Developing tools for learning oriented assessment of interactional competence: Bridging theory and practice

Lyn May
Queensland University of Technology, Australia

Fumiyo Nakatsuhara, Daniel Lam
CRELLA, University of Bedfordshire, UK

Evelina Galaczi
Cambridge Assessment English, UK

Abstract

In this paper we report on a project in which we developed tools to support the classroom assessment of learners' interactional competence (IC) and provided learning oriented feedback in the context of preparation for a high-stakes face-to-face speaking test. Six trained examiners provided stimulated verbal reports (n=72) on 12 paired interactions, focusing on interactional features of candidates' performance. We thematically analyzed the verbal reports to inform a draft checklist and materials, which were then trialled by four experienced teachers. Informed by both data sources, the final product comprised (a) a detailed IC checklist with nine main categories and over 50 sub-categories, accompanying detailed description of each area and feedback to learners, which teachers can adapt to suit their teaching and testing contexts, and (b) a concise IC checklist with four categories and bite-sized feedback for real-time classroom assessment.

IC, a key aspect of face-to-face communication, is under-researched and under-explored in second/foreign language teaching, learning, and assessment contexts. This in-depth treatment of it, therefore, stands to contribute to learning contexts through raising teachers' and learners' awareness of micro-level features of the construct, and to assessment contexts through developing a more comprehensive understanding of the construct.

Keywords

Introduction

The assessment of interactional competence (IC) often plays an important role in speaking tests. Although IC has attracted considerable attention in speaking assessment research and practice (e.g., Galaczi & Taylor, 2018; Plough, Banerjee, & Iwashita, 2018), its theoretical conceptualisation and practical operationalisation have not been fully developed in terms of informing the teaching, learning, and assessing of interactional skills in a comprehensive and learner-friendly way. The aim of this research was therefore to develop an empirically-driven checklist and accompanying descriptions and recommendations that can be used to provide feedback on learners' interactional skills. Such an endeavour is especially important because it taps into two strong learner needs: the need to develop the complex and cognitively demanding skill of participating in interaction and the need to provide learners with micro-level feedback which can inform further learning (Carless, 2007).

Interactional competence

Following on from theoretical debates on IC (e.g., Kramsch, 1986; Young, 2011), there is now a solid body of research in L2 assessment which has provided useful insights about the co-construction of interaction between test-takers, paving the way for a comprehensive definition of the IC construct (e.g., Ducasse & Brown, 2009; Lam, 2018; Galaczi, 2008, 2014; May, 2011; Nakatsuhara, 2013, Plough et al., 2018; Roever & Kasper, 2018). These studies have suggested that IC is the ability to co-construct interaction in a purposeful and meaningful way, taking into account socio-cultural and pragmatic dimensions of the speech situation and event. This ability is supported by the linguistic and other resources which speakers and listeners leverage at a micro-level of the interaction, i.e., aspects of: *topic management*, *turn management*, *interactive listening*, *breakdown repair* and *non-verbal behaviors* (Galaczi & Taylor, 2018). Similarly, Ikeda (2017) synthesized findings from a range of studies identifying aspects of IC associated with higher proficiency L2 learners, and suggested that IC includes being more actively involved in the conversation, providing support to the interlocutor to develop the topic of the conversation, being less dependent on the interlocutor's support, more able to develop other-initiated topics, and more able to recognize a broader range of functions of a discourse marker and use it appropriately. These studies have provided useful insights about the development of IC in learners.

In the wider field of applied linguistics, Hall and Pekarek Doehler (2011) defined IC as "the ability to accomplish meaningful social actions, to respond to co-participants' previous actions and to make recognizable for others what our actions are and how they relate to their own actions" (p. 1). Pekarek Doehler and Berger (2018) further extended this and argued that IC development is manifested in diversification of methods for interaction, improvement of recipient design and context sensitivity, and a higher capacity in monitoring co-participants' talk and using grammatical resources for different interactional purposes.

Over the past few decades, the construct of IC has also been operationalized in interactive speaking test rating scales, with examples found in the *communicative effectiveness* and *interactive listening* scales in the ISE Speaking & Listening test by Trinity College London (2015, p. 44), and the *interactive communication* scale in Cambridge English Speaking tests (2016, p. 85). Similarly, international language benchmarks such as the Common European Framework of Reference for Languages (Council of Europe, 2018) include descriptors covering the construct of IC (i.e., *turn-taking, co-operating, asking for clarification, online conversation and discussion*). All of these scales provide useful information for assessment purposes, but stop short of more in-depth descriptions which can be used for diagnostic purposes and the provision of relevant feedback for learners.

Analysis of candidate talk informed the development of rating criteria for interactional and pragmatic competence in studies by Youn (2015) and Ikeda (2017). Youn's rating criteria

were developed in the context of EAP classroom assessment of L2 pragmatics through roleplays. The five criteria that emerged were: *content delivery*, *language use*, *sensitivity to situation*, *engaging with interaction*, and *turn organization*. Ikeda's study on the assessment of oral pragmatics in the context of university activities involved the use of both monologues and dialogues. Four rating criteria—*social actions to achieve the communicative goal*, *facility with the language*, *language use to deliver the intended message*, *language use for mitigation*—were used for both tasks, while *engagement in interaction* and *turn organization* were additionally used for the dialogic tasks. Lam's (2018) analysis of candidate discourse in a group speaking assessment unpacked the rating descriptors related to *responding* through identifying three specific ways of doing so: formulating, extending, and accounting for (dis)agreement with co-participants' ideas, which has implications for rater training and developing rating descriptors for *interactive listening*.

The theoretical debates on the conceptualization of IC, the empirical work on co-construction of interaction in L2 tests and in classrooms, as well as IC rating scales explicitly defining the IC construct have provided a solid body of knowledge to inform the comprehensive definition of IC in the context of assessment and learning and have provided key areas for meaningful feedback to learners.

Note that the conceptualization of IC in this study and the corresponding feedback offered to learners focus mainly on managing the mechanics of interaction (see Figure 2). We acknowledge that interaction as the exchange of meanings embodies a host of prior knowledge and experience, cultural references, and understandings of social norms that each participant brings with them (cf. Bourdieu's (1991) habitus), which shapes the exchange in complex and nuanced ways. However, our scope was necessarily limited to one view of IC that is immediately relevant to learners within the context of a paired speaking test. This way, the feedback around these aspects would be contextually appropriate, relating and feeding back to the learners' immediate experience of the task.

Learning oriented assessment and IC
The term Learning Oriented Assessment (LOA) has been widely used to "encapsulate the idea that all forms of assessment should promote learning" (Green, 2016, p. 23). In order for assessment to engineer lasting learning, Carless (2007) identified three LOA principles: (i) assessment tasks to stimulate sound learning practices, (ii) learners' active engagement with assessment activities, and (iii) appropriate and timely feedback "which learners can ultimately 'feedforward'" (Carless, 2007, p. 59). LOA moves beyond traditional binary positioning of the formative and summative purposes of assessment to a systemic and dynamic relationship between teaching, learning and assessment (Jones & Saville, 2016; Turner & Purpura, 2016).

Recent studies highlight the potential for standardized speaking tests to provide not only a summative judgement of performance, but also feedback to candidates, in order to connect large-scale examinations with meaningful learning opportunities in language classrooms where test preparation is undertaken (e.g., Green, 2016; Jones & Saville, 2016). The key issue is how assessment information can be used to promote learning in the context of high-stakes examinations, and generate personalized feedback which learners generally find most useful (Dunlop, 2017).

Personalized feedback has also been advocated as key to *deliberate practice*. In the area of education research, Ericsson (1996) contended that in order to improve, a learner needs to spend time focusing specifically on those activities which he/she cannot yet do well. From a cognitive perspective, DeKeyser (2007) argued along similar lines. Deliberate practice is individual-specific and can be challenging to implement at class level. Tools that language teachers can readily use to identify individual learners' developmental needs can therefore be

useful in classroom settings and provide granular diagnostic information, which is essential for determining future learning objectives.

As noted, IC has attracted considerable attention in the field of speaking assessment research over the last few decades (Plough et al., 2018). However, research findings have not been fully applied to teaching and learning contexts in a comprehensive and learner-friendly way. A review of several major English textbooks and course books indicates that interactive skills receive relatively little coverage in comparison to other speaking sub-skills (Nakatsuhara, May, Lam, & Galaczi, 2018).

Drawing on empirical evidence in the development of LOA tools is critical. This includes examiners' and language teachers' understanding of learners' interactional performance (e.g., Ducasse & Brown, 2009; May, 2011). In May's (2011) study, teacher-raters were able to describe micro-level interactional features of paired performance accurately. Nakatsuhara (2013) extended this line of argument and suggested that in addition to test scores raters could give a descriptive report on candidates' performance, commenting on key aspects of each candidate's IC and giving suggestions for improvement. Acknowledging that this may not be viable for high-stakes contexts, Nakatsuhara argued for providing diagnostic feedback in low-stakes classroom assessment contexts. Leaper (2014) further suggested a drop-down menu with key interactional aspects as a feasible feedback method in a testing context.

Research questions

Drawing on theoretical and empirical findings, and the limited availability of comprehensive and practical LOA tools for interactional skills, we aimed to develop a checklist and accompanying descriptions and feedback for learners and teachers. The specific test context chosen for this study is the *Cambridge B2 First* (formerly known as *First Certificate in English*) exam.

Two research questions were addressed.

RQ1: Which features of IC are salient to *B2 First* speaking examiners, who are also experienced teachers, when they award scores for the *B2 First* paired speaking task?

RQ2: How can these features inform the development of an IC checklist that can facilitate meaningful learning oriented feedback to learners?

Research methodology

The project reported here consisted of two stages:

- **Stage 1:** Collecting and analysing data to inform a draft checklist and accompanying materials.
- **Stage 2:** Piloting and refining the draft checklist and accompanying materials.

Stage 1: Participants

Six experienced *B2 First* examiners (Examiner ID: E1–E6) who were also senior examiners responsible for examiner training in different world regions participated in Stage 1 of the study. The average length of ESL/EFL teaching experience was 34.6 years (SD=3.5), and *B2 First* examining experience was on average 31.2 years (SD=6.3). Some of them had contributed to the development of the current IC scales, and most were involved in the production of examiner standardization/certification materials.

Stage 1: Data collection

Materials. Out of a dataset provided by Cambridge English and from publicly available videos of the paired task in the *B2 First* speaking examination, we selected twelve videos as stimuli for examiners' verbal reports. To ensure that the videos cover a range of candidate proficiency levels, we selected balanced numbers of high-IC scoring (Band 4 or 5 on the *B2 First* scale),

mid-IC scoring (Band 3), and low-IC scoring (Bands 1 or 2) candidates. We also reviewed the videos and evaluated the interactional pattern of each discussion based on Galaczi's (2008) classification (i.e., collaborative, asymmetric, and parallel interaction), and ensured the selection covers different interactional patterns. The different patterns reflect variations in turn-taking, topic management, and engagement with each other's talk (ibid.). Finally, we also chose candidates with balanced L1 and gender profiles to avoid potential biases due to test-taker characteristics (see Nakatsuhara et al., 2018, p. 12 for a tabulation of candidate profiles).

In the *B2 First* paired task, candidates are provided with a prompt card with several visual items. They are first required to discuss each visual item in relation to the given topic, and then asked to make a decision (an example of the test, including this task, is available at www.youtube.com/watch?v=-tqeI9t4x9E).

Examiners' verbal comments. Following May (2011), the data collection of verbal reports involved two viewings of videos, employing stimulated recall methodology (Gass & Mackey, 2000). The examiners first received a tutorial for the stimulated recall procedures using written instructions and a practice video. Then, working with the 12 video clips, they individually digitally recorded their verbal comments on each paired discussion performance. They followed three steps:

Step 1: Examiners viewed the performance once without stopping the video. In order to confirm the level of the IC performances selected for this study, they reassessed the paired candidates using the *B2 First* IC rating scale. They recorded an IC score for each candidate, together with a brief justification for the score.

Step 2: Examiners viewed the same performance again, pausing the video at any point to comment on any IC features that influenced their evaluation.

Step 3: After the second viewing, examiners provided recommendations for each candidate to enhance his/her IC performance, drawing upon their teaching and examining expertise.

The order of viewing the 12 videos was counter-balanced to minimize a possible order effect. In total, 72 sets of audio-recorded examiner comments were gathered.

Stage 1: Data analysis
Data processing. The 12 paired interaction performances were transcribed using a simplified version of Conversation Analysis (CA) notation (Atkinson & Heritage, 1984) to enable the researchers to understand examiner comments better by linking each comment with relevant candidate utterances. The 72 audio-recordings of examiner comments were orthographically transcribed and divided into three sub-sections: the summary statement, stimulated verbal recall, and recommendations for candidates. In order to match each examiner comment to the section of test-taker discussion to which the comment referred, we inserted examiners' specific comments from their second viewing into the corresponding locations of the CA transcripts.

Thematic analysis. We thematically analysed the annotated examiner comments using NVivo 11. Following Yin's (2011) recommendation, we used both deductive and inductive approaches to develop the coding scheme. IC themes from the *B2 First* rating scales informed an initial provisional list of codes. The coding was then carried out both deductively, using/modifying these provisional codes, and inductively, with additional codes developed in response to emergent themes in the data. We took care when segmenting the transcripts so that "each segment should be representative of a single, specific process" (Green, 1998, p. 75) to ensure reliable coding and quantification of codes under each theme. This meant that each turn in the recall and the recommendations could generate several segments e.g., *C01F introduces*

a new topic on shops and she develops it adequately was segmented as: *C01F introduces a new topic on shops/ and she develops it adequately/.* The first segment here relates to introducing a topic, whereas the second segment relates to the extent to which the candidate could develop this topic.

From the summary statements and stimulated verbal recalls, eight macro themes and 27 micro themes emerged. We further classified comments within the 27 micro themes into positive and negative categories (Neutral comments were not excluded from the analysis, but were, in fact, rare and when they occurred, they were framed as part of a clearly positive or negative comment.) From the examiner recommendations for improvement, we identified 17 themes.

We double-coded 25% of all the data (examiners' summary statements, stimulated verbal recalls, and recommendations), and inter-coder agreement rates were high, with 95.6% for macro themes (366 of the 383 coding points), 89.0% for micro themes (275 of the 309 coding points) and 100% for examiner recommendations (108 of the 108 coding points). We discussed discrepancies and reached a consensus on all coding decisions. We then revised the coding scheme to reflect these discussions and the remaining data were coded accordingly.

It should be noted that finalising the boundary of each code was not always straightforward, since some micro themes could potentially contribute to more than one macro theme (e.g., interactive listening and body language). While acknowledging that some macro themes are not mutually exclusive, we identified the primary code for each segment, flagged all ambiguous cases, discussed as they occurred throughout the coding process (see Nakatsuhara et al., 2018 for more information) and assigned to one macro category. This process was undertaken to support the objectivity of the coding scheme.

Stage 2: Participants
Stage 2 was conducted in a single focus group format, involving four language teachers (Teacher ID: T1–T4) who had taught *B2 First* preparation and general English classes. They had an average of 22.5 years of experience (SD=12.3) in teaching, examining, and developing testing/teaching materials for interactional skills. Following Krueger and Casey's (2015) recommendation, a group of four was selected to be an appropriate group size to enable the participants to carry out rather complex tasks, discuss their experience in detail and then share in-depth insights in response to a number of semi-structured questions. For the same reason, the focus group employed a *single-category design* (Krueger & Casey, 2015), where all participants were relatively homogeneous – i.e., experienced teachers with great insights into teaching and testing interactional skills. In addition to long teaching careers, they also had varying backgrounds, which we hoped would help offer different perspectives in the focus group. For example, T1 had extensive experience in developing teaching materials for interactional skills; T2 had over 30 years of experience in preparing students for *B2 First*; T3 worked for a large international examination board, developing and validating various tests of interactional skills; and T4 was an item writer for several standardised speaking examinations. The limited resources of the study did not allow us to undertake more than one focus group. However, we considered that that a single focus group would be sufficient as an initial trial of the developed checklist at this stage (see also *Limitations and future research*).

Two of the researchers, who have a PhD in language testing and specialise in speaking assessment research and who were also heavily involved in Stage 1 of the study, were present in the focus group session, in order to moderate and take detailed notes of the discussion.

Stage 2: Data collection
In Stage 2, we piloted the draft checklist and accompanying materials developed in Stage 1 with four teachers during a focus group discussion in order to gain potential users' perspectives

on the applicability and usefulness of the developed materials and to inform further refinements. The focus group involved both evaluation tasks and discussion, and it took over four hours in total.

The four teachers watched six *B2 First* paired task video performances (two high-scoring, two mid-scoring, two low-scoring) one by one, and identified salient performance features by highlighting and checking off the relevant checklist items. After each video, they commented on the ease or difficulty of assessing the performances using the checklist, how relevant the descriptions and feedback were to the categories, and whether the feedback was comprehensible, encouraging and helpful.

The moderators had 28 scripted questions regarding the applicability and usefulness of the checklist and other materials, specific wording used in them and how best these materials could be formatted and presented. Since some of the questions had already been covered during the discussion of individual video performances, the latter part of the session focused only on the questions that had not been covered earlier as well as follow-up questions which emerged during the focus group (for the list of questions explored during the focus group, see Nakatsuhara et al., 2018, p. 42-43).

The teachers provided various suggestions to make the materials more teacher- and learner-friendly. The focus group was audio-recorded, and the two participating researchers took detailed notes during the focus group discussion. Aspects of Stage 2 will be further explained in the section 'Results from Stage 2: Piloting and refining the materials.'

Stage 2: Data analysis
We examined the detailed notes taken during the focus group, and checked the notes against the audio-recording of the session for accuracy. The key suggestions made by the four teachers were then tabulated. We discussed in depth each of the suggestions, agreed on the response to each issue, and noted the rationale for each decision.

Results from Stage 1: Developing the checklist

As examiners had been trained to apply certain criteria, it was unsurprising that they focused on these criterion aspects of the performance in their summary statements and stimulated verbal recalls. These aspects included a candidate's ability to initiate discussions, introduce new ideas, respond appropriately, link contributions to those of other speakers, maintain the interaction, develop the interaction, negotiate towards an outcome, and the extent to which support was needed. The examiners also noted a range of non-criterion (i.e., not included in rating criteria) aspects of the performance, including body language, interactive listening, assertiveness, and whether the interaction appeared to be genuine. Table 1 provides a summary of the themes.

Table 1. Summary of macro and micro themes.

Criterion features	
Macro themes	Micro themes
Initiates discussions, introduces new ideas (204 comments; all raters)	• Starts discussions • Introduces/contributes new ideas/topics, shifts topics
Responds to partner (236 comments, all raters)	• Responds appropriately to partner, including politeness conventions, ability to agree and/or disagree • Links contributions to those of partner; expands, extends partner's ideas
Maintains and develops the interaction (382 comments, all raters)	• Keeps the interaction going, including asking for opinion, clarification, repetition, elaboration, explanation • Contributes adequately to the discussion • Engages with the interaction through relevant, substantive contributions • Extends own idea by explaining, elaborating and/or justifying • Extends the scope of the interaction beyond task • Uses functional/formulaic language, including linking words, cohesive devices • Engages in turn-taking: including invites partner to take a turn, initiates turns, holds on to turns, takes the floor from partner when necessary, uses intonation to facilitate turn-taking
Negotiates toward an outcome (61 comments, all raters)	• Summarizes own and/or partner contributions to facilitate outcome • Explicitly negotiates toward an outcome
Support (42 comments, all raters)	• Needs support from either partner or interlocutor/examiner • Provides support for partner
Non-criterion features	
Macro themes	Micro themes
Interactive listening (40 comments, raters 1,2,3,4, 6)	• Backchannels to indicate comprehension and/or interest • Noted to be concentrating on what partner says
Body language (78 comments, all raters)	• Makes eye contact • Nods to indicate comprehension and/or agreement, signals end of turn • Smiles and/or laughs to indicate agreement, interest and/or rapport • Points to task prompt/visuals • Shows involvement through positioning of body, leans towards task visual and/or partner
Manner of interaction (65 comments, all raters)	• Interacts with ease and confidence • Extent to which interaction seems natural, genuine, authentic

Examiners also reflected on non-IC features, such as the proficiency level of the candidates in terms of their ability to participate in the task (20 comments), the impact of the pairing (six comments) and the impact of viewing the performance for a second time (three comments). For a detailed account and examples of each macro and micro theme, see Nakatsuhara et al. (2018).

In addition to the themes from the verbal recalls, examiners provided recommendations for the candidates regarding improving their IC skills. Their recommendations were wide-ranging, and included the need to make eye contact and use a range of effective body language, engage in active listening, engage substantively with topics, develop the confidence to introduce new topics, ask the partner questions to show the ability to initiate, learn how to hold the floor, move away from overly formulaic expressions, and ensure that there is negotiation towards an outcome (see Nakatsuhara et al., 2018 for full details).

In the interest of space and succinctness, we selected findings from two macro themes to illustrate the examiners' comments on key aspects of IC and their recommendations for candidates: the criterion theme of *maintains and develops the interaction* and the non-criterion theme of *interactive listening*. The former criterion theme emerged as a key focus for

examiners, eliciting 382 comments, while the latter non-criterion theme reflects findings from other IC studies (e.g., Ducasse & Brown, 2009; May, 2011, Ross, 2018).

Maintains and develops the interaction
Although maintaining and developing the interaction are placed together under one theme (reflecting their inclusion under one criterion in the *B2 First* rating scale), examiners at times clearly differentiated them and would comment that a candidate could do one, but not the other. As reflected in the examiners' comments, *maintaining the interaction* is mainly about turn-taking management and sharing vs. dominating the discussion, whereas *developing the interaction* concerns topic development, including sustaining and extending a topic over several turns.

Table 2 summarizes positive and negative aspects noted by all six examiners when commenting on the extent to which candidates could maintain and develop the interaction.

Table 2. Maintains and develops the interaction: features noted.

Theme	Positive	Negative
Maintains and develops the interaction	Seen to be active and effective in both maintaining and developing the interaction	Requires support/relies on partner to maintain and/or develop the interaction
	Actively invites partner in by asking for opinion	Takes an overly passive role
	Effectively turn-takes, shares the floor	Engages in extended turns that are monologic, rather than dialogic and/or dominates the interaction; for example, through interrupting and making it difficult for partner to fully participate
	Carefully considers all points and substantively engages with the topics and task	Describes rather than gives opinion/analyzes/discusses/persuades Responds in ways that are not relevant to partner's point and/or task
	Extends own idea by explaining, elaborating, justifying and/or providing examples	Provides brief/minimal responses that do not effectively develop the interaction and/or does not provide reasons for a response
	Brings the interaction back on track when necessary	Responds with points that are illogical and/or contradictory
	Uses a range of functional language appropriate to interaction	Uses a limited range of appropriate functional language and/or overuses formulaic expressions

Selected illustrative comments are provided below, indicating the examiner (e.g., E1), the candidate (e.g., *C01F*) and the performance (e.g., P1) on which the examiner is commenting.

One essential aspect of maintaining the interaction concerns the extent to which candidates invited their partners into the conversation, which was often done through questions. Examiners evaluated speech positively when candidates asked their partner for his or her opinion or asked for explanation or elaboration. This gives the partner an opportunity to develop the topic further, showed the candidates' interest in their partner's ideas, and provided evidence of turn-taking skills.

> *C01F asks C01M to expand on that and to explain why she thinks that parks are a good idea giving her a chance to develop the topic further.* E1-P1

> *C04D asks C04S questions and endeavours to bring her into the conversation.* E6-P4

In contrast, it does little to maintain the interaction if a candidate offers only their own ideas and does not involve the partner in discussing these. If both candidates were to do this,

the resulting discussion would typically orient to a parallel interactional pattern (Galaczi, 2008), with each candidate introducing their own ideas without substantively engaging with each other. Examiners also evaluate it negatively if a candidate dominates the discussion or interrupts their partner, or, at the other extreme, if a candidate is overly passive and does not speak even when invited to or after a long silence.

> *C07G doesn't say anything, but maybe she hasn't got the opportunity to because C07A's over-extending his turn a little bit and speaking at length here as he did with the first topic as well.* E4-P7

> *And C10O doesn't keep the interaction going, not even with C10J's support.* E1-P10

The manner and effectiveness of turn-taking was noted by examiners, including the ways in which candidates maintained and passed interactional turns.

> *C06N does not give up her turn easily. She lengthens vowels and words ("we," "see") and also uses a lot of rising intonations ("dancing" and "cafeteria" and "flowers") to show that she is holding onto her turn.* E1-P6

> *Once again they toss the ball to and fro between them and there is an agreement. They each put in some new ideas and C04D ends the interchange with a very nice "yes, exactly."* E6-P4

As for developing the interaction, examiners noted whether an idea was developed across turns. Ways of developing the interaction include adding information, offering examples, or elaborating on the idea with additional points. In contrast, the interaction is not effectively developed if the candidate states a choice or opinion without giving reasons, or only gives minimal or brief responses (e.g., "*mm hmm,*" "*yeah, I agree*") when responding to the partner.

> *Once again C04S says "yeah, I think so" in agreement and then she explains a little bit why. She doesn't just stop at "I think so."* E2-P4

> *C08M doesn't really develop the point, just introduces it and passes over to his partner.* E4-P8

Other key aspects that influenced examiner decisions were the quality and relevance of the contributions, including whether candidates consider each idea/option carefully and from different perspectives and the relevance of their contributions. Examiners commended candidates who show a good understanding of what the task requires and could steer the discussion in the right direction, but negatively remarked on instances when candidates mainly described details in a picture without relating them to the discussion point.

> *C02H brings the interaction on track by saying what she believes is useful.* E1-P2

> *They don't really focus on the task. What they are talking about is just what's in the picture. They are not talking about the aspect of helping other people.* E4-P5

Examiners also positively noted the appropriate use of functional language, while overuse or inappropriate use was regarded negatively and seen as 'artificial.'

> *So C10J actually says "let's start with this one," "what do you think about this shoes and gloves?" so I think she may have been trained and done a course and actually understands what the task requires.* E3-P10

> *C07A's throwing in a phrase that he's learnt "I totally agree with you," but a little bit inappropriate there, it's very artificial.* E2-P7

Following on these observations, examiners then provided recommendations for candidates to improve their skills in maintaining and developing the interaction. Examples of recommendations included:

> *[They need to] move on properly to the different topics and relate it back to the task all the time….so keep the overarching topic in focus all the time.* E4-P8

> *I would tell her to use pause fillers… to show she's not giving up her turn, so that she is not interrupted by her partner.* E1-P5

> *Both of them need to find a way to be able to pass on the turn. You can do that through intonation. You can do that by looking at the person. But they don't make any use of that. Pushing the paper over is not a very successful means of saying "it's your turn."* E2-P11

> *I would tell them to develop each topic further, to exhaust the topic before passing on to the next one.* E1-P2

> *I would say that both of them need to add more detail to their ideas, to make their contributions more extensive and justify their ideas.* E4-P8

Interactive listening
Table 3 summarizes positive and negative aspects of interactive listening noted by five of the six examiners.

Table 3. Interactive listening: features noted.

Feature	Positive	Negative
Interactive listening	Demonstrates active listening through being noted to listen carefully, including nodding	Seems to be more focused on what they want to contribute rather than listening to their partner's views
	Uses back-channelling to indicate comprehension, agreement and/or interest	Does not use back-channelling
	Gives partner time to frame contributions	Does not give partner space to talk

Examiners viewed speech positively when candidates displayed interest in what the partner had to say, and thus showed support for the partner to continue talking. Listener support strategies noted by examiners included back-channelling, nodding, smiling, and looking at the partner. Interactive listening and body language were thus closely connected, as Examiner 1 clearly expressed: "they use body language to show they are listening."

> *I think C04D was a very sympathetic listener and a very supportive listener.* E4-P4

> *C08G has been listening attentively to him, has been making eye contact, and also nodding to show that he is in fact listening.* E1-P8

Examiners commented positively on candidates demonstrating interactive listening by developing the partner's ideas in the next turn and by asking follow-up questions, which provided evidence that the candidate listened attentively to the partner's talk. Examiners also valued the listener giving the partner time to formulate ideas, rather than interrupting or finishing the partner's talk prematurely.

> *And C01F comes straight back with "why do you think that, actually?" This is really good interactive communication. It shows she's been listening.* E6-P1

C04S shows that she's actively listening to C04D by nodding and by providing back channelling "yeah, yeah," and commenting briefly on what C04D has said. E1-P4

C08M's sensitive to his partner. He gives his partner space. He listens to his partner. That's good interactive communication in my book. E6-P8

In contrast, examiners commented negatively on candidates whose interaction was mainly self-focused or candidates who did not appear to be listening attentively.

Each one is introducing his or her topic, different topics, but not really listening to their partner's. They are probably listening for the opportunity to grab the turn and say something themselves. E1-P9

Examiners then provided recommendations for candidates to help improve their interactive listening skills. Their suggestions included:

Both of them need to learn to be better active listeners, not just look at the other person…but to nod, to say "yes, yes." E2-P2

He needs to listen a bit more to what his partner is saying, instead of thinking about what he is going to say next. If he were more engaged with his partner, if he listened to what she was saying, he might actually be able to add to that in terms of accepting or rejecting the choice and giving some more extended reasons for doing that. E3-P10

Developing a draft checklist

The examiners' comments on the IC features that they attended to while rating the 12 video performances, together with their recommendations for learners, informed the development of LOA materials for both teachers and learners, consisting of a draft checklist, accompanied by descriptions of the features and feedback for learners. We drafted the first version of the checklist by listing the salient features within each of the eight macro themes. Particular attention was paid to establishing the corresponding positive and negative performance features in the examiners' comments. In so doing, users of the checklist can have a clear idea of 'Dos' and 'Don'ts' for each aspect of interaction. For example:

Start the discussion
+ve: *Do take the initiative or show willingness to start.*
-ve: *Don't wait passively for the partner to start.*

The accompanying description, with teachers as the targeted readership, elaborated on the positive and negative features in each macro theme. We provided examples of functional language (e.g., "*Shall I start?*" "*Perhaps you're right, but…*"), drawing from examiners' comments or retrieved from the paired interaction transcripts corresponding to the relevant examiners' comments.

The feedback section of the checklist drew on the NVivo analysis of recommendations given by examiners, simplified and reworded to tailor to learners as the intended readership. The *Well done!* sub-section gives due credit for the positive features that candidates have displayed, and encourages them to keep incorporating those features in their interactions. It also outlines ways in which they can further enhance their performance. The *Needs more work* sub-section helps learners identify problematic aspects of performance and outlines ways to improve, together with some useful phrases they could draw on. We drafted both sections in keeping with the LOA principles of raising learners' awareness of success criteria, and giving meaningful feedback that helps learners identify what they can and can't do, and how to bridge that gap (Jones & Saville, 2016). This will be further elaborated in the Discussion and Conclusion section. Figure 1 shows an extract of the checklist and accompanying descriptions

and feedback (Note: The way in which ticks are applied—i.e., making a broad judgement for a macro category or evaluating every micro feature—should vary according to the purpose of the checklist application as discussed in Stage 2 below).

Does the learner……? (Please tick (✓) as appropriate. You do not need to tick all boxes if it is difficult to make binary judgments based on observed performances. L indicates a learner on the left and R indicates a learner on the right.)

1.1 Start the discussion?

Positive		Negative		Description	Feedback
1.1a) Negotiate who/how to start in a collaborative manner, involving the partner	L ☐ R ☐	1.1d) Start the discussion with a monologue without involving the partner; or in a way difficult for the partner to contribute	L ☐ R ☐	• A good way for the student to begin the discussion is to involve the partner in deciding who or how to start (e.g. by saying 'Shall I start?' or 'Where should we start?'). It is less desirable to start by delivering a long speech, assigning who to start without negotiation, or starting in a way that makes it difficult for the partner to respond or contribute (e.g. asking a generic question 'what do you think?' or a question that simply repeats the task instructions).	**Well done!** →Great to show that you are ready to start the discussion! Well done also on deciding together with your partner who to start or how to start the discussion. You have also shown that you understand what the task is about and where the discussion should be going.
1.1b) **Take** the **initiative** or show willingness to start		1.1e) Wait **passively** for the partner to start		• It is also good if the student takes the initiative or shows the willingness to start, rather than be hesitant or wait passively for the partner to start speaking.	**Needs more work** → Showing that you are ready to start and also happy to work with your partner are both important. Next time, you may want to:
1.1c) Start in a way that is clearly **relevant** to the **task**		1.1f) Start in a way that seems to **ignore** the task **instructions**		• Moreover, it is important to show a clear understanding of what the task is about (e.g. 'So, we need to decide which activities would make life in a city more enjoyable.') and where the discussion is going. It is less desirable for the student to start by simply describing a picture.	– Decide together with your partner who to start (e.g. 'Shall I start?'/'Would you like to start?') and how to start the discussion (e.g. 'Where should we start?') – Avoid either being too hesitant to start, or starting the conversation on your own delivering a long speech, as your partner may find it difficult to respond to it. – Show that you understand what the task is about (e.g. 'So, we need to talk about/decide…'). Try not to start simply by describing a picture.

Figure 1. Extract of IC Checklist – Full version (Nakatsuhara et al., 2018, p.62)

Results from Stage 2: Piloting and refining the materials

The focus group discussion in Stage 2 provided insights on a number of issues, ranging from the organisation of the IC categories to the layout and language of the checklist and accompanying materials, as well as suggestions for use in classroom teaching and assessment. The key comments/suggestions from the focus group can be summarized as below:

1. *Length:* A long checklist puts high demand on teachers' processing capacity.
2. *Layout:* For each checklist statement, describe the positive feature, with the options Yes/ No / Maybe.
3. *Organisation of IC categories:* Functional language is relevant to different categories rather than only 'Maintain and develop interaction.'
4. *Feedback:* More examples of fixed phrases or formulaic language can be added to the feedback.
5. *Teaching:* Breaking down the checklist for teaching purposes would be very useful.

The following excerpts illustrate comments/suggestions 1, 2, and 5:

Comment/suggestion 1: *It's an exhaustive list with all these micro points* (T1). *This seems to me more like a short book rather than something like teachers' notes… I think when you simplify it, you should maybe do different alternatives* (T2). *It's also the processing capacity, so, how much my brain can process under pressure* (T3).

Comment/suggestion 2: *My brain can really only do the left side [+ve] for two people. I can't do two sides [+ve/-ve] for two people* (T3). *If I was doing this, I would have one thing and it would say "yes?" "no?" so there would be only one description. And that's the only way I could actually handle it* (T2).

Comment/suggestion 5: *It'll be interesting to know how you think this is gonna be used, because I would've recommended the teacher to just do with the class number 1, and then next week, do 1 and 2, and then build it up. Because for a teacher to pick this up and try and cope with this, it's actually quite a lot to deal with* (T2).

The group also discussed whether it is possible to evaluate each micro feature or whether it is more realistic to ask users to make an overall judgement on macro categories only. A consensus was reached that it should depend on the purpose of the checklist application; that is, if a teacher intends to focus only on a few macro categories to assess and teach in a lesson, it makes sense to evaluate every micro feature under that category.

Taking on board these insights, the research team produced a condensed version of the checklist and feedback as an example for teachers to use with learners. The concise version (see Figure 2) consisted of four main categories and included features (a–m) extracted from all categories in the original, detailed checklist. We selected these components based on the number of examiner comments received in Stage 1, and the teacher feedback from Stage 2. We adopted the "one feature, one statement" suggestion to improve ease of processing the checklist in real time (see Nakatsuhara et al., 2018 for more information on the selection process). It should be noted that this concise version is not meant to replace the full version (Figure 1). Instead, the concise version is an example of how the detailed checklist can be adapted to suit individual learning and testing settings, and how it can dynamically evolve rather than staying as a static product.

Please tick (✓) as appropriate. You do not need to tick all boxes if it is difficult to make binary judgments based on observed performances.

The student can…	Interaction strategies	Learner 1			Learner 2		
		Well done!	Needs more work	Comments	Well done!	Needs more work	Comments
1. Initiate new ideas	a) *New ideas:* Take initiative to contribute relevant new ideas						
	b) *Right time for new ideas:* Contribute new ideas after the current idea has been adequately discussed						
	c) *Language:* Use a range of appropriate language to initiate new ideas and/or shift from one idea to another						
2. Keep the discussion going over several turns	d) *Develop (own idea):* Extend your own ideas sufficiently						
	e) *Develop (partner's idea):* Extend the partner's ideas by linking their own contribution to the partner's and giving more than just a token response						
	f) *Invite:* Actively invite your partner if needed (e.g. asking questions, helping complete a sentence where necessary, prompting partner to say more)						
	g) *Listen:* Show listener engagement through back-channelling and short responses (e.g. 'exactly', 'right', 'OK')						
	h) *Be collaborative:* Keep a natural and collaborative flow to the interaction (e.g. no long pauses within/between turns, no dominating interruptions)						
	i) *Language:* Use a range of appropriate language (e.g. agreeing, disagreeing, explaining, elaborating, justifying, providing examples)						
3. Negotiates towards an outcome	j) *Joint decision:* Proactively work towards making a joint decision (e.g. inviting the partner to make a choice, showing willingness to compromise)						
	k) *Language:* Use a range of appropriate language (e.g. summarising, evaluating, comparing, prioritising points raised in the discussion)						
4. Use body language appropriately	l) *Body language:* Use appropriate body language (e.g. nodding, smiling) to show interest in your partner's contributions and/or signal change of speakers						
	m) *Eye contact:* Keep eye contact with partner						

Figure 2. IC checklist- concise version (Nakatsuhara et al., 2018, p.67)

Another significant change concerned the category of "using functional language." This was originally one of the micro features subsumed under "maintains and develops the interaction" (see Tables 1 and 2). However, the focus group participants pointed out that using appropriate functional language is relevant to a range of different aspects of interaction (e.g., initiating, responding, developing). Therefore, we made it a macro category in the full checklist, and included it as a micro-feature (*c, i, k*) under each of categories 1-3 in the concise version, to help teachers and learners to focus on it as a key aspect of IC.

Discussion and conclusion

In this study we aimed to bridge theory and practice in the learning oriented assessment of IC through developing an empirically-driven tool for providing meaningful feedback to learners on their interactional skills in the context of a discussion task. This undertaking involved 1) bringing together complementary perspectives: the theoretical and research literature of IC,

examiners' comments and recommendations to candidates, and teachers' views on using the tool; 2) distilling the insights from various data sources; and finally, 3) crafting a tool supporting learning. In so doing, we aimed to address the gap that IC skills are underrepresented in existing textbooks, and that feedback to learners on IC skills had received less attention in research than formal aspects of language such as vocabulary and grammar.

RQ1: Identifying salient IC features
A key aim of this study was to identify IC features relevant to performances on the *B2 First* paired discussion task, enabling us to develop the checklist and feedback materials. Examiners attended to both criterion and non-criterion IC features, as evidenced by the mentioning of these features in their comments and recommendations to candidates. The macro categories, identified through thematic analysis of examiner comments that were further modified after the focus group, capture the extent to which a learner can:

- Start the discussion and contribute new ideas
- Respond to partner
- Maintain and develop the interaction
- Negotiate towards a common decision
- Provide or need Support
- Demonstrate interactive listening
- Use body language
- Use effective functional language for interaction
- Interact confidently and naturally (For the full checklist, see Nakatsuhara et al., 2018.)

The nine categories were generally in line with IC features identified in the L2 IC assessment and development literature (e.g., Galaczi & Taylor, 2018; Young, 2011). For example, the two macro themes discussed and illustrated above (*maintaining and developing the interaction* and *interactive listening*) echo the findings of Galaczi (2014) on topic management, and keeping the discussion relevant to the task resonates with the notion of recognizing and being sensitive to the ongoing context (Ishida, 2011; Youn, 2015). As for the theme of *interactive listening*, back-channelling as an interactional feature displaying listener support and engagement has been identified in a number of previous studies (Galaczi, 2014; Ross, 2018; Youn, 2015); whereas micro-features such as giving the partner time to formulate ideas versus interrupting or prematurely finishing a partner's talk relate to a key dimension of IC development: increasing capacity to monitor the co-participants' prior talk (Pekarek Doehler & Berger, 2018). We must acknowledge that some salient features that characterize L2 IC development such as diversification of methods for actions (*ibid.*) and more preliminary work in making requests (Roever & Kasper, 2018) were not directly captured in the current study, possibly due to task type and proficiency range (see the *Limitations* section).

RQ2: Developing learning-oriented feedback materials
RQ2 of this study relates to our aim of developing a checklist and accompanying materials to provide meaningful feedback on IC features in the *B2 First* discussion task. Feedback to learners is consistently identified as a key dimension in LOA frameworks (Carless, 2007; Jones & Saville, 2016; Turner & Purpura, 2016). When formulated and delivered effectively, feedback helps learners identify strengths and weaknesses within their current performance and provides them with guidance on how to improve. This resonates with the idea of noticing in SLA as a key process in language learning (Schmidt, 1990) and the principle of focusing one's attention on areas of performance needing deliberate practice (Ericsson, 1996).

The checklist and the accompanying feedback developed in this research represents an initial step towards building the infrastructure for LOA of IC in discussion tasks, and the materials were designed in keeping with LOA principles. Firstly, the checklist and accompanying materials provides information about *specific* aspects of the learners' performance (Wiliam, 2011). The checklist (both the full and concise versions) identifies more IC features than the rating scale descriptors, captures each feature to several levels of granularity (micro-features within macro categories), and the accompanying materials describe and exemplify the features in detail. Secondly, we recognized the importance of presenting feedback in a manner and at a level of detail appropriate to learners (Hattie & Timperley, 2007). The feedback section drew on the thematic analysis of examiners' recommendations and built on the descriptions of IC features, but was simplified and tailored to learners – e.g., shortened sentence length, simplified syntax, technical jargon removed. Thirdly, and perhaps most importantly, the feedback was constructed in a forward-looking way (Carless, 2007). Both the *Well done!* and *Needs more work* sections aim to draw learners' attention to aspects of their current performance and outline ways to improve. For example, under the heading *Keep the discussion going* in the concise version (see Nakatsuhara et al., 2018, p. 68), one of the feedback comments for a learner who *Needs more work* is:

> ***Develop (own idea)***: It is important not to state your choice (e.g., "I think a café is good.") or simply describe an idea or picture and stop there. Next time, try to ***extend your ideas*** by giving reasons for your choice, giving examples, or providing more details.

This helps the learner to identify weaknesses in the current performance and gives recommendations for improvement. Similarly, for learners getting "Well done!" under *Initiate new ideas*, the feedback comment reads:

> It is great that you bring new ideas into the discussion. To do this even better, make sure you think about 1*) **when** to bring in a new idea (has the last one been fully discussed?), 2) **how** to introduce it, and 3) **how relevant** it is to the task (ibid., p. 68).

Thus, the feedback highlights what the learner has done well and how s/he can achieve an even more skilled performance. This way of providing feedback is supported in Dunlop (2017), who found that learners devote most attention to personalized feedback, particularly "information about how to move forward and improve" (p. 184). The teacher can make use of this feedback and further personalize it to individual learners in accordance with their performance.

Implications for learning and teaching
In summary, the materials developed in this project have potential for LOA use in a range of ways. First, they raise learners' awareness of success criteria (Turner & Purpura, 2016), and help them notice any problematic areas needing more focused practice. In fact, an extension of the checklist and providing success criteria is offered by Lam (2019) who provides annotated video examples of interactional features captured in the checklist. The checklist, as an observation record of learner performance, also gives teachers diagnostic information on the basis of which they can modify learning objectives (Jones & Saville, 2016). Secondly, the simplified, concise checklist can be used as a tool by learners themselves to engage in self or peer assessment, which promotes self-regulated and autonomous learning (Carless, 2007). Finally, the checklist, and particularly the detailed version with the accompanying materials, constitutes a rich resource for use in lesson planning or materials writing (e.g., Week 1: starting the discussion; Week 2: responding to the partner; Week 3: maintaining the interaction).

Implications for assessment

The checklist and materials developed in this study were based on examiners' comments on candidates' paired interactions in the *B2 First* Speaking test. As such, the findings also carry implications for the *B2 First* paired interaction task, its rating scales and rater training. First, by identifying eight macro and 27 micro themes related to IC features, this study confirms that the paired task in this test elicits a rich variety of interactional features, and five of the eight main categories are reflected in the current rating scale. Moreover, the checklist provides enhanced resources for examiner training in awarding IC scores, thus contributing to the scoring validity for the *B2 First* Speaking test.

The IC features identified, especially the non-criterion features, also provide construct information for rating scale development. Interactive listening, a salient IC feature emerging in previous research (e.g., Ducasse & Brown, 2009; Galaczi, 2014, Ross, 2018), was well attended to by the examiners and constituted a macro theme in the analysis. Interactive listening is often implicitly captured in the rating scale through descriptors focusing on development of the interaction. On the basis of our findings, we would recommend explicitly including this feature in IC-related rating scales. Of course, there is the caveat of how it could be measured/evidenced, considering that backchannelling and nodding are displays of interactive listening but do not necessarily entail comprehension. (May, 2011; Lam, 2018). One suggestion for assessing candidates comprehension of co-participants talk, based on Lam (2018), is to look at whether candidates provide next-turn responses which are contingent on previous speaker contribution (e.g. paraphrasing or summarizing, giving reasons for (dis)agreements, or extending previous speakers ideas). Rater training can involve sensitizing raters to notice these types of responses.

Limitations and future research

The IC checklist and feedback materials developed in this study were based on performance on one particular test and one task type: a decision-making discussion, which may affect the generalisability of the findings. They also focused on IC features at B2 and adjacent B1/C1 levels (Cambridge English, 2016). Other interactional task types including a different range of functional language (e.g. arguing/debating, complaining, delivering bad news) may invoke context-sensitive IC repertoires (c.f., Hall & Pekarek Doehler, 2011; Young, 2011) with features both overlapping with and distinct from the current checklist. They would be worthwhile avenues to explore in future research.

Another limitation concerns the tension between construct coverage and practical usability. We had aimed at a more comprehensive coverage of the IC construct reflecting the full range of examiner comments and features not readily captured by the rating scales descriptors. However, the focus group confirmed that the detailed checklist would place heavy processing demands on teachers using it in classroom assessment to evaluate learners' interactive performance in real time. The development of the concise checklist represented an effort to reconcile the tension between breadth/depth and usability by teachers/examiners. We should, however, keep in mind that while the concise checklist was developed based on the salience perceived by examiners and feedback provided by teachers, less salient features might also be equally important to successful IC performance in certain contexts. It is therefore necessary to explore the usefulness of the concise version as well as the full version in different settings and amend them as necessary. As highlighted earlier, these checklists are not meant to be static products; they were developed to be further enhanced and modified to suit different teaching and testing contexts. In a similar vein, the checklist is aimed at providing some basis for teachers to offer feedback, but it is not meant for them to stop there. We recommend that teachers should draw on and build on the checklist features flexibly and in context-sensitive ways that respond to their students' performance. We stress that the feedback can, and should,

be adapted accordingly. Finally, the reception and use of the checklist by learners was beyond the scope of this study, but is certainly an important area for future research.

These issues point to the need for continued research efforts in describing the rather fluid and complex construct of IC. This study represents one such effort to define the construct precisely in the context of a discussion task, and present it to learners, teachers, material writers and examiners, who all engage in the teaching, learning, and assessment of this important aspect of second language ability.

References

Atkinson, J. M., & Heritage, J. (1984). *Structures of social action.* Cambridge, UK: Cambridge University Press.

Bourdieu, P. (1991). *Language and symbolic power*. Cambridge, MA: Blackwell.

Cambridge English. (2016). *Cambridge English: First*. Retrieved from http://www.cambridgeenglish.org/images/167791-cambridge-english-first-handbook.pdf

Carless, D. (2007). Learning-oriented Assessment: Conceptual Bases and Practical Implications. *Innovations in Education and Teaching International*, *44*(1), 57–66. doi:10.1080/14703290601081332

Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment – Companion volume with new descriptors*. Retrieved from https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

DeKeyser, R. M. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology.* Cambridge, UK: Cambridge University Press.

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

Dunlop, M. (2017). *Maximizing feedback for language learning: English language learners' attention, affect, cognition and usage of computer-delivered feedback from an English language reading proficiency assessment*. Unpublished doctoral dissertation, University of Toronto, Canada.

Ericsson, K. A. (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. Mahwah, NJ: Erlbaum.

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. doi: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. doi:10.1093/applin/amt017

Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, *15*(3), 219–236. doi: 10.1080/15434303.2018.1453816

Gass, S., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Erlbaum.

Green, A. (1998). *Verbal protocol analysis in language testing research*. Cambridge, UK: UCLES/Cambridge University Press.

Green, A. B. (2016). Assessment literacy for language teachers. In D. Tsagari (Ed.), *Classroom-based assessment in L2 contexts* (pp. 8–29). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

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. 1–15). Bristol, UK: Multilingual Matters.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. doi:10.3102/003465430298487

Ishida, M. (2011). Engaging in another person's telling as a recipient in L2 Japanese: development of interactional competence during one-year study abroad. In G. Pallotti, & J. Wanger (Eds.), *L2 Learning as social practice: Conversation-analytic perspectives* (pp. 45–85). Honolulu, HI: National Foreign Language Resource Center, University of Hawai'i.

Ikeda, N. (2017). *Measuring L2 Oral Pragmatic Abilities for Use in Social Contexts: Development and Validation of an Assessment Instrument for L2 Pragmatics Performance in University Settings*. Unpublished PhD thesis, University of Melbourne, Australia.

Jones, N., & Saville, N. (2016). *Learning oriented assessment: A systemic approach*. Cambridge, UK: UCLES/Cambridge University Press.

Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal, 70*(4), 366–372. doi:10.2307/326815

Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research (5th ed.)*. Thousand Oaks, CA: Sage Publications.

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

Lam, D. M. K. (2019). Enhancing learning-oriented feedback for Cambridge English: First paired interactions. *Research Notes*, *75*.

Leaper, D. (2014) *Consistency in performance in the group oral discussion test: An interactional competence perspective*. Unpublished PhD thesis, Macquarie University, Australia.

May, L. (2011). *Interaction in a Paired Speaking Test: the Rater's Perspective*. Frankfurt am Main, Germany: Peter Lang.

Nakatsuhara, F. (2013). *The co-construction of conversation in group oral tests*. Frankfurt am Main, Germany: Peter Lang.

Nakatsuhara, F., May, L., Lam, D., & Galaczi, E. (2018). Learning oriented feedback in the development and assessment of interactional competence. *Research Notes*, *70*, 1–68.

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. doi:10.1093/applin/amw021

Plough, I., Banerjee, J., & Iwashita, N. (2018). Interactional competence: Genie out of the bottle. *Language Testing*, *35*, 427–445. doi:10.1177/0265532218772325

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. doi:10.1177/0265532218758128

Ross, S. (2018). Listener response as a facet of interactional competence. *Language Testing, 35*(3), 357-375. doi:10.1177/0265532218758125

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*, 129–158. doi:10.1093/applin/11.2.129

Trinity College London. (2015). *Integrated Skills in English (ISE) Guide for Teachers — ISE II (B2)*. Retrieved from http://www.trinitycollege.co.uk/site/?id=3195

Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in the classroom. In D. Tsagari & J. Banerjee (Eds.). *Handbook of second language assessment* (pp. 255–273). Boston, MA: DeGruyter Mouton.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Yin, R. K. (2011). *Qualitative research from start to finish*. New York, NY: Guildford Press.

Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing, 32*(2), 199–225. doi:10.1177/0265532214557113

Young, R. (2011). Interactional competence in language learning, teaching, and testing. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning Volume II* (pp. 426–443). New York, NY: Routledge.