Interactional Competence: Conceptualisations, Operationalisations, and Outstanding Questions

Evelina Galaczi & Lynda Taylor

To cite this article: Evelina Galaczi & Lynda Taylor (2018): Interactional Competence: Conceptualisations, Operationalisations, and Outstanding Questions, Language Assessment Quarterly, DOI: 10.1080/15434303.2018.1453816

To link to this article: https://doi.org/10.1080/15434303.2018.1453816

Published online: 30 Apr 2018.

Article views: 54

View related articles

View Crossmark data
**Interactional Competence: Conceptualisations, Operationalisations, and Outstanding Questions**

Evelina Galaczi and Lynda Taylor

*Cambridge Assessment English, University of Cambridge Assessment; Centre for Research in English Language Learning and Assessment, University of Bedfordshire*

**ABSTRACT**

This article on interactional competence provides an overview of the historical influences that have shaped theoretical conceptualisations of this construct as it relates to spoken language use, leading to the current view of it as involving both cognitive and social dimensions, and then describes its operationalisation in tests and assessment scales, and the challenges associated with this activity. Looking into the future, issues that need to be dealt with include developing a fuller representation of the construct and of more contextually relevant assessments, deciding upon additional assessment criteria and the appropriate interpretation thereof, and determining how technology can be applied in assessment practice and the extent to which technology fundamentally changes the construct itself. These all have implications for testing if it is to be relevant and fit for purpose.

Spoken interaction is a fundamental but also complex endeavour. It is dynamic and co-constructed, it evolves and emerges, and is shared between interlocutors. It is reciprocal and those involved are both pro-active and re-active at the same time, simultaneously deconstructing messages as listeners and constructing their own message as speakers. Interaction is not necessarily linear, predictable, or tidy, and it is strongly shaped by diverse personal cognitive and contextual factors. It is, from a sociological perspective, the “primordial site of sociality” (Schegloff, 1986, p. 112), and, from a cognitive psychology perspective, something we are wired for: “humans are designed for dialogue rather than monologue” (Garrod & Pickering, 2004, p. 8).

Spoken interaction in a second/foreign (L2) language has assumed greater importance in the last two decades, in light of the growing role of the communicative approach to language teaching, learning, and assessment. As speaking tests have evolved to capture interaction (e.g., through using paired or group formats) and as technology has started to play a significant role in assessment, a need has developed to more accurately understand and describe the construct of interactional competence. Our current understanding of interactional competence, with its implications for L2 teaching, learning, and assessment, has not received extensive and focused attention, apart from a handful of theoretical papers and empirical studies (e.g., Chalhoub-Deville, 2003; East, 2016; Galaczi, 2008, 2014; He & Young, 1998; Kramsch, 1986; May, 2009; Nakatsuhara, 2013; Taylor & Wigglesworth, 2009). In this article we focus on the historical evolution and current understanding of interactional competence in L2 speaking assessment. Our aim is twofold: (a) to seek a comprehensive definition of interactional competence drawing upon theoretical and empirical work to date and (b) to highlight some challenges and issues for further consideration, as well as areas of investigation in the assessment of L2 spoken interaction. In addressing these twin aims, we ask more questions than we answer, in the hope that they will provide food for thought for future theoretical and empirical endeavours in our field.
Historical factors influencing theoretical conceptualisations of interactional competence

The role of interaction in L2 assessment and the path to our current understanding of interactional competence have been shaped by a number of pedagogical and sociopolitical developments over the past 100 years. These have directly and indirectly given impetus to influential theoretical ideas and useful empirical endeavours that have in turn informed our understanding of the construct of interactional competence.

One of the first key influences on the role of interactional competence in assessment can be traced back to language pedagogy in the 1880s when the Direct Method (with its focus on oral fluency) and the Reform Movement (with its focus on phonetics and pronunciation in language classrooms) began to take their place alongside the classical Grammar-Translation Method. The primary emphasis on speaking in these pedagogic approaches paved the way for a change of focus in L2 assessment at the beginning of the 20th century. That change can be seen in the introduction of tests that included a conversational component, such as the Cambridge Certificate of Proficiency in English (CPE) in 1913 and the Lower Certificate in English (LCE) in 1939 (both in the United Kingdom), and the Army Specialized Training Program (ASTP) in 1943 in the United States. These pedagogic changes did not appear in a vacuum but were supported by the sociopolitical climate of the 1930s and 1940s, which, as a result of two world wars and more widespread travel, fostered language learning and highlighted the shortage of English L2 speakers (Weir, Vidakovic, & Galaczi, 2013).

The next significant influence on interaction in L2 assessment came in the 1970s, the decade that saw the work of Hymes (1974) on performance vis-à-vis competence, Halliday (1975) on the functional aspect of language, and van Dijk (1977) on discourse and text/context relationships. Hymes’s (1974) ethnography of communication highlighted the central role of the speech situation (e.g., a party), speech event (e.g., a conversation during the party), and speech act (e.g., a joke in the conversation), and the need for these contextual factors to be seen as shaping conversation. These influential shifts in thinking paved the way for the model of communicative competence proposed by Canale and Swain (1980) and further expanded by Canale (1983), Bachman (1990), and Bachman and Palmer (1996), which broadened the focus in L2 assessment, going beyond grammatical competence to considerations of sociolinguistic, discourse and strategic competence. Kramsch’s work in 1986—the first to make use of the term interactional competence—further expanded the construct of interactional competence in her discussion on the “dynamic process of communication” (Kramsch, 1986, p. 368). Kramsch noted that “interaction always entails negotiating intended meanings, i.e. adjusting one’s speech to the effect one intends to have on the listener. It entails anticipating the listener’s response and possible misunderstandings, clarifying one’s own and the other’s intentions” (1986, p. 367). The focus on interaction challenged established notions of communicative competence and brought into contrast, on the one hand, conceptualisations of competence residing within an individual and, on the other, competence that is manifested in co-constructed interaction between individuals.

Thus, throughout the 1980s and 1990s we saw a growing awareness of the interactional nature of spoken language ability. Theoretically, the construct of interactional competence in L2 assessment became broader and more nuanced, expanding to a more social view where communicative language ability and the resulting interactional performance reside within a social and jointly constructed context. Supporting this line of thought, McNamara (1997) and McNamara and Roever (2006) argued that interaction can be defined in two contrastive ways: as a psychological construct residing within an individual and as a social construct that involves joint construction of interaction between individuals, who each have their own sociocultural identity. Models of communicative language ability, for example, Bachman and Palmer (1996), include interaction in their model, although they do not explicitly discuss its role, so their model represents a psychological conceptualisation of interaction. In contrast, an interactionist approach involves explicit recognition of the importance of social context and its influence on the interaction produced. Conceptualisations of interactional
competence, as proposed by Kramsch (1986) and more recently He and Young (1998) and Young (2008), argue for a construct that explicitly includes the co-construction of interaction and goes beyond individual ability. Therefore, it encompasses elements such as awareness of the roles of the individuals in the interaction and the context, as well as interactional resources, such as the appropriate use of speech acts, managing turn-taking, repairing conversation breakdown, and the ability to use visual behaviours, such as eye contact, posture, and facial expressions effectively (Young, 2008). In this line of thinking, speaking is viewed both as a cognitive and a social interactional trait, with emphasis not just on the knowledge and processing dimension of language use, as seen in the Bachman and Palmer (1996) model, but also on the social, interactional nature of speaking, which has as its primary focus the individual in interaction. As such, the interlocutors and the host of variables they bring to the interactional event become part of the construct of L2 interaction and have implications for the validity considerations supporting the assessment. As Chalhoub-Deville argued: “individual ability and contextual facets interact in ways that change them both” (2003, p. 369).

**Insights into interactional competence from theoretical and empirical research**

As models of communicative competence found their operationalisations in different speaking test formats (individual, paired, and group), two important theoretical and empirical research strands emerged with implications for test validity and construct definition. These two strands found expression in concurrent debates on authenticity and variability. In addition, methodological advances in Conversation Analysis, as well as in Corpus Linguistics, played a role in building a more in-depth understanding of the construct of interactional competence. We briefly explore each of these factors in turn.

**The debate over authenticity**

Individual oral tests (often referred to as Language Proficiency Interviews) arrived on the scene in 1913, with the introduction of the Certificate of Proficiency in English by the University of Cambridge Local Examinations Syndicate (Weir et al., 2013); such tests became widespread in L2 assessment from the 1950s onwards (as seen in the Foreign Service Institute (FSI) test, the Interagency Language Roundtable (ILR) test and American Council on the Teaching of Foreign Languages Oral Proficiency Interview (OPI) tests in the United States, and the Cambridge Proficiency in English and First Certificate in English in the United Kingdom).

All these tests involved one examiner and one test taker engaged in an interview-like interaction, with the examiner leading the test and the test taker responding. As this test format became widespread, it attracted research attention in the 1980s, which highlighted its strengths as well as certain limitations. A key question to emerge was whether a single type of interaction in an oral test (i.e., interview-style interaction) is sufficient to assess oral proficiency and whether it is representative of everyday, naturally occurring spoken interaction. Thus, test authenticity became the focus of attention and empirical investigation.

Research found the individual format to be limited in interactional roles and range of language functions (e.g., Brooks, 2009; ffrench, 2003; Johnson, 2001; van Lier, 1989) because information questions, which were observed to be central to language proficiency interviews, did not have such a prominent role in natural non-test conversation (Lantolf & Frawley, 1988). Individual oral tests were also found to be limited in alignment with L2 communicative classrooms, which were becoming more widespread during the 1980s and which emphasised interaction in different contexts. As Savignon (1985, p. 132) remarked, “among the many contexts not sampled in an individual format are small group discussion, playing a game, or conducting a survey, contexts requiring very different discourse strategies; strategies that teachers often encourage, or would like to encourage, in their classrooms.” This body of empirical research was the catalyst for the development of more refined
interactional speaking assessments and was influential for the development of the conceptualisation of interactional competence, because it signalled that even though useful for assessment purposes, the language generated in individual interview-format language tests is not spoken interaction as we typically understand it.

Following the now classic appeal by van Lier (1989) to look inside the language proficiency interview (i.e., to analyse the actual discourse produced and not just look at test scores), a solid body of literature from the 1990s revealed what language this test format can and cannot generate. Lazaraton (1992) used Conversation Analysis techniques to investigate the language generated in individual oral tests and concluded that even though this test format shared some features with natural interaction, the pre-specified system of turn-taking in this format aligned it with the interview speech event. Ross and Berwick, in a series of studies that focused on the types of accommodations employed by examiners during the test, suggested that the language proficiency interview is “a hybrid of interview and conversational interaction” (Ross & Berwick, 1992, p. 160). Young and Milanovic (1992) investigated the notion of control and contingency in the language proficiency interview and concluded that oral interviews are asymmetrically contingent (i.e., one interlocutor leads, the other one follows); as such, the authors concluded, they bear little resemblance to the co-constructed nature of interaction. Johnson and Tyler (1998) and Johnson (2001)—perhaps the most strident critics of the Oral Proficiency Interview—focused on the turn-taking behaviour of the participants in an individual oral test and suggested that the OPI could not be considered a valid example of a typical, real-life interaction because salient features of natural conversation, such as turn-taking management and topic negotiation, were not present. They added: “naturally occurring conversation is by its very nature interactive, and . . . a crucial part of this interactiveness is a sense of involvement or reactivity among interlocutors” (Johnson & Tyler, 1998, p. 48). This, the authors maintained, is not what happens in the OPI.

The empirical focus on the nature of the interaction in individual oral tests was paralleled by theoretical developments focusing on the notion of authenticity, as seen in Bachman’s (1990) and Bachman and Palmer’s (1996) distinction between situational and interactional authenticity and a decade later in Weir’s (2005) distinction between context and cognitive validity. These conceptualisations offered a twofold distinction: on the one hand, the degree of correspondence between the test task and the non-test domain (Bachman’s situational authenticity and Weir’s context validity) and, on the other hand, the degree of correspondence between the cognitive processes triggered by a test task and a non-test task from the target language use domain (Bachman’s interactional authenticity and Weir’s cognitive validity). A test by definition would have limited situational authenticity (as Stevenson (1985) reminded us, this is a test, not a tea party); however, depending on the tasks and types of interaction generated, a test could have different degrees of interactional authenticity.

The 1990s saw a steady growth in the use of paired and group oral assessment; this was partly as a response to the move toward a more communicative approach in language teaching and learning and partly as a reaction to some of the perceived limitations of the individual speaking test format, such as the relatively restricted range of tasks and types of interaction and the unequal distribution of rights and responsibilities between the examiner and test taker. The paired/group test format, which creates opportunities for peer-peer interaction, was seen as a viable alternative to the individual language proficiency interview, offering more balanced conversational rights and responsibilities and a wider spectrum of functional competence.

As paired and group testing became more widespread, empirical attention turned to this test format and showed that, in contrast to the individual speaking test format, paired/group oral tasks result in more symmetrical interaction possibilities (Együd & Glover, 2001; Galaczí, 2008; Iwashita, 1998; Kormos, 1999; Lazaraton, 2002); they elicit a wider sample of learner performance with a broader range of speech functions (ffrench, 2003; Galaczi & ffrench, 2011) and provide more opportunities for test takers to display conversational management skills (Brooks, 2009; Gan, Davison, & Hamp-Lyons, 2009; Kormos, 1999; O’Sullivan, Weir, & Saville,
Brooks (2009), for example, found more complex interaction between participants in the paired configurations than in individual ones: she observed more prompting, elaboration, finishing sentences, referring to a partner’s ideas, and paraphrasing in the paired format. Gan et al. (2009) found that equal power distribution and a range of speech functions associated with topic negotiation (e.g., initiating, expanding and closing or discarding a topic) were present in the group test performances they analysed. Another body of literature also indicated that the paired/group test construct encompasses interaction management features, such as turn-taking management, topic initiation, negotiation and development, and interactive listening (Ducasse & Brown, 2009; Galaczi, 2014; May, 2009; Riggenbach, 1998; Storch, 2002). From a cognitive perspective, differences between paired and individual tests have been reported as well, such as having to process input from several interlocutors (including a peer), keeping track of different points of view and topics, having familiarity with each other’s L2 variety, and forming judgements in real time about the extent of accommodation to the partner’s language. Such cognitive demands are largely absent in individual interview tests (Field, 2011). Positive washback on classroom teaching in exposure to a wider range of real-life interactions was found to be a further feature of the paired format (Együd & Glover, 2001), and paired/group oral tests were seen as either mirroring good practice in the classroom or encouraging it (East, 2016), thus promoting language acquisition, both from a cognitive perspective (e.g., Long’s [1996] Interaction Hypothesis) and a sociocultural perspective (e.g., Vygotsky’s [1978] social constructivist theory).

The empirical insights provided by a growing body of literature on test authenticity have been significant for the definition of L2 interactional competence, because they have indicated that the construct underlying paired/group oral tests is broader than the individual speaking test construct and is more closely aligned with the conceptualisation of interactional competence as a sociocognitive construct.

**The debate over variability**

The findings from empirical studies on oral tests in the 1980s and 1990s highlighted a further fundamental characteristic of speaking tests, namely, the *variability* associated with them, which results from the unpredictability of co-constructed interaction. Test variability quickly became a controversial issue in interactional speaking tests, largely because of the tension between authenticity (a validity asset) and reliability (a validity threat).

In examiner variability in individual tests, Lazaraton (1996) focused on the language generated between speaking examiners and test takers, and she documented deviations from the examiner script for the Cambridge Assessment of Spoken English test (CASE). The deviations consisted of different interviewer support and accommodative behaviour across tests, such as repeating questions at a slower rate, echoing and correcting responses, giving evaluative responses, and supplying vocabulary in some cases. Brown (2003), in a now seminal paper, also focused on variability across interviewers in the ways they elicited test takers’ speech, and she documented that different interviewing techniques could affect a test taker’s score. Using Conversation Analysis as the methodological approach, the author analysed two individual oral tests involving different interviewers but the same test taker and illustrated “how intimately the interviewer is implicated in the construction of test taker proficiency” (Brown, 2003, p. 1). Brown showed that a test taker came across as a “willing and responsive interlocutor” and was given a higher score when interviewed by an examiner who used more explicit questioning techniques and topic development, regularly provided feedback, and showed interest in what the test taker was saying. The same test taker was perceived as “unforthcoming or uncooperative” and had a lower score when interviewed by an examiner who used implicit closed questions and echoed the test taker’s speech; the latter were misinterpreted by the test taker as interactional exchange closures, rather than prompts for more information, and so the test taker did not elaborate.
The last few decades have also seen a focus on variability through the lens of test reliability. A body of research that focused on group oral tests and rater agreement indicated lower inter-rater correlations than those found in individual interview tests (e.g., Reves, 1981) and differences in severity between raters in group tests, even in the case of experienced and trained raters (Bonk & Ockey, 2003; van Moere, 2006). This brought into question the scoring validity of this test format and supported the notion that “the less interaction in a task, the greater the degree of control over the rating process” (van Moere, 2006, p. 435). Or, put differently, in the tug-of-war between reliability and construct coverage, if test developers prioritised test reliability, then interaction tasks, and especially paired and group tasks, were seen as problematic, because by design they have lower degrees of control and therefore higher degrees of variability.

A related body of research in the 1990s and early 2000s highlighted another fundamental source of test variability, namely, the sociocultural identity and personal characteristics that interlocutors (both test takers and examiners) bring to the speaking test. The so-called “interlocutor effect” (O’Sullivan, 2002) highlighted the influence of the interlocutors’ personal characteristics on the interaction produced in speaking tests (e.g., Berwick & Ross, 1996; Katona, 1998; Nakatsuhara, 2013; O’Loughlin, 2002; O’Sullivan, 2002; Young, 2008).

The interlocutor effect was found to be present in both individual and paired/group tests, but it was potentially more pronounced in paired/group tests. The body of literature that has investigated the effect of interlocutor variables in paired/group speaking tests has documented a variety of interlocutor factors influencing test takers’ performance, such as: test takers’ acquaintanceship and gender (Chambers, Galaczi, & Gilbert, 2012; O’Sullivan, 2002), personality (Berry, 1993; Ockey, 2009), extroversion (Nakatsuhara, 2013), talkativeness (van Moere & Kobayashi, 2004), and language proficiency (Davis, 2009). The key insight to emerge from this body of research was that despite the clear influence of personal characteristics on the co-constructed interaction in an L2 oral test, the magnitude or direction of that influence was found to be less clear and not directly predictable (Brown & McNamara, 2004). For example, Berry’s (2004) analysis, using a group test, indicated an advantage for both extrovert and introvert students, depending on the group: introverts performed better in high extrovert groups, and extroverts performed better in high introvert groups. Nakatsuhara (2013), also focusing on a group oral test, concluded that the extroversion level of test takers had some influence on test performance, but—importantly—was closely related to task type (among other factors, such as group size). The author found that introverts performed better in structured, highly prompted tasks, whereas extroverts received higher scores when given a higher degree of freedom. This echoed Berry’s assertion that the effect of personality could be dominant when “either extreme is placed in their least favoured situation” (Berry, 2004, p. 502). The role of test taker’s familiarity was a further empirical strand, as seen in O’Sullivan’s (2002) investigation of the effect of test taker’s familiarity on the performance of Japanese test takers in paired speaking tasks. The author found evidence of an “acquaintanceship effect,” reporting that test takers achieved higher scores when working with a friend. The author also found that the gender of the interlocutor played a role and further speculated that the effect of interlocutor familiarity and gender may be culturally specific. Chambers et al. (2012) supported the contention that cultural context influences interlocutor behaviour in their investigation of the role of test taker’s familiarity in a Swiss context. On the basis of quantitative score data and qualitative questionnaire and interview data, the authors found that test taker’s familiarity played a significant role in the Swiss context of their study (carried out with Italian, French, and Swiss German learners) but had a small effect size. A key finding to emerge from this body of research, therefore, was that the role of personal characteristics is dependent on a host of personality and contextual factors. As Brown and McNamara (2004) argued in the context of gender-related effects, interlocutor personal variables “compete in the context of an individual’s social identity” and no linear, clear-cut behaviours based on interlocutor characteristics can be claimed (p. 533). Such findings about variability in interactional speaking tests are not altogether surprising, because sociolinguistic research has unequivocally indicated that characteristics, such as gender, cultural/L1 background, personality, acquaintanceship, and the role of the
participants can affect the amount and quality of interaction (Beebe, 1980; Wolfson, 1989). Who one talks to, in other words, is not unimportant, because the characteristics of the interlocutor affect the way we speak. We are all, so to speak, linguistic chameleons.

The dynamic two-way influence of the interlocutors has now become central to the construct of interactional competence. The key construct-related questions, therefore, have shifted from “Is there variability in those test formats?” (we now know that there is, because variability is inherent in interaction) to “What should test developers do about such variability?” and “Should they try to eliminate such variability altogether or should they develop tests which tap into the construct of interactional competence?” In other words, the key question becomes whether the variability in interactional oral tests (with all their inherent limited psychometric control and unpredictability) should be seen as construct-irrelevant variance or whether it should be seen as part of the construct. Swain (cited in Fox, 2004) provides some useful insights: the variability related to different characteristics of conversational partners is “all that happens in the real world. And so they are things we should be interested in testing” (p. 240). van Moere extends this argument: the variability reported in interactional tests and associated effects “only threaten the validity of this test format as long as interactional competence is not part of the construct being measured” (van Moere, 2006, p. 415).

Developing an argument along similar lines, East (2016, p. 27) notes: “spoken communicative proficiency cannot be fully determined without reference to some kind of interactional ability, and this interactional ability presupposes the ability to deal with the unexpected.” The argument here is for the need for a speaking test construct defined in social terms that taps into interaction and goes beyond a purely psychological conceptualisation, and that encompasses a certain degree of unpredictability and less control than the individual interview format speaking test. Engaging in interaction successfully with different interlocutors is now seen as a fundamental part of real-life interactional demands, and it can therefore be argued that the associated variability is part of the construct underlying communicative language tests.

**Insights from Conversation Analysis and Corpus Linguistics**

Finally, our understanding of conversational organisation, and therefore interactional competence, was greatly influenced by advances in Conversation Analysis (henceforth CA), which started to gain momentum in the 1970s and had a significant impact on the nature of studies of interaction. CA was influenced by Garfinkel’s ethnomethodology and Goffman’s interaction analysis (Schiffrin, 1994), and within linguistics it took shape in the work of Sacks, Schegeloff, and Jefferson (1974). A basic belief of CA is the view that interaction is structurally organised, and its methodological lens focused on elements of interaction, such as the organisation of turn-taking, the adjacency pair, preference structure, the organisation of repair, topic management, conversational dominance, and the role of sociological variables in interaction.

CA proved a useful methodology in L2 assessment research, and the resultant body of research, mostly carried out in the first decade of the 21st century, moved understanding of the interactional competence construct forward through focusing on its microlevel features. This body of research explored general notions of topic initiation and response and moved to more fine-tuned and specific aspects of interactional competence, such as interactional means to develop topics across speakers and turns, topic shifts, listener involvement, and turn-taking strategies (Galaczi, 2008, 2014; Gan, 2010; Gan et al., 2009; Lazaraton, 2002; May, 2009; Nakatsuhara, 2013).

The more widespread availability of spoken language corpora has also enabled a growing body of literature to emerge that analyses large-scale corpus data to investigate interactional features. For example, Tao and McCarthy (2001) investigated the role of syntactic elements in linking interactional turns and found that vague language (e.g., “or something”) often triggers a change of speaker (in L1 interaction). Evison, McCarthy, and O’Keefe (2007) looked at how vague language can create a shared social space. Tao (2003) extended this line of research by analysing the use of turn-opening tokens (e.g., “yeah” or “yes, sure”) in interaction and demonstrated their
role in signaling mutuality in topic development across turns. For now, this work is still in the realm of native-speaker interaction, but is a useful stepping stone to interactional features in learner language, especially because work on learner spoken corpora is already making strides (Gablasova, Brezina, McEnery, & Boyd, 2015).

**Defining a construct of interactional competence**

Thus far we have explored conceptualisations and operationalisations of interactional competence as seen through the lenses of theoretical modelling, empirical research, and assessment practice. This overview has confirmed the complex and multifaceted nature of interactional competence, as well as the challenge of attempting to articulate a working definition that resists oversimplification and that serves to advance the operationalisation of the construct for the purposes of L2 speaking assessment.

A broad macrolevel definition to emerge from the discussion so far is that interactional competence is the ability to co-construct interaction in a purposeful and meaningful way, taking into account sociocultural and pragmatic dimensions of the speech situation and event. This ability is supported by the linguistic and other resources that speakers and listeners leverage at a microlevel of the interaction, namely, aspects of topic management, turn management, interactive listening, breakdown repair and non-verbal or visual behaviours. This view is consistent with Young’s assertion that interactional competence is “distributed across participants and varies in different interactional practices” (2011, p. 430).

It is difficult to represent the construct of interactional competence by means of visuals, such as a table, line diagram, or flowchart, because these tend to suggest fixed hierarchical or causal relationships between elements, which are actually flexible, fluid, and subject to fuzzy boundaries. For this reason we have chosen a more metaphorical visual representation (in the form of a tree) in the hope that it suggests plausible but dynamic relationships between the various elements of interest. A tree illustration is also more consistent with the organic and dynamic nature of spoken interaction; like a tree, spoken interaction can grow out of various types of terrain, and it can take various shapes according to a multiplicity of internal and external factors impacting its development.

**Figure 1** offers a visual representation of the phenomenon of interactional competence as currently understood. It draws together various elements that the theoretical and empirical literatures have identified as underlying the construct, locating them simultaneously within the macrolevel context of the speech situation and the microlevel context of the speech event and speech act. The main trunk of the tree represents the interlocutors, who are colocated, as a pair or a group, within a shared time and space, regardless of whether they are interacting face-to-face or online; Young refers to them as “interactional partners” (2011, p. 428). Their respective interactional skills of topic management, turn management, etc. are shown as larger limbs of the tree from which emanate smaller branches representing microfeatures of each skill (e.g., topic management encompasses initiating, extending, shifting, and closing down topics). Representation of the microfeatures that reflect interactional competence is not intended to be exhaustive. **Figure 1** seeks to illustrate the current state of play in the definition of the construct and awareness of associated microlevel features relevant to L2 teaching, learning, and assessment. Though many of the microfeatures are beginning to find a place within teaching syllabuses and assessment scales, as we shall see in the next section of this article, there are some microfeatures that are yet to be acknowledged and accommodated within the construct definition; furthermore, some microfeatures are still emerging as empirical research on L1 and L2 spoken interaction begins to reveal their presence and importance. The unlabelled branches in **Figure 1** acknowledge this reality and offer space for such microfeatures to be added in over time as empirical investigation confirms their relevance.
Issues and challenges in operationalising and assessing the construct of interactional competence

Fine-tuning the construct

A key issue in our discussion must be to what extent the current definition of interactional competence (as illustrated in Figure 1) is sufficiently comprehensive for the purposes of language teaching, learning, and assessment, and whether it needs further specificity. Several related questions emerge: how far does the current construct conceptualisation take account of the full range of elements, which research now suggests are core to interactional competence? What additional elements of interactional competence might need to be represented, and thus sampled and evaluated for the purpose of assessment, as well as teaching and learning? Furthermore, what are the characteristics of learners’ interactional competence across proficiency levels? How do they vary across levels?

Figure 1. Defining interactional competence.
Some additional elements suggested by microlevel analyses of interaction (Galaczi, 2014; Gan et al., 2009) might be part of this construct but have not thus far received adequate empirical investigation and are therefore not typically part of assessment scales. They include interactional microfeatures identified in Conversation Analysis research (Hutchby & Wooffitt, 2008) and in recent research on the role of conversation in language learning and the interaction generated in interactional speaking tests (Leaper, 2014). Examples include the following:

- holding the conversational floor, e.g., through pausing or pitch
- assigning conversational rights, e.g., through asking questions or syntactic means
- use of deixis and ellipsis for between-turn cohesion
- use of vague language
- collaboratively completing turns.

Some of these features appear sporadically in the empirical literature on language assessment (e.g., Galaczi, 2008, 2014; Gan et al., 2009; Leaper, 2014; Tao, 2003), but they have not been explicitly articulated within the interactional competence definition.

A further issue is whether these additional micro-interactional features are located in our assessment scales and/or examiner training materials, and whether existing scales and training materials do justice to the richness of interactional competence. In most cases the answer is “only partially.” Scales capturing interactional competence have contributed greatly to operationalising a broad construct, but they (and/or supporting training materials for raters) could benefit from further development to capture other essential features. For example, the Common European Framework for Reference for languages (CEFR; Council of Europe, 2001) and its scales relevant to interactional competence draw attention to concepts, such as “initiating, maintaining and closing” conversations, “taking turns,” “confirming comprehension,” “inviting others in,” “keeping the floor,” and “relating contributions to those of other speakers” (2001, pp. 28–29, 86). These concepts are undoubtedly at the heart of the construct of interactional competence, but they may need some further refinement if they are to be usable for relevant stakeholders.

The American Council on the Teaching of Foreign Languages (ACTFL) OPI test, which includes interaction between an examiner and test taker, is assessed on a holistic scale, and that scale includes references to interactional skills, such as “engage in conversation in a clearly participatory manner,” “are able to handle with ease and confidence a large number of communicative tasks,” “participate actively in most informal and some formal exchanges” (American Council on the Teaching of Foreign Languages, 2012). Here again we see the need for more fine-tuning of the interactional skills displayed by test takers.

In many of the Cambridge English tests, interactional competence is assessed as part of a set of analytic scales, and the “Interactive Communication” criterion includes references to concepts, such as “initiates and responds appropriately, linking contributions to those of other speakers,” “maintains and develops the interaction,” and “keeps the interaction going with very little prompting and support” (Cambridge English, 2016). The Integrated Skills in English tests offered by Trinity College London include scales on “Communicative Effectiveness” and “Interactive Listening” in the Speaking and Listening test. Such scales include some fundamental features of the construct of interactional competence (e.g., active listening and the co-constructed nature of interaction), but they stop short of a full depiction of the construct.

Other interactional features identified in empirical endeavours from a range of disciplines, do not tend to appear in conceptualisations of interactional competence and/or empirical research at all. For example:

- genre awareness, e.g., sharing personal stories or exchanging ideas (Paltridge, 2001)
• sequencing practices in speech acts, especially where “face” is involved; e.g., the use of “face-saving pre-sequences” to avoid dis-preferred responses, as seen in responding to an invitation with a refusal (Pomerantz, 1984)
• politeness control (Brown & Levinson, 1987)
• nonverbal features, such as laughter, posture, gaze, and gestures (Ducasse & Brown, 2009; Gan & Davison, 2011; May, 2011).

Clearly, many of these features may be cross-culturally determined and may not be easy to capture within a universal definition of interactional competence. But all of these features contribute to good-quality interactional competence skills, so more exploration into how they could be operationalised in assessment scales would be beneficial.

Also of relevance to the construct definition of interactional competence is a recent upsurge in discussions of mediation, as seen in the recent Companion Volume to the CEFR (Council of Europe, 2017). Mediation involves the transfer of spoken (or written) language from a source language into a target language and is conceptualised as a distinct fourth mode of communication alongside reception, interaction, and production (Council of Europe, 2017). The exploration of mediation in the Companion Volume to the CEFR (Council of Europe, 2017) has implications for future discussions of the definition of interactional competence.

Construct coverage and reliability challenges

One consequence of a broad and comprehensive interactional construct is the multiple sources of variance, which in turn leads to some test reliability concerns. As van Moere (2006, p. 414) notes: “language testers are intuitively wary of assessing students in interactive groups because of the sheer number of uncontrollable variables and unknown effects associated with these variables.” Reliability is an essential aspect of test validity because the inferences made on the basis of scores need to be rigorous. As we have seen, test variability in interactional tests is an integral feature of such tests. Therefore, further exploration of the tension between reliability and the resultant need to increase control of tasks and narrow the construct coverage is needed. This is conflict with the authentic—and often unpredictable and therefore variable—performance produced during an interactional test (East, 2016). Thus, a key question emerges: how can interaction be captured in speaking tests without compromising the fundamental assessment requirement for reliability? Clearly, test design and scoring models could play a role here, for example, Rasch models, which take into account different sources of variability in the assignment of scores (Bonk & Ockey, 2003) or test design and the use of different interactional formats (e.g., a speaking test comprising a question-and-answer interview task, a monologue task, and some paired/group discussion tasks). This would allow test developers to optimise the benefits of an interactional construct while minimising some of the associated reliability caveats (Galaczi & ffrench, 2011).

Communicative context

Savignon (1983) described communication as context-specific: “communication takes place in an infinite variety of situations, and success in a particular role depends on one’s understanding of the context and on prior experience of a similar kind” (pp. 8–9). There has been, therefore, a growing recognition of the context-specific nature of spoken interaction. To what extent is “context relevance” or “context dependence” satisfactorily accounted for in the construct definition and operationalisation of interactional competence? Or put another way: are our speaking assessment tasks sufficiently context-rich? In some speaking tests much more attention is paid to this area than in others. For example, if we take the Occupational English Test targeted for health care professionals, a test taker (who is a professional doctor) would be expected to engage with the following task in a speaking test (Figure 2):
This assessment tool is very context-rich. The question, therefore, arises: if we wish to sample and evaluate interactional competence effectively, should we design our speaking test tasks to be more context-specific/context-relevant/context-rich, especially at higher proficiency levels (such as the C levels on the CEFR)? On a strong-weak performance assessment continuum (along the lines of McNamara, 1996), should we be designing interactional tests closer to the strong end of the continuum, where context and task completion is essential? There are, of course, also limitations to bringing this type of context-richness to all assessments, such as the role of personality traits, which might have to be considered part of the construct.

### Scalability and discriminability

A further question relates to scalability. How scalable are certain criteria, such as genre awareness or sequencing practices? Do they only merit a basic 3-point scale (appropriate/somewhat appropriate/not appropriate at all) or can they be assessed with a longer scale? Politeness and pragmatic competence is another notoriously difficult construct to measure in a valid and reliable way, largely because judgements about what might be deemed appropriate pragmatic behaviour can be highly culturally relative (Roever, 2013; Sydorenko, Maynard, & Guntly, 2014). But just because a construct is difficult to scale, does not necessarily mean we should not try.

There is also the question of discriminability and differentiation. What elements of interactional competence are best judged by a human rater using eye and ear? And which by automated scoring systems?

### Individual versus joint scores

The inherent interactional interdependence of the two or more test takers in paired/group tests poses further challenges for the assigning of scores. Both Swain (2001) and McNamara (1997), in their thought-provoking articles, raised the need to address the contentious issue of interpreting individual scores.
scores based on jointly constructed interaction distributed between/among participants. Chalhoub-Deville and Deville (2005) commented as follows: “If we view language as co-constructed, how can we disentangle an individual’s contribution to a communicative exchange in order to provide a score or assess a candidate’s merit for a potential position?” (p. 826). It is interesting that some researchers have since then argued for the awarding of shared scores for interactional competence in paired tasks (May, 2009; Taylor & Wigglesworth, 2009). They argue that we may have to design and use different assessment scales and criteria, some aimed at the assessment of individual performance, and some aimed at joint performance. This is a question that future research endeavours and academic discussions would need to shed light on, and some research has already started to address the issue: Nakatsuhara (2013) argued that joint scoring would be unfair to those who keep inviting quiet partners to the interaction but still fail to involve them, ending up with an asymmetrical interaction.

**Technological advances**

The use of technology in the assessment of speaking risks fundamentally changing the underlying test construct. In contrast to the co-constructed nature of interactional speaking tests discussed at the beginning of the article, computer-delivered speaking tests are unidirectional and lack the element of co-construction. Performance is typically elicited through technology-mediated prompts, and the conversation has a predetermined course, which the test taker typically has no influence on. As such, computer-delivered speaking tests draw on a psycholinguistic definition of the speaking construct, which places emphasis on the cognitive dimension of speaking and monologue speech.

Current developments in speaking assessment and technology are providing examples that preserve the construct of interactional competence through the use of technology, such as video conferencing systems (e.g., Skype and Facetime), which involve internet-based communication between speakers who do not have physical proximity. A technology-mediated remote face-to-face test preserves the interactional co-constructed nature of face-to-face speaking tests while at the same time offering the practical advantage of connecting test takers and examiners who could be continents apart. As such, it reduces some of the practical difficulties of face-to-face tests while preserving the interactional advantage. Computer-mediated interactional competence poses a new challenge for testers, however: what are the implications for construct definition and operationalisation? Some empirical efforts are underway to investigate the nature of interaction in a remote test setting (e.g., through video conferencing systems, such as Skype or Facetime), and they promise to provide some useful insights and initiate a discussion focusing on innovative ways of using technology in oral assessment and the implications for the construct of interactional competence (Davis, Timpe-Laughlin, Gu, & Ockey, 2016; Nakatsuhara, Inoue, Berry, & Galaczi, 2017). The “Online Interaction” scales recently added to the CEFR (Council of Europe, 2017) also hold useful insights about defining a new technology-enhanced construct of interaction.

That is not the only technological challenge. Another challenge arises from the rapid developments that are occurring in Automated Speech Recognition (ASR) systems and their capacity to discern and measure features of talk, as seen, for example, in the Pearson Test of English Academic and the Educational Testing Service (ETS) Speech-rater (Bernstein, van Moere, & Cheng, 2010; van Moere, 2012; Xi, 2010; Xi, Higgins, Zechner, & Williamson, 2012). A further narrowing down of the psycho-linguistic construct is seen in automated speaking tests that are both delivered and scored by computer and that aim to tap into “facility in L2” (Bernstein et al., 2010, p. 356) and “mechanical” language skills (van Moere, 2010, p. 93) (i.e., core linguistic knowledge that every speaker of a language has mastery of and that is independent of the domain of use).

It is worth asking the question: which features of interactional competence can be meaningfully rated by a machine and which features can be at least measured automatically to inform a human  

---

1We draw a distinction here between a speaking test where the computer delivers a set of prompts to the test taker and a computer-based speaking test that allows an examiner and one or more test takers to interact together online.
judgement on quality and interactional competence? Perhaps prosodic patterns (measured by machine) could demonstrate something about interactional competence, because they might show elements of flow and confluence. Work in corpus linguistics on interactional features that signal mutuality in spoken interactions (e.g., Evison et al., 2007; Tao, 2003) could provide potentially promising interactional features that build reciprocity and create “confluence” in interaction (McCarthy, 2010), and that could be captured in automated systems. The automated evaluation of facial expressions or gestures (as explored by Chen, Leong, Feng, & Lee, 2014) could also be a development as automated assessment approaches to speaking continue to evolve. Thus, automated assessment tools could be used in partnership with human raters to track and evaluate the more complex and sophisticated features of interactional competence.

Looking to the future in exploring interactional competence

We hope to have shown in our discussion that as far as assessing interactional competence is concerned, there is scope to expand the theoretical definition of the construct. There is also scope to analyse the nature of computer-mediated interactions, given the rapid growth of social media. There is certainly scope for developing and optimising automated assessment systems. And there is also a strong case for increasing interdisciplinarity in research: cross-cultural studies, computational linguistics, linguistics ethnography, behavioural and linguistic studies in professional and organisational domains, all have a strong contribution to make.

A systematic investigation of the questions and caveats discussed in this article holds implications not just for theoretical definitions but also for the practical pursuits of test and scale development, because it could contribute to a more accurate understanding of interactional speaking tests and the design of appropriate and meaningful assessment scales. Test takers are also learners, and a more precise delineation of interactional competence would bring benefits for teachers and learners, because it would provide guidance for the development of interactional skills in a communicative classroom. This is especially important now, because technology is making a mark in language learning, often offering attractive user experiences, but dated pedagogies that prioritise an atomistic and receptive view of learning, as opposed to a constructivist approach to learning in which interaction is fundamental.

To paraphrase Bachman (2000), is everything that counts in interaction accounted for in our existing speaking tests? We need speaking tests that tap comprehensively into a rich construct of interactional competence. How well do our existing speaking tests do this, and how might they do it better?

ORCID

Evelina Galaczi http://orcid.org/0000-0001-8269-8462
Lynda Taylor http://orcid.org/0000-0003-1366-0254

References


