



Lam, D. M.K., Galaczi, E., Nakatsuhara, F. and May, L. (2023) Assessing interactional competence: exploring ratability challenges. *Applied Pragmatics*, 5(2), pp. 208-233.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<https://eprints.gla.ac.uk/284799/>

Deposited on: 7 November 2022

Enlighten – Research publications by members of the University of Glasgow
<https://eprints.gla.ac.uk>

Assessing interactional competence: Exploring ratability challenges

Daniel M. K. Lam¹, Evelina Galaczi², Fumiyo Nakatsuhara³ and Lyn May⁴

¹University of Glasgow ²Cambridge University Press and Assessment ³CRELLA, University of Bedfordshire ⁴Queensland University of Technology

This paper is positioned at the interface of second/foreign language (L2) assessment and Conversation Analysis-Second Language Acquisition (CA-SLA). It explores challenges of ratability in assessing interactional competence (IC) from three dimensions: an overview of the conceptual and terminological convergence/divergence in the CA-SLA and L2 assessment literature, a micro-analytic Conversation Analysis of test-taker interactions, and the operationalisation of IC construct features in rating scales across assessment contexts. It draws insights from these dimensions into a discussion of the nature of the IC construct and the challenges of IC ratability, and concludes with suggestions on ways in which insights from CA research can contribute to addressing these issues.

Keywords: interactional competence, language assessment, conversation analysis, test-taker interactions, rating scales

1. Introduction

Interaction is positioned at the intersection of many second/foreign language (L2) acquisition theories and frameworks, whether they be psycholinguistic, sociocultural or interactional, and is considered to be a prerequisite not just for the exercise of existing linguistic resources, but also for the development of new resources (Mackey, 1999; Kasper, 2006; Markee, 2008). Interactional competence (IC) has, as such, emerged as a central focus in L2 acquisition, learning, and assessment, and specifically within Conversation Analysis-Second Language Acquisition (CA-SLA) and L2 assessment of interaction.

The collective efforts of CA-SLA and L2 assessment have culminated in a substantial body of research on L2 IC development as an object of inquiry (e.g., Pekarek Doehler, 2018, 2019), an objective of language teaching/learning (e.g., Waring, 2018, 2019), and a construct in L2 proficiency assessment (e.g., Galaczi & Taylor, 2018; Roever & Kasper, 2018). Findings from this body of research have implications for the construct definition of IC in L2 assessment and for practical guidance for CA-SLA researchers interested in designing assessment tasks and developing IC performance descriptors as alternative ways to triangulate the tracking and analysis of L2 IC development in learning contexts.

In this paper we attempt to further extend this body of research through exploring theoretical and operational *challenges of ratability* in assessing IC. Our focus is three-fold: (i) the terminological convergence and divergence in CA-SLA and L2 assessment vis-à-vis IC, (ii) the nature and noticeable features of IC recoverable from microanalytic CA of test-taker interactions, and (iii) the operationalisation of IC in rating scales across different assessment contexts. We believe that the complementarity of insights gained from micro-level CA of test-taker interactions alongside a macro-level overview of IC rating scales in use will, despite their differing

orientations, priorities, and usage, contribute to a more in-depth understanding of the challenges inherent in assessing this important L2 construct and potential ways of addressing them.

2. IC in CA-SLA and L2 assessment – A synthesis of terminological orientations

The past two decades have witnessed a surge in theoretical (e.g., Galaczi & Taylor, 2018; Hall & Pekarek Doehler, 2011; Young, 2008) and empirical works (e.g., Galaczi, 2014; Pekarek Doehler & Berger, 2018; Nakatsuhara, May, Lam & Galaczi, 2018) that define the construct of IC and provide fine-grained descriptions of IC features in order to support L2 learning and assessment.

The CA-SLA literature typically describes L2 learners' interactional ability in terms of four fundamental CA organisational mechanisms of interaction (Schegloff, 2007):

- *Turn-taking*: how learners take, time, and distribute interactional turns
- *Action formation*: how they design or format their turns to accomplish different actions, taking into account the context and the recipients, which includes an awareness of *preference organisation*
- *Sequential organisation*: how they organise series of turns pertaining to different actions (e.g., request sequences) or stages (e.g., opening/closing) of the interaction
- *Repair*: how they identify and resolve problems in production, hearing, and understanding of talk – their own or others

These IC organisational mechanisms (with some nuances in terminology) have been the primary analytical lens for a number of seminal CA-SLA works (e.g., Kasper, 2006, Markee, 2008, Young 2008, Hall & Pekarek Doehler, 2011, Pekarek Doehler, 2019), see also Barth-Weingarten/Freitag-Hild (this special issue).

Alongside these interactional mechanisms, CA-SLA researchers have also identified a range of *resources* deployed for organising interaction and coordinating actions. Kasper (2006) and Markee (2008), for example, highlight the use of non-verbal/embodied resources in conjunction with verbal ones, while Hall and Pekarek Doehler (2011) and Pekarek Doehler (2019) identify “prosodic”, “linguistic”, “sequential” and “non-verbal resources” as part of learners' repertoires. The key point, as Pekarek Doehler (2019, p. 37) notes, is the refinement of methods for interaction as learners progress linguistically: “the development of L2 IC can be usefully understood as a progressive diversification of methods for action, comprising an increased sensitivity for such issues as recipient design, context sensitivity and ... preference organization”. The enablers for such a diversification are the resources which learners employ during interactions.

Empirical studies of IC in the majority of L2 speaking assessment literature (published in journals related to language testing and assessment between 2008 and 2021) have concentrated on the role of each individual test-taker in successfully accomplishing the test task (or not) through different interactional behaviours. Such interactional behaviours, often termed IC features in this literature, broadly fall under the following categories:

- *Turn-taking management*: how test-takers take, time, and distribute turns
 - *Topic management*: how they initiate topics, respond to their interactional partner, and maintain (self-initiated) and develop (other-initiated) topics
 - *Interactive listening*: how they demonstrate involvement through listenership displays such as backchannelling and collaborative completions of turns
-

This is an Accepted Manuscript of an article to be published by John Benjamins Publishing Company in Applied Pragmatics in 2023. The publisher should be contacted for permission to re-use or reprint the material in any form.

- *Breakdown repair*: how they deal with conversational trouble, and ask for or provide interlocutor support
- *Non-verbal behaviour*: how they use embodied actions (e.g., gaze, facial expressions, gestures, laughter)
- *Manner of interaction*: how confident and interactionally proactive test-takers present themselves

The above list of IC features has been informed by micro-level analyses of test-taker interactions (e.g., Burch & Kley, 2020; Galaczi, 2008, 2014; Gan, Davison, & Hamp-Lyons, 2009; Hırçın Çoban & Sert, 2020; Lam, 2021; Nakatsuhara, 2013) and by raters' perceptions of what is important in IC assessment (e.g., May, 2011, Nakatsuhara et al., 2018).

The two overlapping areas of interest in CA-SLA and L2 assessment share a number of alignments. *Turn-taking management* and *repair* are explicitly referred to in both approaches, with similar conceptualisations. In contrast, *topic management* and *interactive listening* are explicitly awarded key prominence in L2 assessment research only. These two interactional behaviours are essentially about producing sequentially relevant turns which follow on from prior talk, and therefore have clear overlaps with the CA categories of *action formation* and *sequential organisation*. Interactive listening also relates to the CA notion of reciprocity, where participants display to one another, verbally or non-verbally, their “awareness of being a recipient [listener] during ongoing turns by other co-participants” (Xu, 2014, p. 35), and to intersubjectivity (Burch & Kley, 2020). In L2 assessment, however, the orientation-to-others focus is often seen through the prism of topic management, since learners' ability to sustain topics over multiple turns has been identified as a key distinguishing feature between higher and lower IC levels (Gan, 2010; Galaczi, 2014). Orientation to others and the action of response, therefore, become critical in IC assessment, since they provide *observable evidence* of IC. A close parallel here can be seen in CA's “next-turn proof procedure” (Hutchby & Wooffitt, 1999, cf. Sacks, Schegloff, & Jefferson, 1974, p. 728–729). In CA, the next-turn proof procedure is the key to understanding how participants have interpreted a specific turn. In L2 assessment, it is the key to evaluating test-takers' IC. A further parallel to the L2 assessment focus on topic development can be seen in the CA notion of topic progressivity (Svennevig, 1999), and, more generally, progressivity – i.e., “keeping the interaction moving forward” (Youn & Burch, 2020, p. xv).

A further category – *non-verbal behaviour* – is seen in CA-SLA as part of the “methods” for performing the four main interactional mechanisms. In contrast, in L2 assessment, it is singled out as a distinct feature of the IC construct, although one that presents challenges for assessment (see Section 5).

The last category – *manner of interaction* – is somewhat different from the others, as it relates to test-takers' attitudes and the impressions they give to raters and co-participants. This category has emerged from several rater studies in the L2 assessment literature (e.g., May, 2011, Nakatsuhara et al., 2018) and encapsulates raters' holistic evaluation of test-takers' confidence and willingness to participate in the interaction.

Table 1 provides a summary of the concepts discussed so far.

Table 1. Juxtaposition of concepts used to describe IC in CA-SLA and in L2 assessment

CA-SLA		L2 assessment
Resources	Turn-taking	Turn-taking management
	Action formation	Topic management <ul style="list-style-type: none"> - Topic initiation - Responding to partner - Topic maintenance (of self-initiated topics) - Topic development (of other-initiated topics)
	Sequential organisation	
		Interactive listening
		Non-verbal behaviour
	Repair	Breakdown repair/Providing or needing interlocutor support
	Manner of interaction	

Taken together, the features identified in Table 1, whether they are from a CA-SLA or L2 assessment perspective, could all be said to reflect techniques or methods-for-action related to three key dimensions of interaction: (i) coordination of actions among interlocutors, (ii) development and management of intersubjective understanding, and (iii) maintenance of progressivity in interaction. These aspects of interaction are, in turn, enabled and shaped by a set of linguistic and embodied resources which participants bring to the interaction. We will now turn to exploring these aspects of interaction and accompanying resources through a dual lens: a CA-informed micro-analysis of learner interactions and an analysis of IC assessment criteria used in a range of contexts. The aim in both is to identify challenges with rating IC, and also lay the ground for potential ways of addressing these challenges.

3. A microanalytic approach to IC in test-taker interactions

In this section, we will illustrate two IC features discussed in Section 2 – *topic management* and *interactive listening* – which have prominence in L2 assessment research. We will present excerpts of two paired interactions from the Cambridge Assessment English speaking tests: Excerpt 1 from B1 Preliminary and Excerpt 2 from C1 Advanced¹. Both excerpts involve a discussion task based on a question with six accompanying ideas (presented in pictures). Each idea/picture therefore forms a *potential topic* for test-takers to initiate and develop in the discussion.

The analysis takes an “applied CA” approach used in language assessment research (cf. Lazaraton, 2002; Galaczi, 2014) – using CA concepts and focusing on micro-level interactional features but not engaging in line-by-line sequential analysis. This provides better alignment with our purpose: to illustrate the criterial features distinguishing test-takers' talk at the two levels, and to provide an account of the interactional resources test-takers draw upon.

In Excerpt 1, the two learners (at CEFR Level B1; Council of Europe, 2001) have been asked to discuss different items they would take to help them pass the time while waiting in a long queue for concert tickets. The choices of items include a mobile phone, food, books, and a guitar. (See Appendix 1 for transcription notation.)

Excerpt 1: V and M

(each potential topic from the task prompt is given in bold)

1 M: If you have to be (.) in a:: (.) place for
 2 many hours, (.) eh: it's good,
 3 if you take with you, som:e (.) **+food,** +
 v +nods slightly+
 4 **an:d *water, * or another *drink.**
 m *looks at V* *looks at V again--->
 5 (0.5)
 6 V: Yeah.* An::d, +it's- it's good, +
 v +turns to M and back+
 m --->*

7 I think that uh for our friends,
 8 it's important that uh (.) uh they have
 9 +a:+ **mobile *phone,** * a::nd (.) +maybe **+um*brella,**
 v +points to pic+ +points +looks at M
 m *nods slightly* *nods--->

10 because maybe (.) uhm raining,* a::nd (.)
 --->*

11 +uh- or- uh+ %of course % a **+guitar,** +
 v +looks at M+ +looks at M+
 v %points to pic%

12 because has important +concert. +
 +turns to examiner+

13 (0.6)
 14 M: Mm hmm, +hh
 v +turns to M--->

15 a:nd if you <don't be::> wor-+ ehm (.) worry,
 v --->+

16 in (.) in- in the- in the- in the: (..)
 17 >with the +people?< +
 v +nods slightly+

18 you ca::n (.) *use your **+walkman?*** (.)
 m *looks at V-----*
 v +nods--->

19 Δa:nd+ %listen the:% music that (.) you prefer,
 --->+
 v Δlooks at M--->
 v %smiles %

20 (.) *orΔ +pla:y* in a flittle **ninf *tendo.** *
 m *points to pic*
 m fglances at examinerf *looks at V*
 v --->Δ +looks at pics on test booklet--->

21 V: ↓Yeah ↓sure. Uh (.) a:nd uhm (.) uh-
 22 it's important,+ +the **food.**=*I think+ that uh:m *
 v --->+ +looks at M-----+
 m *nods slightly-----*

23 uh: he need to- +to eat *some (what),*+
 v +looks at M-----+
 m *nods slightly*

24 +a:nd (.) + **cola or +chocolate,**+
 v +points to pic+ +looks at M+

25 *eat an **apple,** is (.) +>very good.<+*

```

m      *nods continuously-----*
v      +turns to M--+
26 M:  Mm hmm,

```

In terms of topic management, even just a brief perusal of Excerpt 1 suggests that M and V's exchange moves through ideas fairly quickly, resulting in short-lived topics. These short-lived topics appear to be a joint product of one participant doing little to establish each idea as a topic, *and* the other participant receipting each idea without expanding on it in the following turn.

Of particular interest is V's turn at lines 6-12, in which the action formed through the concerted deployment of linguistic and embodied resources takes the shape of a *listing* action, with three ideas constructed as a list, rather than each being established as a topic. After receipting M's turn and her suggestion *food, water, or another drink* (one picture) with the minimal response *Yeah*, V suggests three potential items (three pictures) in a row (lines 9-11). Notice how *mobile phone, umbrella* and *guitar* are strung together as a list – syntactically through the frame *it's important that they have [X, and Y, and Z]* (lines 8-11); prosodically through a continuing intonation after each item (lines 9-11), together with the conjunction *and* prolonged and followed by a pause; and non-verbally through pointing to the picture just one or two syllables prior to naming the item (lines 9, 11). The turn thus designed does little to sequentially project a second pair part (SPP) from the co-participant to each suggestion (as asking a question such as “What do you think about X?” would). Indeed, M's responses, which receipt and/or affiliate with V's suggestions, are given by the embodied action of nodding (line 9), contiguously, immediately after the suggestion is proffered, but not verbally in the next turn (an otherwise less contiguous position). Similar patterns of listing items by V and receipt through embodied actions by M are observed at lines 22-26. Thus, we see how the short-lived topic development is interactionally co-constructed, and, arguably, the responsibility falls not only on the (co-)participant not ratifying and developing a topic in the following turn, but also on the first participant's failure to create a sequential environment conducive to ratifying and developing a topic (i.e., projecting a contiguous SPP allowing the co-participant to comment on the suggestion offered; cf., Sacks, 1987).

Turning now to interactive listening, it can be seen that many of the recipient actions are performed with embodied resources by both V and M (e.g., V's receipting of M's suggestions through nods at lines 3, 17, 18; V's displaying of attention through gaze and smile at lines 19-20; M's affiliating with V's ideas through nods at lines 9, 22, 23). Co-construction of listenership is also observed at lines 22, 23, and 25, where M is seen to be nodding in response to V's soliciting of listenership through gazeⁱⁱ. Notably, verbal minimal responses by the listening participant such as *mm hmm* (line 14) and *yeah* (lines 6, 21) are positioned turn-initially following speaker change, but not during the speaking co-participant's ongoing turn. There are also no substantive next-turn responses which demonstrate (not just claim) understanding of the previous speaker's contribution (Lam, 2021).

In contrast to Excerpt 1, Excerpt 2 illustrates a more refined repertoire of IC observed in learners with higher-level linguistic resources (CEFR Level C1). The participants C and L have been asked to discuss different ways in which we use the world around us and environmental problems such practices may cause. They have been provided with accompanying ideas presented in pictures, one of which is a tourist beach.

Excerpt 2: L and C^{iv}

1 C: %Uhm +I th*ink this is showing like+ the: (.)
 c %looks at diff pic-->
 c +points to diff pic +
 l *looks at pic C is pointing at-->

2 uh (effects) of like- (.) tourist in like a
 3 >beautiful area,=+for example, this+ woul-
 c +points again at same pic+

4 would have been a **beach**< in like +(1.0) +
 +raises palm then knock on desk+

5 I don't know (.) pfft for exam+ple Hawaii an:+
 +opens palm-----+

6 it's like (0.5) before (.) humanity started to
 7 co- come there,=it was like (.) maybe (.)
 ve+ry% + +peaceful:,*
 c +sweeps palm flat across+
 -->% +looks at L-->
 l -->*

8 an[d like +(.)=
 c -->+sweeps palm-->

9 L: [(*°oh yes yes°*)
 l *nods *

10 C: =(sea:), and NOW+ YOU can +see that there's
 c -->+ +leans towards pic-->

11 a lot of (.) hot[el:s,] [a:nd]
 12 L: [°mm hmm°] *[and-] and- (.)+
 l *points tw C
 c -->+

13 maybe it could be +also (.) Δbe a problem
 c +looks at L
 Δvideo displays task prompt-->

14 for (0.5) uh:m (.) for the water and- and
 15 how can I [say, uh:m]

16 C: [YEAH exactly.]
 17 L: not only water but=
 18 C: =the EN[VIRONMENT IN ↑GENE[↓RAL,
 19 L: [Ah yes yes [YES INDEED=
 20 C: =I mean (.) t- (.) carsΔ +are starting to co:me, and like=
 -->Δ
 c +gesticulates while looking at pic-->

21 L: =Mm hmm=
 22 C: =exhaust fumes+ and stuff like that.
 -->+

23 L: Yes, and every+thing in- in (.)+ *maybe it was a *
 c +looks at L-----+
 l *points to pic w palm*

24 wood before or (.)
 25 C: +Ye[ah! +
 c +raises palm momentarily+

26 L: [something like that, that they had to (.)
 27 uh::m <burn down>, (0.5) uh: and- and where are-
 28 uh the animals +who who lived+ there in the past¿=

rating recorded performances) under time pressures. Alternatively, such conversational features might be noticed, but interpreted differently, due to their nuanced and highly context-embedded use – the so-called “Rashomon effect” (Seedhouse & Satar, 2021).

4. IC constructs featured in speaking assessment scales

We now shift to a 'macro' lens to gain an overview of IC features captured in various rating scales, and then to a discussion of what challenges there might be for raters to map their observations of test-takers' performance to rating scale descriptors. With that aim in mind, we scrutinised the content of a range of IC-related rating scales that are currently used in standardised and university in-house speaking assessments. A total of 23 rating scales were gathered for this purpose, of which 18 are publicly available scales published by international or national examination boards (e.g., The Michigan Examination for the Certificate of Proficiency in English (ECPE)), and five are in use at a single educational institution (e.g., The Kanda English Proficiency Test (KEPT)) or are the product of research to develop tasks and rating scales (e.g., Ikeda, 2017). A full list of the rating scales can be found in Appendix 2.

The coding of the IC descriptors proceeded in four main steps:

Step 1: Three sets of descriptors per scale – those for the highest, pass/borderline, and low-level performances – were selected for analysis. All the descriptors at those levels were examined in order to identify the descriptors relating to IC.

Step 2: The selected descriptors were collated in table format in Microsoft Word, alongside the IC features presented in the “L2 assessment” column of Table 1. Given the modest sample size and to allow ease of comparing all 23 scales, manual colour-coding of the descriptors was applied.

Step 3: Two of the researchers independently coded the selected descriptors against the list of IC features, and then discussed their coding. Only minor discrepancies between the two coders were found, which were discussed until full agreement was reached.

Step 4: The researchers also noted down features emerging from the descriptors beyond the features in Table 1. Those features were discussed and a consensus was reached whether they could be considered stand-alone (feature) categories or part of the initial (Table 1) categories. All the descriptors were then reviewed, the coding was checked in light of the new understanding of the coding scheme, and re-coding was done where relevant.

The analysis indicated that *understanding of situation and context* emerged as an IC feature additional to the “L2 assessment” column of Table 1 (see Table 2, last row), which resonates with the IC conceptualisation of Hall and Pekarek Doehler (2011). As part of the analysis, we also broadened the definition of interactive listening to encompass *listening-in-speaking*, where the focus is on listening comprehension. How this comprehension could be determined is not specified in some of the rating scales, but as evidence of comprehension, some recent empirical studies suggest contingent responses, collaborative completions (Lam, 2021), and displays of stance using laminating intonation and embodied resources (Burch & Kley, 2020). Drawing on these studies, these forms of evidence of comprehension became part of the coding for interactive listening.

A caveat noted throughout the coding process is that our intention was to discern and extract underlying core IC categories from differently worded descriptors in the selected scales, in order to understand broad trends in IC assessment scales. Each rating scale was developed in relation to the operationalisation of speaking constructs at different proficiency levels, through specific tasks,

This is an Accepted Manuscript of an article to be published by John Benjamins Publishing Company in Applied Pragmatics in 2023. The publisher should be contacted for permission to re-use or reprint the material in any form.

and for different contexts/purposes. For example, relevant IC features vary with test format (e.g., candidate-examiner vs. paired/group), task (e.g., discussion vs. role-play) and test type (e.g., standardised test vs. classroom assessment). It is also often the case that only a subset of tasks within an assessment is suitable for eliciting specific IC features. As such, the coverage of IC aspects is not intended to be indicative of the quality of the test. Rather, this scale analysis aims to showcase the range of IC features in assessment scales, and not evaluate or equate them.

The results of the coding of IC features captured in the selected rating scales are summarised in Table 2.

Table 2. Features of IC identified in speaking assessment rating scale

Test	ISE (4 levels)		Cambridge English (5 levels)	ECPE	OPI	TEAP	CET 4 and CET 6	HK SBA	Spanish 1 Final Oral test	EAP Speaking test	Ikeda (2017)*	Youn (2015)*	KEPT
Scale	Communicative Effectiveness	Interactive Listening	Interactive Communication	Discourse & Interaction	Global Tasks & Functions	Interactional Effectiveness	Flexibility & Appropriateness	Assessment Criteria for Group Interaction	Interaction	Interactional Effectiveness	Social actions to achieve the communicative goal; Engagement in the interaction; Turn organisation	Content delivery; Sensitivity to situation; Engaging with interaction; Turn organisation	Communicative Skills/strategies
Tasks [Example level selected for suites of tests]	[ISEIII] 1.Topic 2.Collaborative 3.Conversation		[C1 Advanced] 1.Interview 2.Long turn 3.Discussion (in pairs) 4.Three-way discussion	1.Introduction 2.Explaining & recommending 3.Consensus reaching 4.Presenting & convincing 5. Justifying & defending	1.Introduction 2.Warm-up 3.Interview 3.1.Level Checks 3.2.Probes (role play) 4.Cool-down	1.Interview 2.Role-play 3.Monologue 4.Discussion	[CET 6] Paired discussion on a given topic	Group discussion on a given topic	Paired discussion on three given topics	Structured discussion in pairs on a given reading texts	Monologic recorded message tasks & roleplay tasks (request, explanation) with examiner taking role of peer & professor	Role plays with classmate & professor	Group discussion based on a prompt
Target level**	A2 – C1		A2 – C2	C2	A1 – C2	Below A2 – B2	Levels 4 – 6 in China's Standards of English	Approx. B1 – C1	Beginners	Approx. B2-C1	IELTS 5: 'near native' level	TOEFL iBT scores 65-111	Approx. A2-B2
IC categories													
Turn-taking management	√		√					√	√		√	√	√
Topic management	√		√		√			√				√	√
- Topic initiation	√		√		√			√	√	√	√		√
- Responding to partner	√	√	√	√	√	√		√	√	√	√		√
- Topic maintenance	√	√	√	√	√	√	√	√	√		√	√	√
- Topic development	√	√	√	√	√	√		√					√
Interactive listening		√		√		√		√	√	√	√	√	
Non-verbal behaviour								√	√			√	
Breakdown repair/ interlocutor support	√	√	√	√		√		√			√		
Manner of interaction			√	√	√	√							√
Understanding of situation and context				√			√				√	√	

*Experimental study to develop and validate tasks and rating scales for assessing L2 interactional pragmatics.

**All given as CEFR levels, unless specified otherwise.

4.1 Overall trends

The following broad observations emerge from Table 2:

- The most widely targeted IC features are *Topic maintenance* and *Responding to partner*
- *Turn-taking management*, *Topic initiation*, *Topic development*, *Interactive listening*, and *Breakdown repair/Interlocutor support* are also features frequently denoted in the scales
- Less frequent reference is made to *Non-verbal behaviour*, *Manner of interaction*, and *Understanding of situation and context*
- All scales position specific IC features as ratable according to each *individual's contribution* to the co-constructed interaction

We now turn to a more detailed look at the ways in which IC is operationalised in examples from two different contexts: one standardised test (Cambridge C1 Advanced Speaking test) and one classroom-based test (Spanish 1 Final Oral test). They are chosen to exemplify different approaches to the range of IC features which could be included.

4.2 Cambridge C1 Advanced Speaking test

We start with the C1 Advanced test, from which Excerpt 2 in Section 3 was taken. In one of the four parts of the test, two test-takers engage in a paired discussion task and are assessed by two examiners on their individual performance (UCLES, 2019). The examiner who manages the test gives a holistic *global achievement* score, and the observing examiner awards analytic scores for five criteria: *grammatical resource*, *lexical resource*, *discourse management*, *pronunciation*, and *interactive communication*. Table 3 illustrates Bands 1, 3, and 5 of the *interactive communication* criterion.

Table 3. Cambridge C1 Advanced interactive communication rating scale

5	Interacts with ease, linking contributions to those of other speakers. Widens the scope of the interaction and negotiates towards an outcome.
3	Initiates and responds appropriately, linking contributions to those of other speakers. Maintains and develops the interaction and negotiates towards an outcome.
1	Initiates and responds appropriately. Maintains and develops the interaction and negotiates towards an outcome with very little support.

A close scrutiny of these descriptors suggests that the majority of the IC features from the terminological discussion in Section 2 and the micro-analysis in Section 3 are represented in these descriptors. In particular, while test-takers' contributions are individually evaluated, their engagement in collaborative topic development, as observed in Excerpt 2, is explicitly denoted (e.g., *linking contributions to those of other speakers*). However, neither non-verbal behaviour nor interactive listening are included in the scales, although aspects of the latter could be assumed to underlie the ability to *respond appropriately*. Additionally, in contrast to the micro features identified in the CA excerpts, the broad-brush nature of the descriptors' wording is noteworthy. This is understandable, given the need to strike a balance between construct coverage and usability of rating scales – rating descriptors have to cover the measured construct as fully as possible, but also be succinct and easily interpreted by raters.

4.3 Spanish 1 Final Oral test

Our second example comes from an Australian university beginner-level Spanish course, which was the context for Ducasse's (2010) empirical IC scale development. The test featured a paired discussion task used in a classroom-based assessment, where learners were asked to discuss three given topics drawn from course content (e.g., family, public holidays, friends). Students could begin speaking on any of the three topics, with the onus on them to introduce, maintain and change between these topics. Working from rater verbal reports made on the specific cohort of students, Ducasse identified three IC features that raters attended to: non-verbal interpersonal communication (including gaze and body language); interactive listening (including supportive listening and comprehension); and interactional management (including within and across turns). These features were then reflected in an empirically derived, binary-choice, boundary definition (EBB) rating scale (Upshur & Turner, 1995).

As we can see from Table 4, the initial question focuses on non-verbal aspects of IC (*Supportive body language?*) and is the first point of difference used to place learners into higher/lower-scoring groups. Different questions are then used for the upper group *Supportive listener?* and the lower group *Relevant questions/answers are given?*, and this process continues until the learner is awarded a score from 1–5.

Table 4. Interaction scale for Spanish 1 Final Oral test (reproduced from Ducasse, 2010, p. 124)

Question 1	answer	Question 2	answer	Question 3	answer	rating
Supportive body language?	yes	Supportive listener?	yes	Questions/Replies mostly show cohesion between and within topics?	yes	5
					no	4.5
			no	Relevant questions/answers are given?	no	Reasonable turn length?
	yes	no			3.5	
		no			Relevant questions/answers are given?	yes
	no		2			
			no			1

It is notable that the first question concerns non-verbal behaviour, while the second question for stronger students is based on whether they have demonstrated supportive listening skills. Foundational aspects of the IC construct relevant to beginners are thus foregrounded in a way that distinguishes this scale from others, as non-verbal behaviour is not commonly featured in most IC rating scales (see Table 2).

This assessment approach demonstrates the ways in which classroom-based assessment of IC can be sensitive to aspects such as non-verbal behaviour in ways that may not be possible in large-scale assessment of IC, where the focus is on a set of tangible criterial features that raters can be trained to recognise and interpret as uniformly as possible. In classroom assessments, teachers can develop bespoke, context-sensitive speaking assessment tasks reflecting their learners' current proficiency level and specific interactional behaviours meaningful to their particular group of learners, thus providing opportunities for engaging with a wider range of IC features and through a longitudinal lens.

5. Discussion: The nature of the IC construct and ratibility challenges

The micro-analysis of test-taker interactions and the analysis of IC rating scales in the previous sections exemplify three characteristics that are in the very essence of IC, but nonetheless present ratability challenges in assessment contexts: (i) the shared *and* individual elements of IC, (ii) the multi-modal/multi-semiotic nature of IC, and (iii) the high-inference nature of IC observations.

5.1 Interactional competence: shared or individual?

Our synthesis of conceptual and terminological IC orientations in CA-SLA and L2 assessment identified the notions of intersubjectivity and co-construction, which are fundamental to IC, and which characterise IC as essentially a *shared* accomplishment. We also noted the role of resources associated with individual participants, which in turn, allows the interpretation and evaluation of IC as an *individual* ability. The analysis of Excerpts 1 and 2 demonstrated how topic management is a shared responsibility, and displays of listener support can be co-constructed (solicited and proffered). However, individual test-takers at different proficiency levels were also seen to mobilise wider/narrower ranges of resources for topic management and interactive listening, demonstrating observably different degrees of diversification of methods and ability to monitor co-participants' talk (Pekarek Doehler & Berger, 2018). The analysis of IC rating scales also provided evidence that IC is commonly operationalised as an individually ratable trait. This tension presents obvious challenges for the ratability of IC.

Theoretical insights about the shared/individual IC tension are featured in Waring (2018) and Hall (2018), who elucidate the ambiguity of “competence” conceptualised as a *psychological* or a *social* entity. Pekarek Doehler (2019, p. 50) extends this line of argument: “CA-SLA research ... warns against an individualistic view of IC as ‘belonging’ to a participant, as being simply ‘brought along’ to social interaction, rather than being brought about in concord with others.” This tension takes centre stage in L2 assessment, as seen in debates about the psycholinguistic vs. socio-cognitive nature of Communicative Language Ability (Bachman & Palmer, 1996). McNamara and Roever (2006, p. 51) argue that institutional needs are a key driver for the psycholinguistic orientation, since “what is required is not a faithful account of the interaction but a score about individual candidates that can then be fed into the institutional decision-making procedures”. Institutional needs aside, to position IC as only a shared achievement and non-existent within the individual might mean that there is “little that language testers can do with the construct” (Fulcher, 2010, p. 112), since the measurement of an individual's language ability and the generalisability of that measurement across domains is fundamental to L2 assessment.

A middle-ground position that reconciles the two views is the position we argue for in this paper: a hybrid individual/social view, which recognises interaction as a co-constructed, shared achievement that is enabled by individuals' linguistic and embodied resources. Our position aligns with Nguyen (2019: 401), who contends that “what participants do in social interaction, ... how they negotiate the employment of interactional resources, is co-constructed, but the abilities to recognize and mobilize interactional resources must also be individual”. This reconciled position allows for (some) interactional features to be individually oriented and scored. It also implies that shared IC scores could be given in certain contexts and for certain IC features, as suggested by May (2011), who argues that aspects of mutual accomplishments that deserve shared scores include topic management, turn management, and breakdown repair, as well as the quality and authenticity of the interaction. Of key importance is the need to tailor the operationalisation of IC features in accordance with test purpose and high/low stakes. For example, individual scoring of test-takers' IC would be necessary for certifying learners'

proficiency for gatekeeping purposes; whereas joint scores for IC are more feasible in contexts of assessing learners for diagnostic and feedback purposes.

5.2 Interactional competence: meaning-making through diverse semiotic systems

The second ratibility challenge stems from IC being a fundamentally multi-modal and multi-semiotic construct, with meaning created through the interplay of multiple semiotic systems. Bull and Anstey (2010) delimit five such systems: *linguistic*, *visual*, *aural*, *gestural*, and *spatial*, which are intertwined together to shape meaning. Goodwin (2017, p. 134) argues along similar lines:

Human beings are able to build action by joining together semiotic materials with quite different properties, such as talk with its language structure, prosody, and the visible organization of embodied participation... Within an interactive field the visible body is not mute linguistically.

Through the microanalytic lens of CA in Section 3, we illustrated how the activity of “listing” ideas is accomplished through the orchestration of embodied and prosodic resources together with linguistic ones (Excerpt 1), and how participants provide listener support and evidence their understanding through a curated range of verbal and non-verbal resources sensitive to the local sequential context (Excerpt 2).

Indeed, the IC literature in both CA-SLA and L2 assessment (Section 2) has identified non-verbal resources as integral to IC and demonstrated the skilful coordination of resources across diverse semiotic systems. In L2 assessment, this position is further supported by studies probing into rater perceptions about what is important to them when evaluating test-takers’ IC (e.g., Ducasse, 2010, May, 2011) and studies investigating the impact of non-verbal resources on scores (e.g., Jenkins & Parra, 2003). In the latter study, which involved a 20-minute interview test evaluated for pronunciation, grammar, fluency, and comprehensibility, participants with high proficiency in these categories passed the test regardless of their non-verbal actions; importantly, however, so did participants with lower proficiency who could employ effective non-verbal resources. Roever and Kasper (2018) argued further and posited non-verbal aspects of interaction not as simply coping strategies when language knowledge is deficient but integral aspects of successful interactions.

Inevitably, a construct drawing on diverse semiotic systems presents ratibility challenges (and questions of remit) in language assessment contexts, since it extends beyond the linguistic system. With these challenges in mind, arguments for the role of IC in L2 assessment have been presented from opposing sides in the context of L2 speaking assessment: at the reductive end of the spectrum, van Moere (2012) argues for a predominantly psycholinguistic focus on language “facility” in L2 speaking assessment and questions the focus on IC; at the expansive end, Galaczi and Taylor (2018) maintain the importance of a broad socio-cognitive multi-layered construct of IC. These positions capture the conundrum of operationalising IC in L2 assessment. On the one hand, some consider non-verbal resources or actions to be outside the remit of a *language* test, and they constitute unwanted measurement error that threatens validity – a position which may partly account for our finding in Section 4, whereby embodied resources are generally *not* captured in IC assessment scales. On the other hand, excluding non-verbal elements from evaluations of test-takers’ spoken language proficiency seems to be asking raters to consciously oversimplify what they assess – as if one is asking an audience to evaluate their experience at a singer’s concert while consciously ignoring the band, the choreography, and lighting effects.

This ratability challenge relates to the integrated role and coordinated use of non-verbal semiotic resources together with linguistic ones evidenced in this paper and the wider CA-SLA and L2 assessment literature. While raters, through their training and experience, are conversant with noticing formal linguistic features and perhaps some aspects of IC operationalised in current rating scales, they may be less sensitive to noticing the ways in which multi-semiotic resources are coordinated to accomplish interactional actions, and variations in such coordination across different proficiency levels.

5.3 Interactional competence: high- vs. low-inference features

In light of the earlier theoretical discussions and definitions of IC, the findings from the analyses of test-taker interactions (Section 3) and IC rating scales (Section 4) have relevance for the level of inference required. Fulcher (1993, p. 132) calls those assessment scales which relate to specific observable behaviour *low-inference* scales, and those that require inferences about why the observed behaviour is occurring *high-inference* scales. For instance, a low-inference scale may refer to the total number of long pauses observed, while a high-inference scale may concern why test-takers paused on each occasion (e.g., looking for ideas or for language, struggling to understand co-participants' utterances or to determine appropriate next moves, or inserting hesitations to mark dispreferred responses). While "low-inference" scales would generally lead to increased reliability, Fulcher warns that they could trivialise the measured construct by making raters' judgement less context-dependent – in the worst case, resulting in mere box-ticking activities. The rating of IC could be argued to be high inference in nature, since any given IC performance needs to be interpreted in conjunction with the specific moment of talk-in-interaction. The collated IC rating scales referred to here (Table 2) suggest that the more low-inference aspects of IC – *Topic maintenance* and *Responding to partner* – lend themselves to being "noticed" in ways that enable a rater to interpret relevant IC descriptors in a rating scale and apply them to a test-taker's performance. The more high-inference aspects of IC – *Non-verbal behaviour*, *Understanding of situation and context*, and *Interactive listening* (including listenership displays and contingent responses) – may add complexity and uncertainty that is difficult for raters to deal with, particularly in standardised tests. The coordination of linguistic, prosodic, and embodied resources to accomplish actions (e.g., taking a turn, establishing a topic) that are illustrated in Excerpts 1 and 2 contribute to the high-inference nature of IC. The joint responsibility among participants for whether and how a topic develops (see Excerpt 1) may also not be immediately apparent during the test and could also be regarded as high inference.

6. Conclusion

In this paper we have explored ratability challenges for assessing IC through two complementary but fundamentally different lenses: micro-level CA and macro-level IC scale descriptors. Evaluating IC through the lens of CA is like viewing synchronised diving in slow-motion during a video playback; whereas evaluating IC in operational rating contexts is like viewing and scoring the same synchronised diving performance in real time as it unfolds. This, in essence, has been the dual focus of this paper: through a detailed CA we have identified the richness of micro-level resources which learners/test-takers employ as they co-construct interaction, and we have concurrently explored the possibilities and constraints of IC evaluation instruments used in real time. This micro/macro approach has highlighted three key IC ratability

challenges: the shared vs individual nature of the construct, its embedding in diverse semiotic systems, and its high-inference nature. The ultimate challenge is for raters to map co-constructed and multi-semiotic IC practices and resources in test-takers' performance onto more abstract rating scale categories/descriptors, and to do so both faithfully and efficiently (see also Barth-Weingarten/Freitag-Hild, this special issue).

What innovations in assessment practice might help address such ratability challenges? We believe that strengthening rater training has a key role to play. Raters can only evaluate what they notice, which points to the importance of clearly defining the construct and raising raters' awareness of salient construct features. One way in which CA-SLA could inform IC assessments is in raising awareness of the multitude of verbal and non-verbal resources and their coordinated use in interaction, especially resources which might otherwise be perceived as symptomatic of inadequate L2 proficiency (e.g., delays/hesitations adumbrating dispreferred responses in Roever & Kasper, 2018).

We are not suggesting that a microanalytic approach should be incorporated into live rating procedure. Rather, we believe that CA can be a useful tool for rater awareness-raising to address some of the ratability challenges discussed above. As part of rater training, reviewing transcripts of interactive performances and employing the next-turn proof procedure to describe test-takers' conduct (Hırçın Çoban & Sert, 2020; see also Schirm et al., this special issue) could facilitate discovery of more interactional resources and practices, and their coordinated use in accomplishing actions, which might not be readily noticeable in live rating. Such a CA-based training activity might also be useful in bringing to raters' awareness the interactional import of the same conversational object (e.g., "yeah") in different interactional contexts (e.g., within prior speaker's turn space vs. at the beginning of a speaker's own turn; following a turn that makes relevant agreement vs. following one seeking compliance). Indeed, a recent investigation by Seedhouse and Satar (2021) has demonstrated the value of a technology-mediated granular investigation of features of speech in an interactional (test) task. Such an approach, underpinned by technology, could provide a practical solution to raising awareness of conversational detail in assessment and classroom contexts. The widespread recent use of video-call technology in assessment could feasibly enable such a complement to rater training by providing access to video-recorded interactional data, including non-verbal behaviours. In closing, we would also like to highlight the importance and potential fruitfulness of more empirical research on rating IC performances. Based on the current discussion, we call for further studies of rater cognition when assessing IC – e.g., what are perceived as high- and low-inference IC features; how do raters interpret different linguistic and embodied resources (as linguistic deficiency or as interactional resources); and empirical investigations of the potential usefulness of a CA-based rater training activity for assessing IC. Moreover, as the field of L2 assessment increasingly embraces AI, the potential usefulness of spoken dialogue systems in eliciting IC performances and AI technology in evaluating IC features could also be worthwhile avenues to pursue (e.g., Ockey & Chukharev-Hudilainen, 2021). Complementary strengths of machine scoring systems and human raters can also be further explored (e.g., Davis & Papageorgiou, 2021). A hybrid-rating model where formal linguistic features are scored by a machine and IC features are assessed by human raters may lead to optimal use of rater resources, enabling a deeper engagement with evaluating high-inference IC features.

Such avenues for further exploration are not possible without an in-depth understanding of IC. We hope that our study, through its complementary use of a deep-dive into CA transcripts and a birds-eye overview of IC assessment scales, has expanded such an understanding and has provided a springboard for further research into assessing IC. That pursuit will benefit not just

the academic community, but also assessment developers and raters and, ultimately, language learners.

Notes

- ⁱ Limitations of our sampling must be acknowledged: these publicly available samples only represent part of a wider range of test-taker interactions with different interactional patterns and dynamics (cf. Galaczi, 2008).
- ⁱⁱ As a reviewer noted, the side-by-side seating arrangement in this paired speaking test may have an influence on the ways in which test-takers display listenership to one another. Research into the affordances/constraints of seating arrangements and their impact on interaction (cf. Auer & Zima 2021) in assessment settings would be useful in the future.
- ⁱⁱⁱ Layering different semiotic resources together to make meaning and build actions (Goodwin, 2013).
- ^{iv} Embodied actions at lines 14–19 not transcribed because the video displayed the task prompt at the time.

References

- Auer, P., & Zima, E. (2021). On word searches, gaze, and co-participation. *Gesprächsforschung*, 22, 390–425.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford University Press.
- Bull, G., & Anstey, M. (2010). *Evolving Pedagogies: Reading and Writing in a Multimodal World*. Education Services Australia.
- 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.
- Council of Europe (2001). *Common European Framework of References for Languages: learning, Teaching, Assessment*. Cambridge University Press.
- Davis, L. & Papageorgiou, S. (2021). Complementary strengths? Evaluation of a hybrid human-machine scoring approach for a test of oral academic English, *Assessment in Education: Principles, Policy & Practice*. Advance online publication.
- Ducasse, A. M. (2010). *Interaction in Paired Oral Proficiency Assessment in Spanish*. Peter Lang.
- Fulcher, G. (1993). *The construction and validation of rating scales for oral tests in English as a foreign language* [unpublished PhD thesis]. Lancaster University.
- Fulcher, G. (2010). *Practical Language Testing*. Hodder Education.
- 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. <https://doi.org/10.1080/15434300801934702>
- Galaczi, E. D. (2014). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests?. *Applied Linguistics*, 35(5), 553–574. <https://doi.org/10.1093/applin/amt017>
- Galaczi, E. D., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, 15(3), 219–236. <https://doi.org/10.1080/15434303.2018.1453816>
-

- Gan, Z. (2010). Interaction in group oral assessment: A case study of higher- and lower-scoring students. *Language Testing*, 27(4), 585–602.
- Gan, Z., Davison, C., & Hamp-Lyons, L. (2009). Topic negotiation in peer group oral assessment situations: A conversation analytic approach. *Applied Linguistics*, 30(3), 315–334. <https://doi.org/10.1093/applin/amn035>
- Goodwin, C. (2013). The co-operative, transformative organization of human action and knowledge. *Journal of Pragmatics*, 46(1), 8–23. <https://doi.org/10.1016/j.pragma.2012.09.003>
- Goodwin, C. (2017). *Co-operative Action*. Cambridge University Press.
- Hall, J. K. (2018). From L2 interactional competence to L2 interactional repertoires: Reconceptualising the objects of L2 learning. *Classroom Discourse*, 9(1), 25–39. <https://doi.org/10.1080/19463014.2018.1433050>
- Hall, J. K., & Pekarek Doehler, S. (2011). L2 interactional competence and development. In J. K. Hall, J. Hellermann & S. Pekarek Doehler (Eds.), *L2 Interactional Competence and Development* (pp. 1–18). Multilingual Matters.
- Hırçın Çoban, M., & Sert, O. (2020). Resolving interactional troubles and maintaining progressivity in paired speaking assessment in an EFL context. *Papers in Language Testing and Assessment*, 9(1), 64–94.
- Hutchby, I., & Wooffitt, R. (1999). *Conversation Analysis: Principles, Practices and Applications*. Polity Press.
- Ikedda, 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* [unpublished PhD thesis]. University of Melbourne.
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation Analysis: Studies from the First Generation* (pp.13–31). John Benjamins.
- Jenkins, S. & Parra, I. (2003). Multiple layers of meaning in an oral proficiency test: The complementary roles of nonverbal, paralinguistic, and verbal behaviors in assessment decisions. *The Modern Language Journal*, 87(1), 90–107. <https://doi.org/10.1111/1540-4781.00180>
- Kasper, G. (2006). Beyond repair: Conversation analysis as an approach to SLA. *AILA Review*, 19, 83–99. <https://doi.org/10.1075/aila.19.07kas>
- Lam, D. M. K. (2021). Don't turn a deaf ear: A case for assessing interactive listening. *Applied Linguistics* [Advance articles], 1–25. <https://doi.org/10.1093/applin/amaa064>
- Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests*. Cambridge University Press.
- May, L. (2011). Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly*, 8(2), 127–145. <https://doi.org/10.1080/15434303.2011.565845>
- Markee, N. (2008). Toward a learning behavior tracking methodology for CA-for-SLA. *Applied Linguistics*, 29(3), 404–427. <https://doi.org/10.1093/applin/amm052>
- Mackey, A. (1999). Input, interaction, and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition*, 21(4), 557–587. <https://doi.org/10.1017/S0272263199004027>
- McNamara, T. F., & Roever, C. (2006). *Language Testing: The Social Dimension*. Blackwell.
- Nakatsuhara, F. (2013). *The Co-construction of Conversation in Group Oral Tests*. Peter Lang.
- Nakatsuhara, F., May, L., Lam, D., & Galaczi, E. D. (2018). Learning-oriented feedback in the development and assessment of interactional competence. *Research Notes*, 70, 1–68.
-

- Nguyen, H. T. (2019). Deepening roots and broadening horizons in interactional competence research and praxis. In M. R. Salaberry & S. Kunitz (Eds.), *Teaching and Testing L2 Interactional Competence* (pp. 397–412). Routledge.
- Ockey, G. J., & Chukharev-Hudilainen, E. (2021). Human versus computer partner in the paired oral discussion test. *Applied Linguistics* [Advance articles], 1–21. <https://doi.org/10.1093/applin/amaa067>
- Pekarek Doehler, S. (2018). Elaborations on L2 interactional competence: The development of L2 grammar-for-interaction. *Classroom Discourse*, 9(1), 3–24. <https://doi.org/10.1080/19463014.2018.1437759>
- Pekarek Doehler, S. (2019). On the nature and the development of L2 interactional competence. In M. R. Salaberry & S. Kunitz (Eds.), *Teaching and Testing L2 Interactional Competence* (pp. 25–59). Routledge.
- Pekarek Doehler, S., & Berger, E. (2018). L2 interactional competence as increased ability for context-sensitive conduct: A longitudinal study of story-openings. *Applied Linguistics*, 39(4), 555–578. <https://doi.org/10.1093/applin/amw021>
- 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>
- Sacks, H. (1987). On the preference for agreement and contiguity in sequences in conversation. In G. Button & J. Lee (Eds.), *Talk and Social Organization* (pp. 54–69). Multilingual Matters.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735. <https://doi.org/10.2307/412243>
- Schegloff, E. A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis*. Cambridge University Press.
- Seedhouse, P., & Satar, M. (2021). The Rashomon effect: which features of a speaker’s talk do listeners notice? *Classroom Discourse*, 1–23. Advance online publication.
- Svennevig, J. (1999). *Getting acquainted in conversation: A study of initial interactions*. John Benjamins. <https://doi.org/10.1075/pbns.64>
- UCLES (2019). *C1 Advanced: Handbook for Teachers*. Retrieved from: <https://www.cambridgeenglish.org/images/167804-cambridge-english-advanced-handbook.pdf>
- Upshur, J. A., & Turner, C. E. (1995). Constructing rating scales for second language tests. *ELT Journal*, 49(1), 3–12. <https://doi.org/10.1093/elt/49.1.3>
- van Moere, A. (2012). A psycholinguistic approach to oral language assessment. *Language Testing*, 29(3), 325–344. <https://doi.org/10.1177/0265532211424478>
- Waring, H. Z. (2018). Teaching L2 interactional competence: Problems and possibilities. *Classroom Discourse*, 9(1), 57–67. <https://doi.org/10.1080/19463014.2018.1434082>
- Waring, H. Z. (2019). Developing interactional competence with limited linguistic resources. In M. R. Salaberry & S. Kunitz (Eds.), *Teaching and Testing L2 Interactional Competence* (pp. 215–227). Routledge.
- Xu, J. (2014). Displaying status of reciprocity through reactive tokens in Mandarin task-oriented interaction. *Journal of Pragmatics*, 74, 33–51. <https://doi.org/10.1016/j.pragma.2014.08.008>
- 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>
- Young, R. F. (2008). *Language and Interaction: An Advanced Resource Book*. Routledge.
-

Address for correspondence

Daniel M. K. Lam
School of Education
University of Glasgow
11 Eldon Street, Glasgow G3 6NH, UK
Daniel.Lam@glasgow.ac.uk
ORCID: <https://orcid.org/0000-0003-3451-7158>

Co-author information

Evelina D. Galaczi
Cambridge University Press and Assessment
University of Cambridge, UK
ORCID: <https://orcid.org/0000-0001-8269-8462>

Fumiyo Nakatsuhara
Centre for Research in English Language Learning and Assessment (CRELLA)
University of Bedfordshire, UK
ORCID: <https://orcid.org/0000-0002-0429-7417>

Lyn May
Queensland University of Technology, Australia
ORCID: <https://orcid.org/0000-0002-2611-7210>

Appendices

Appendix 1: Transcription notation

This paper uses Jefferson's conventions for transcribing talk (Jefferson, 2004), and Mondada's conventions for transcribing embodied actions (<https://www.lorenzamondada.net/multimodal-transcription>). Some are illustrated below for readers' reference.

= =	Latching of successive talk, of one or more speakers, with no beat of silence between the talk
(.)	An untimed short pause. Number of dots indicates relative length of the pause.
(0.5)	A timed pause (in seconds)
!	Emphatic intonation
¿	A weaker rise in intonation compared to ?
>word<	Parts of talk faster / slower than the surrounding talk
<word>	
↑ ↓	Marked rise / fall in pitch
* *	Embodied action synchronised with a particular stretch of talk
+ +	
△ △	
% %	
£ £	
*--->	Action continues across lines until the same symbol is reached
--->*	

Appendix 2: IC-related scales

IC-related scales from standardised tests

- Scales 1-4: Integrated Skills in English (ISE) examination: ‘Communicative Effectiveness’ across 4 levels (Foundation, I, II, III) – Trinity College London, <https://www.trinitycollege.com/qualifications/english-language/ISE/ISE-results-and-certificates/ISE-rating-scales>
- Scales 5-8: Integrated Skills in English (ISE) examination: ‘Interactive Listening’ scales across 4 levels (Foundation, I, II, III) – Trinity College London, <https://www.trinitycollege.com/qualifications/english-language/ISE/ISE-results-and-certificates/ISE-rating-scales>
- Scales 9-13: Cambridge General English examination: ‘Interactive Communication’ scales across 5 levels (A2 Key, B1 Preliminary, B2 First, C1 Advanced, C2 Proficiency) – Cambridge Assessment English, <https://www.cambridgeenglish.org/exams-and-tests/qualifications/general/>
- Scale 14: The Examination for the Certificate of Proficiency in English (ECPE): ‘Discourse and interaction’ scale - Michigan Language Assessment, <https://michiganassessment.org/michigan-tests/ecpe-details/#speaking> ; <https://michiganassessment.org/wp-content/uploads/2020/02/20.02.ECPE-Speaking-Scale.pdf>
- Scale 15: Oral Proficiency Interview (OPI): ‘Global Tasks and Functions’ scale - American Council on the Teaching of Foreign Languages (ACTFL), <https://www.actfl.org/sites/default/files/assessments/OPI%20Familiarization%20Guide%202020.pdf>
- Scale 16: The Test of English for Academic Purposes (TEAP): ‘Interactional Effectiveness’ scale – Eiken Foundation of Japan, https://www.eiken.or.jp/teap/construct/sp_rating_crit.html
- Scale 17: The College English Test (CET): ‘Flexibility and appropriateness’ scales across 2 levels (CET 4, CET 6), <http://cet.neea.edu.cn/html1/folder/16113/1588-1.htm>; <http://cet.neea.edu.cn/res/Home/1704/55b02330ac17274664f06d9d3db8249d.pdf>
- Scale 18: Hong Kong Diploma of Secondary Education Examination- English Language School Based Assessment: ‘Communication Strategies’ and ‘Ideas & Organisation’ scales, <https://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/SBAhandbook-2019-ELANG.pdf>

IC-related scales from university in-house tests and validation research

- Scale 19: University X Spanish speaking test for beginners: Interaction EBB scale. (Ducasse 2010)
 - Scale 20: University X ELICOS Centre English for Academic Purposes Speaking test: ‘Interactional Effectiveness’ scale (May, 2011)
 - Scale 21: L2 Oral Pragmatic Abilities for Use in Social Contexts(Ikeda, 2017)
 - Scale 22: Rating scales for ‘interaction informed pragmatic performance’ in an EAP classroom-based assessment context (Youn, 2015)
 - Scale 23: The Kanda English Proficiency Test (KEPT): ‘Communicative skills/strategies’ – Kanda University of International Studies, Japan, https://kuis.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=1294&item_no=1&page_id=13&block_id=17
-