic competence. In S. J. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp. 311–326). Palgrave Macmillan.
Sato, E., Chen, J. C. C., & Jourdain, S. (2017). Integrating digital technology in an intensive, fully online college course for Japanese beginning learners: A standards-based, performance-driven approach. *The Modern Language Journal, 101*(4), 756–775. https://doi.org/10.1111/modl.12432
Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis.* Cambridge University Press.
Seedhouse, P., & Nakatsuhara, F. (2018) *The discourse of the IELTS speaking test*. Cambridge University Press.
Taguchi, N., & Kim, Y. (2018). Task-based approaches to teaching and assessing pragmatics: An overview. In N. Taguchi & Y. Kim (Eds.), *Task-based approaches to teaching and assessing pragmatics* (pp. 2–24). John Benjamins.
Taguchi, N. (2024). Technology-enhanced language learning and pragmatics: Insights from digital game-based pragmatics instruction. *Language Teaching*, *57*(1), 57–67. https://doi.org/10.1017/S0261444823000101
Tang, X., & Taguchi, N. (2021). Digital game-based learning of formulaic expressions in second language Chinese. *The Modern Language Journal, 105*(3), 740–759. https://doi.org/10.1111/modl.12725
Timpe-Laughlin, V., Sydorenko, T., & Dombi, J. (2022). Human versus machine: Investigating L2 learner output in face-to-face versus fully automated role-plays. *Computer Assisted Language Learning, 37*(1–2), 149–178. https://doi.org/10.1080/09588221.2022.2032184
Timpe-Laughlin, V., Dombi, J., Sydorenko, T., & Sasayama, S. (2023). L2 learners' pragmatic output in a face-to-face vs. a computer-guided role-play task: Implications for TBLT. *Language Teaching Research*, Article 13621688231188310. https://doi.org/10.1177/13621688231188310
Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, *32*(2), 199–225. https://doi.org/10.1177/0265532214557113
Youn, S. J. (2020). Managing proposal sequences in role-play assessment: Validity evidence of interactional competence across levels. *Language Testing, 37*(1), 76–106. https://doi.org/10.1177/0265532219860077
Youn, S. J., & Burch, A. R. (2020). Where conversation analysis meets language assessment: Toward expanding epistemologies and validity evidence. *Papers in Language Testing and Assessment, 9*(1), 3–17. https://doi.org/10.58379/DYKI2461

Author Bio

Chie Ogawa, Ph.D. is an associate professor in the Faculty of Cultural Studies at Kyoto Sangyo University, Japan. Her research focuses on speaking assessment, task-based language teaching, and language teacher education. She has published in international journals on L2 speaking development and assessment and is actively engaged in an English teacher education program.

Nancy Shzh-chen Lee, Ph.D. was born in Taiwan and later migrated to Australia with her family. She is currently a faculty member at Osaka University. She is interested in speaking proficiency evaluation and development. Recently, her research interests have expanded to include trilingual education and learning.

Shiming Xia, Ph.D. is an assistant professor at the Professional College of Arts and Tourism in Japan. His research interests include gender studies, individual differences in language learning such as personality, anxiety, and tolerance of ambiguity. He is also actively engaged with teaching and researching languages other than English (LOTE).

Appendix A

Role-Play Cards in Japanese

ロールプレイタスク
Task 1：教授に推薦状を依頼する
状況：あなたは日本語学科の大学生です。日本に大学院進学に行きたいので、応募書類として推薦状が必要です。締め切りは１週間後です。大学院への提出方法は、紙媒体かウェブサイトからの電子送信かどちらかを選んで提出する必要があります。
タスク：先生に推薦状を依頼してください。

Task 2：面談の時間を依頼する
状況：あなたは授業の課題論文についてトピックが決められていないので悩んでいます。教授に相談をして適切なトピックを決めたいです。相談は時間がかかりそうなので、面談を設定して質問したいと思います。
あなたのスケジュール：月・火曜は１０時～１３時まで授業。水木金は午後が空いている。

タスク：先生と面談のアポイントをとってください。

Task 3：依頼を断る
状況：あなたは3日後に大きな期末テストがあり、その授業の勉強を毎晩しています。とても難しい科目なので、テストに向けて猛勉強をする予定です。そんな時に、別の授業を担当する日本語の先生から頼み事をされました。
タスク：先生からの依頼を断ってください。

Role-Play Cards Translated in English

Task 1: Request a letter of recommendation from your professor.
Situation: You are a university student in the Japanese language department. You want to go to graduate school in Japan and need a letter of recommendation for your application. The deadline is in one week. You need to submit the letter to the graduate school either in paper form or electronically through the website.
Task 1: Request a letter of recommendation from your professor.

Task 2: Request time for a meeting
Situation: You are struggling to decide on a topic for a class paper assignment. You would like to discuss with your professor to determine an appropriate topic. The discussion may take some time, so you would like to set up a meeting to ask questions.
Your schedule: Classes are from 10:00 to 13:00 on Monday and Tuesday. You are free in the afternoon on Wednesday, Thursday, and Friday.
Task: Make an appointment with your teacher for an interview.

Task 3: Refuse the request.
Situation: You have a big final exam in three days and you study for that class every night. It is a very difficult subject, and you plan to study hard for the test. At that time, your Japanese teacher, who is in charge of another class, asks you for a favor.
Task: Please refuse the request from the teacher.

Appendix B

Rating Rubrics in Japanese (Adapted from Youn, 2015; 2020)

スコア	流暢さ	言語使用	状況についての認識	やりとりでの関わり	ターン
3 うまく行った	・スムーズに依頼・断りを進められる ・スムーズにトピックを開始できる #1　推薦状の依頼、提出方法の選択 #2　面談の依頼と時間の設定 #3　教授の依頼への応答と，状況説明	・語用論的に適切な表現ができる（～ませんか、～していただけますか、もし良ければ、おねがいできますか） ・語彙と文法が正しい	・内容説明において、状況を把握して説明できる #1推薦状について説明できる。１週間の締め切り・提出方法の選択を説明できる。 #2面談の依頼の説明ができる #3断りの理由を説明できる	・相手の発話を理解している（お互い理解をしている） ・返答に、相槌・確認・質問などの反応がしっかりある。	・ぎこちない沈黙や突然の発話の重なりがない Note: #3断る前の沈黙はOK ・やりとりの返答が隣接ペアとなっている（例：質問—答え、挨拶—挨拶）
2 普通	・時々スムーズではない時がある（言い直し、沈黙、不必要に長い言い回し） ・内容を唐突に始める。	・時々、語用論的な言語表現が不適切である。 ・時々、語彙・文法が不正確な場合がある。	・時々状況を把握しておらず十分な説明がなされない。 ・説明が受け身である #1推薦状の説明ができても、締め切りの説明ができない。#2, 3理由の説明ができない	・時々、前の発話ターンを理解していない時がある。 ・時々、相槌・確認・質問などの反応がない。	・時々、発話ターンが遅れる ・時々、相手の発話を遮断して話す ・隣接ペアではない（例：質問—お礼）
1 うまく行かない	・スムーズではない。（途切れ途切れの話し方で、言い直しが多い）（語学力の低さから）	・語用論的な言語表現が不適切である。 ・語彙・文法が不正確で意味を伝えることができない。 ・ロールプレイカードを読んでいるだけで表現が直接的（～が必要です。～する予定です）	・状況に関して、ほとんど説明ができていない #1締切について認識していない #3断る際の理由が不十分など	・やりとりにおいて誤解が生じている ・返答に、相槌・確認・質問などの反応が全くない。	・突然会話を遮断する ・沈黙が不自然 （例：ターン間に、ひときわ長い沈黙がある、#3反対や断りの前後に沈黙をしない）

#1, 2, 3 = Task 1, 2, 3の意味です。*ロールプレイカードを読んでいるだけ、の場合は、減点してください。

Translated in English (Adapted from Youn, 2015; 2020)

Score	Fluency	Language Use	Sensitivity to situation	Engage in interaction	Turn taking
3 successful	• Clear and fluent speech act delivery (requests and refusals) • Smoothly initiate topics. #1 a letter request & letter submission option #2 need for a meeting & decide a time #3 respond to professor's request & explain a situation	• Pragmatically appropriate expressions (do you mind~~, would you do..? If you don't mind...,) • Correct vocabulary and grammar which does not obscure meaning	• Consistent evidence of awareness and sensitivity to situation and explain. #1: request along with explanation about the application, acknowledge a short due #2: explanations for a meeting request #3: handle a face-threatening refusal with acceptable reasons	• A next turn shows understandings of a previous turn (shared understanding). • Evidence of engaging with conversation (e.g., backchannel, confirmation, or clarification questions, nodding included)	• No awkward silences or abrupt overlap Note: #3 Silence before refusal is acceptable. • Complete adjacent pairs (e.g., question-answer, greeting-greeting).
2 moderate	• Generally smooth but sometimes occasional rephrase, silences, unnecessarily wordy. • Abrupt topic initiation	• Pragmatic language expressions are occasionally inappropriate. • Vocabulary and grammar are limited and obscure meaning.	• Inconsistent of evidence of awareness and sensitivity to situations (e.g., they do not give sufficient explanations.) • Explanations are given after the professor asks for it. #1: explain the letter of recommendation, but not acknowledge a short due day. #2&3: Unable to explain reasons.	• Some evidence of engaging with the conversation but not consistent. • Sometimes, there is no response such as a backchannel, confirmation, or clarification questions.	• Some turns are delayed. • Sometimes abruptly cut off the a previous turn. • Not in adjacent pairs (e.g., question - thank)

(*Continued*)

(*Continued*)

Score	Fluency	Language Use	Sensitivity to situation	Engage in interaction	Turn taking
1 unsuccessful	• Not smooth, delivery is choppy and minimal. (due to lack of language competence)	Inappropriate pragmatic language expression. • Limited vocabulary and misuse of grammar obscure meaning. • Read the role-play cards and the expression is direct (~ is necessary. I need).	• Little evidence of situational sensitivity #1: not acknowledge due day #3: lack of explanation for refusal	• Misunderstandings in the communication. • There is no response to the response, such as confirmation, clarification questions or no nodding.	• Noticeable cutoff between turns • Noticeably long pauses (e.g., there is a very long silence between turns, no silence before or after #3 disagreement or refusal)

#1, 2, 3 = Task 1, 2, 3. *Please deduct a point if students are just reading the role-play cards.

Appendix C

Abbreviations Used in Gloss Translations

CP	copula
FP	sentence final particle
LK	linking particle
N	nominalizer
NEG	negative morpheme
O	object marker
P	particle
PAST	past tense
Q	question marker
QT	quotation marker
S	subject marker
SF	sentence filler
TP	topic marker

Appendix D

Transcription Conventions (Atkinson & Heritage, 1984)

:	Lengthening of the preceding sound
-	Abrupt cutoff (.) Very short untimed pause
> < talk	Talk surrounded by this bracket is produced more quickly than neighboring
[	Point of overlap onset
=	No gap between adjacent utterances
word	Speaker emphasis
CAPITALS	Especially loud sounds relative to surrounding talk
○ ○	Utterances between degree signs are noticeably quieter than surrounding talk
(3.5)	Intervals between utterances (in seconds)