

Reconceptualising Interactional Competence for Language Testing

David Wei Dai, UCL Institute of Education, University College London,

david.dai@ucl.ac.uk

<https://orcid.org/0000-0002-3575-131X>

Carsten Roever, School of Languages and Linguistics, University of Melbourne,

carsten@unimelb.edu.au

<https://orcid.org/0000-0002-1055-6331>

This is an author-produced PDF of a book chapter accepted for publication in *Assessing speaking in context: Expanding the construct and its applications*. R. Salaberry & R. Burch (Eds.). Multilingual Matters, following peer review. The page numbers of this version are formatted to match the ones of the definitive publisher-authenticated version, which can be found at: <https://multilingual-matters.com/page/detail/Assessing-Speaking-in-Context/?k=9781788923804>

To cite (APA): Roever, C., & Dai, D.W. (2021). Reconceptualizing interactional competence for language testing. In R. Salaberry & R. Burch (Eds.), *Assessing speaking in context: Expanding the construct and its applications* (pp. 23–49). Multilingual Matters.

2 Reconceptualising Interactional Competence for Language Testing

Carsten Roever and David Wei Dai

There has recently been strong interest in the assessment of interactional competence (IC) as witnessed by a special issue of the journal *Language Testing* (Plough et al., 2018) devoted entirely to IC, a special issue on the employment of conversation analysis in assessing IC of the journal *Papers in Language Testing Assessment* (Youn & Burch, 2020) and the prominent role of IC assessment in the special issue of the journal *Language Assessment Quarterly* on speaking assessment (Lim, 2018). However, to this date no major language test assesses IC, and this chapter is intended to contribute to ending this lamentable state of affairs. We will do so by discussing why the absence of IC assessment is problematic, why simply assessing proficiency does not give information about test-takers' IC, what particular challenges IC test designers face and how the IC construct can be broadened to incorporate social role enactment as a rating criterion via membership categorisation analysis (MCA).

1 Why Test IC?

In the widely used argument-based approach to assessment (Kane, 2006, 2012), the purpose of a test is to generate desirable consequences or, more specifically, to provide information in the form of scores about the strength of an attribute of interest in test-takers, such as language skills, to enable decisions about these test-takers (see Chapelle et al., 2008, 2010, for examples). In the case of language tests, decisions informed by language test scores might include such high-stakes decisions as admission to a foreign university, suitability for practising as a medical professional or permission to settle permanently in the host country. A test is arguably doing its job well if it enables good decisions, e.g. foreign medical graduates really do have the necessary language skills to communicate with patients and fellow professionals.

In the overwhelmingly typical case where test-takers' real-world language use involves interacting with others, it seems clear that their ability to do so should be a core part of the information gathered on their language ability. Language is a tool to make private thoughts public and visible (audible) to others, obtain access to their private thoughts and thereby enable coordinated social actions. Language users' ability to deploy language for accomplishing social actions has been conceptualised as their interactional competence, defined by Hall and Pekarek Doehler (2011: 2) as the 'ability to accomplish meaningful social actions, to respond to co-participants' previous actions and to make recognisable for others what our actions are and how these relate to their own actions'. This includes behaviours by which social roles are enacted in a given context, context-specific ways of organising turn-taking and communicative practices, as well as the use of linguistic and non-linguistic resources to accomplish these goals. While writing and reading are also to some degree forms of interaction, the immediate coordination of social actions between people on a moment-by-moment basis relies most strongly on speaking and listening. Language users' social actions are specifically designed with regard to their recipient (Drew, 2013) and every utterance accomplishes a social action which provides a context for the interlocutor's subsequent actions (Heritage, 1984): this is the core meaning of inter-action.

It seems logical that language tests should mirror this interactive use of language in order to obtain a representative picture of what the user can do in the real world and allow extrapolation from the sample of language use situations in the test to real-world language use. Alas, they do not. Language tests frequently do not include interaction in a second language and, where they do, they do not assess it, instead just using the resulting talk as a language sample to be rated on non-interactive criteria. For example, TOEFL and PTE contain monologic speaking tasks where test-takers react to prompts and input materials without engaging with an interlocutor. Tests like the ACTFL OPI and the IELTS speaking test involve a live interlocutor but their rating scales do not include measure of interactional abilities, and their main purpose is simply the elicitation of samples of spoken language. Various empirical studies have demonstrated the non-equivalence of interview-based speaking tests and natural conversation (e.g. Johnson, 2001; Lazaraton, 1992; van Lier, 1989).

The Cambridge English scales for assessing speaking performance (Cambridge English, n.d.) used for the Cambridge main suite exams go a step further and contain a rubric for interactive communication. However, this rubric is strongly influenced by the test format and focuses on maintaining interaction, responding to the interlocutor and linking contributions to the interlocutor's, thereby likely abbreviating the construct of interaction. Finally, the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001, 2018) has a number of scales

for different interactive activities (casual conversation, interview, negotiation) and 'interaction strategies' (taking the floor, ensuring understanding, repair), but pays little attention to social action implementation and social role enactment. Tests of languages other than English, such as the Test of German as a Foreign Language (TestDaF) or the Chinese Standard Exam (Hanyu Shuiping Kaoshi, HSK) also do not score IC.

This lack of attention to IC means that most language tests privilege speaking over talking, to use Roever and Kasper's (2018) parlance. Roever and Kasper (2018) view talking as interactive language use, including designing utterances for a specific interlocutor and comprehending implied social actions. By contrast, speaking is simply a monologic response to a stimulus not designed to achieve a social action vis-à-vis an interlocutor. While people usually talk to others and rarely just speak to nobody, the latter is exactly what the vast majority of language tests assess.

Given that large-scale, high-stakes language tests do not assess what people use language for, there is a serious risk that their results are flawed and that construct underrepresentation threatens the defensibility of decisions and conclusions based on scores (Messick, 1989). There would be no problem if speaking were the same as talking, i.e. if people who are good at speaking are invariably good at talking, and people who are good at talking are invariably good at speaking. We will start our deliberations by showing that speaking is in fact not the same as talking, and then we will discuss ways of making language tests more representative of what people do with language.

2 Are Talking and Speaking the Same?

Speaking ability following the classic model by Levelt (1989, 1999), which explicitly informs tests such as PTE, requires fast access to vocabulary, automatised of grammatical knowledge and high-speed phonetic encoding and articulation. This enables fluent, smooth, easily comprehensible speech, captured well in the construct of facility in L2, which Bernstein et al. (2010: 356) freely admit 'provides a measure of performance with the language without reference to any specific domain of use'. Similarly, the IELTS band descriptors (IDP IELTS, 2021) for the speaking test rate test-taker performance in four categories: fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation. There is no mention of interactional abilities, even though IELTS speaking involves a face-to-face interview, and research exists on the international construction of this interview (Brown, 2003; Seedhouse & Egbert, 2006; Seedhouse & Nakatsuhara, 2018).

Roever and Kasper (2018) term this the psycholinguistic view of speaking, and it sees speaking ability as consisting of a set of components that can be assessed without reference to their conversational use. Still, it assumes that such an assessment will provide useful information since

these abilities are required across a wide variety of contexts. This is probably not entirely unreasonable: if speakers have ready access to vocabulary and can implement a broad range of grammatical functions under real-time conditions, this will help them under any circumstances, be it a conversation in a pub, a job interview or a classroom discussion.

However, this view is akin to assessing driving by having a candidate drive alone on a closed course, and then assuming that they will do equally well in rush hour traffic. This is clearly a daring assumption: while the basic skills needed for successful driving (such as accelerating, braking, steering) are called upon in both situations, driving in rush hour traffic goes beyond basic driving skills. It requires coordinating one's actions with others, predicting what they might do, reacting to their actions and adapting to changing conditions.

While driving on a closed course is like speaking, driving in rush hour traffic is like talking: you need to be able to speak in order to talk (although not in all circumstances, as we will elaborate on below), but talking requires more than speaking. In addition to basic speaking skills, talking also requires understanding of the interlocutor, and adjusting of one's speaking to the interlocutor. Both are crucial, so let us take them in turn.

Understanding of the interlocutor involves real-time decoding of incoming language on a semantic level of word meaning, as well as on a pragmatic level of social action and interpersonal meaning. To be able to respond, interlocutors must understand the content of what is being said, but to be able to respond appropriately they must also understand what action is being done and how the utterance frames the relationship between the interactants. All of this happens before the interlocutor has even finished speaking, as otherwise there would simply not be enough time, given that assembling a response takes about 0.6 seconds but gaps between turns in English-language conversation are only about 0.2 seconds (Levinson & Torreira, 2015).

Knowing what social action is being performed is crucial for designing a response, as social actions tend to only have a limited range of typical response. For example, in responding to the informal greeting, 'How is it going?', a response like 'Not bad, you?' is quite typical because it consists of a second pair part commonly associated with this type of greeting. However, an account of one's health status ('How is it going?' – 'I've had really bad hay fever recently.') is atypical, as it recasts the social action of greeting as an inquiry after well-being. Finally, a response like 'I'm going by train.' indicates a semantic lack of understanding and would most likely be taken as a mis-hearing, which may be repaired or ignored.

In addition to responding to a social action with a type-fitting action, utterances must be recipient-designed. In the CA literature (e.g. Drew, 2013), recipient design refers mostly to taking into account shared knowledge between interlocutors. For example, when a speaker talks about their

partner, Mike, to a stranger who can be assumed not to know Mike, they are likely to refer to Mike as 'my partner'. However, when they talk to a close friend who knows Mike well, they are likely to refer to him as 'Mike'; in fact, referring to him as 'my partner' would be decidedly odd.

Talk can be specifically designed for a recipient in other ways as well. A great deal of research in pragmatics, based in speech act theory (Austin, 1962; Searle, 1969) and taking off from Brown and Levinson's (1987) seminal work on the social context of interaction, has focused on social factors that impact how people talk to each other, with most research focusing on politeness in making requests. Brown and Levinson posit three major factors that influence the politeness level of utterances: Power, Social Distance and Degree of Imposition. Power refers to the degree that one interlocutor can exert control over the other one's behaviour; e.g. in a workplace situation, a manager would have power over an employee they supervise. Social Distance is the degree of acquaintanceship or common membership in a social group between interlocutors; e.g. close friends have low social distance whereas strangers have high social distance. Finally, the Degree of Imposition describes the 'cost' to the hearer in terms of money, time, effort or social sanction for complying with the speaker. For example, asking someone for the time is a low imposition request, but asking them to go out of their way to help carry a heavy item is high imposition. It is worth noting that these factors do not deterministically govern how people talk to each other; other factors come into play as well. For example, Curl and Drew (2008), taking a CA approach, showed that the perceived degree of entitlement for making a request in emergency calls strongly affects request formulation.

Finally, work in sociolinguistics has identified a number of other factors that impact talking, most famously encapsulated in Hymes's (1974) SPEAKING model, which takes into account (among others) the physical setting of the interaction, its overall tone, the channel through which it is conducted and the cultural norms through which interlocutors make judgments about meanings.

Given the numerous additional factors that impact talking as opposed to speaking, it is difficult to imagine that simply measuring speaking would give testers a strong basis for making inferences about the ability to talk. However, this is really an empirical question. If strong speaking ability invariably leads to strong talking ability, it is sufficient to test speaking and then simply infer talking. This may seem unlikely, and speaking will probably not account for all the variance in talking, but even if it just accounts for a large amount of variance, that may be sufficient for a test. However, if a dissociation between speaking and talking can be demonstrated, where you can be good at speaking but not talking and vice versa, talking would need to be tested separately to enable defensible inferences from test scores.

2.1 Can you be good at speaking but not talking?

It certainly appears that way, as anecdotal evidence abounds about test-takers who do well in the testing situation but not so well in real-life interaction, e.g. in medical communication (Eggly et al., 1999; Hall et al., 2004). In fact, end-user complaints about apparent mismatches between candidates' scores on the Occupational English Test (OET) taken by medical professionals and those test-takers' real-world performance were a major impetus in the revision of the OET to include a stronger focus on role-appropriate interaction in the rating criteria (Pill, 2016). This divergence is likely due to tests being rated on language-focused criteria, which tend to privilege speaking, and not criteria indigenous to users (see Jacoby, 1998), which tend to privilege talking. Sato (2014) supports this conclusion, showing that different aspects of speaking performance matter to naïve judges assessing a performance on their general impression of the test-taker as a skilled communicator, as compared to trained raters assessing on language-focused criteria. This dissociation between the criteria used in language tests and real-world language requirements also likely accounts for the less-than-overwhelming confidence of students, academics and employers in the predictive value of language tests for real-world performance (see Murray et al., 2014, for IELTS; Malone & Montee, 2014, for TOEFL).

It appears that speaking ability as measured by such a test is not a good predictor of the ability to talk in the real world, which does not bode well for language tests. However, two counter-arguments to this line of reasoning could be adduced to defend general language tests: first, it could be claimed that specific-purpose language tests like the OET not only require a general ability to interact but a role-specific ability to interact. To put it simply, not only do you have to talk well, you have to talk recognisably like a doctor, nurse, dentist, etc. A potential argument is that this ability is not required in general proficiency tests. Second, it could be argued that no test performance ever extrapolates perfectly to real-world performance, and so the apparent gap between speaking in a test and talking in the real world is simply an unavoidable gap between eliciting performance in a controlled test environment and performing 'in the wild'. In our view, both points are akin to desperate rear-guard battles trying to stave off the inevitable loss.

Role-specific interactional abilities are always required in talking. There is no such thing as language use unbounded to a social role. Interlocutors are always speaking as a friend, colleague, supervisor, customer, partner, neighbour, student, with all the social requirements of talking appropriately to that role. These requirements can be subtle: Bella (2014) found that although advanced learners of modern Greek had control of the pragmalinguistic tools for performing refusals in roleplays, they overused some and underused others when compared to native speakers. In other words, just because you have grammar and vocabulary that you can deploy in speaking

does not mean you can deploy it in conventional ways when talking from the perspective of a particular social role. This is more obvious in specific- purpose assessment, but no different in other language tests.

While language use in tests is not the same as language use in the real world, strengthening the extrapolation inference should be the main mission of language testers. O’Sullivan (2019) lays out the problem very clearly, and the larger the gap between test performance and real-world performance (i.e. the weaker the extrapolation inference in Kane’s, 2006, framework), the less value test scores have for end-users because they do not enable the decisions for which end-users need them. Tests based on a universe of generalisation that is mostly chosen to be practically measurable will not do well on extrapolation. This is precisely the issue with measuring speaking versus measuring talking: in particular, tests that use monologic measures and thereby measure speaking have an advantage in terms of administration, standardisation and scoring, but they do not measure the skills associated with talking, such as designing talk for the recipient, responding, organising talk sequentially, enacting befitting social roles, etc. Similarly, tests that simply use interaction to elicit ratable samples of language to be rated without any reference to interactional abilities sell the talking construct short. Scores from these tests run precisely the risk of being mismatched with end-user impressions of ability, which can eventually bring down the whole language testing enterprise: Why should end-users go through the trouble of obtaining test scores if these scores do not tell them what they want to know? This is a broader question than just speaking versus talking, but because of the pervasiveness and everyday necessity of being able to talk, speaking versus talking is a particularly important aspect of this problem in language testing.

2.2 Can you be good at talking but not speaking?

It may seem counter-intuitive that a language user could be better at talking than speaking since we have so far portrayed talking as ‘speaking plus’: interactants need basic speaking skills, but to talk they need to be able to configure and deploy them within a particular physical, social and interactional context. However, it is possible to talk with little or even no language proficiency. Levinson (2006) describes an interaction during field research with a deaf signer on Rossel Island, an island in the far south-east of Papua New Guinea. Levinson and his interlocutor shared no language and little background knowledge but successfully managed a storytelling. While this may seem like an extreme case, it is actually fairly mundane: anybody who has travelled to a country without speaking the local language knows that it is possible to communicate to some extent through pointing, miming and gesturing. This is not sufficient for discussing abstract, complex topics, but it demonstrates that talking in the sense of communicating is possible with very little speaking proficiency.

A case closer to the experience of applied linguists is the well-known case study of Wes (Schmidt, 1983). Wes is a native Japanese speaker who migrated to the United States as an adult, and Schmidt collected tape recordings and field notes of Wes's use of English over several years. He found little and stagnating grammatical development, but rapid development and high levels of performance in spoken discourse. Wes had an active social life with many friends and managed everyday interactions in English successfully. While Schmidt described Wes's language ability in order to test Schumann's acculturation theory (1978), the disjunct he found between grammatical ability and interactional/sociolinguistic abilities is a good illustration of a highly interactionally competent language user with low grammatical ability. Schmidt writes: 'If language is seen as a means of initiating, maintaining, and regulating relationships and carrying on the business of living, then perhaps Wes is a good learner. If one views language as a system of elements and rules, with syntax playing a major role, then Wes is clearly a very poor learner' (Schmidt, 1983: 164). Wes was good at talking but not so good at speaking, and it is probably safe to say that he would not have performed well on formal language tests where his lack of grammatical accuracy and limited vocabulary range would have likely been penalised.

To conclude, the above discussion offers theoretical and conceptual evidence that speaking and talking are two overlapping but still distinctive forms of competence. This lends support to assessing talking/IC in its own right in testing settings so as to better gauge test-takers' ability to interact in real life. But what are some of the caveats in talking/IC assessment?

3 How Can Talking Be Tested Better?

The second part of this chapter focuses on four practical concerns in operationalising a 'talking'/IC construct in testing settings. The first three concerns are target domain delineation, IC construct validity and IC marker selection and the last one is social role enactment. Although previous IC developmental and assessment studies have interrogated the first three concerns to varying degrees, we argue that these concerns warrant more thorough deliberation if test designers want to develop a defensible interpretative framework for an IC construct. The last concern, social role enactment, is unexplored in existing IC research and we aim to demonstrate its relevance in an IC construct with the analytical toolkit of MCA.

3.1 Target domain delineation

Galaczi and Taylor (2018: 227) offer a tree-shaped visual representation of IC which is helpful in framing our discussion of IC assessment. Their IC tree grows into branches and leaves, with branches indexing general IC markers such as 'turn management' and 'interactive listening',

and leaves indexing finer IC markers such as ‘maintaining turns’ and ‘pausing’, under the branch of turn management. Going downwards, the authors specify the roots for the IC tree in three concentric circles: speech acts (or social actions in CA parlance), speech events and speech situations. Although such contextual factors situate, nourish and support the investigation of IC, they are rarely given the consideration they deserve. In terms of the first two circles, although previous L2 IC research has examined a range of social actions such as requests (Al-Gahtani & Roever, 2015), refusals (Al-Gahtani & Roever, 2018), storytelling (Waring, 2013; Watanabe, 2016) and disagreements (Pekarek Doehler & Pochon-Berger, 2011), the rationale for researchers’ choices of social actions is rarely provided. Understandably for developmental IC studies, social actions merely serve as vehicles for analysts to observe differential IC, so what social action is used to elicit L2 speakers’ performances is a minor concern. However, from a test design perspective, which social action to include in or exclude from the target domain needs to be a carefully considered and empirically justifiable decision. For example, there needs to be a rationale for testing L2 Arabic speakers’ ability in launching a request, instead of a disagreement or a complaint.

When we move beyond speech acts and speech events in Galaczi and Taylor’s (2018) IC tree, we encounter the largest circle, speech situations where social actions take place. The authors’ conceptualisation of speech situations draws from Hymes (1974) where speech situations are decidedly sociocultural. This further complicates the picture for IC assessment. Cross-linguistic CA research has offered evidence that speakers adopt different methods and sequentially structure their talk differently across languages and cultures when they conduct social actions (Golato, 2002; Huth, 2006). Developmental L2 IC studies have also noted that L2 speakers increasingly diversify their methods to adapt to and align with routinised interactional patterns commonly found in the host culture (Cekaite, 2007; Pekarek Doehler & Pochon-Berger, 2015). The cultural specificity of social action implementation requires test designers to go one step further: they need to demonstrate the relevance of their chosen social actions to the particular L2 context. In other words, the question now is not just choosing between requests, disagreements or complaints for an unspecified L2 speaker population, but it is to decide between these social actions for an IC test specifically targeting, for example, L2 Arabic speakers. Why should requests be singled out in an L2 Arabic test? Is it because requests in Arabic are particularly frequent, complex or difficult? Or do L2 Arabic speakers struggle with requests in particular due to some sequential, interactional or sociocultural differences in how requests are formatted in the host community? Fortunately, there is a ready remedy for this problem: needs analysis.

Needs analyses can assist L2 IC test designers in their depiction of the IC target domain and identify social actions that are most pertinent to or

challenging for their specific test-taker population. Youn (2015) followed this procedure by grounding the design of her IC test in a needs analysis on the English for academic purposes (EAP) domain. Her needs analysis, later published as Youn (2018), triangulated the perspectives from programme administrators, instructors and students and utilised two data elicitation methods: interviews and questionnaires. Her data revealed a wide range of social actions where EAP students struggle, which then fed into the two interactional tasks she designed. The two tasks in Youn (2015) require test-takers to roleplay with a professor and a classmate and elicit actions such as making a request for a recommendation letter and agreeing on a meeting time. In a similar vein, Dai (2019) conducted a task-based needs analysis of the interactional needs of L2 Chinese speakers for the design of his IC test. Utilising Socratic-Hermeneutic interviews (Dinkins, 2005) and follow-up written communication and triangulating the perspectives from L2 Chinese speakers, Chinese teachers and native speakers who frequently interact with L2 speakers, Dai elicited rich qualitative data on where L2 Chinese speakers encounter problems when interacting in Chinese in China. Dai then designed nine roleplay tasks targeting the top-ranking social actions that L2 Chinese speakers struggle with the most (Dai & Roever, 2019). To conclude, although every social action is worth investigating from a developmental perspective, due to the limited resources in testing settings, test designers need to define their target domain clearly and select social actions that are most germane to their respective L2 groups. A methodic needs analysis serves this purpose and can assist test designers in narrowing down the target domain to items in a test.

3.2 IC construct validity

The second concern links back to our previous discussion on differentiating talking from speaking. If an IC construct claims to cover ‘talking’ variance not already covered by existing tests that measure ‘speaking’ or general proficiency, it needs to be able to demonstrate it in a statistical sense. Having this goal in mind and then looking at existing research in L2 IC, the issue of differentiating between IC and speaking, or proficiency in general, quickly becomes a chicken-and-egg question. Longitudinal studies on L2 IC development document L2 speakers’ changing methods or interactional patterns but it is difficult to tease apart how much of those changes are attributable to increase in IC or increase in general proficiency (Hellermann, 2007, 2008; Pekarek Doehler & Pochon-Berger, 2016). Cross-sectional studies, on the other hand, start with pre-grouping L2 speakers by proficiency and investigate if speakers from different proficiency levels mobilise different methods in implementing the same social action (Al-Gahtani & Roever, 2018; Pekarek Doehler, 2018; Pekarek Doehler & Pochon-Berger, 2011). Their grouping criteria are either proficiency frameworks based largely on the speaking construct (e.g. CEFR)

or researchers' intuition. The few IC assessment studies that exist also adopt a cross-sectional design and although they provide statistical evidence that L2 speakers do differ on researchers' IC measures, we cannot know for sure if proficiency or 'speaking variance' can already account for such IC differences (Galaczi, 2013; Ikeda, 2017; Youn, 2015).

Nevertheless, there is some incipient evidence that IC does measure unique variance not encompassed by speaking, or a psycholinguistic conceptualisation of proficiency. Lee and Hellermann (2014) offer cross-sectional data where a low-proficiency speaker, Larissa, demonstrates the capacity to launch a storytelling sequence despite her lack of linguistic resources. In their data Larissa, who is from Russia, is conversing with another low-proficiency speaker Jamie, who is from Mexico, and before the turn that launches the storytelling sequence in focus, a sequence on Thanksgiving was completed with an agreement token by Larissa. After a 2.5-second gap, Larissa self-selects, and launches her storytelling sequence with 'mm husband #uh::# call my uncle.' (Lee & Hellermann (2014: 772). Lee and Hellermann argue that although Larissa's storytelling announcement turn lacks explicit time reference devices such as 'yesterday', which are common in high-proficiency L2 speakers' talk, she still designs her turn to adumbrate a forthcoming story. Accompanying this turn, Larissa employs body language mimicking a person making a call, suggesting that her husband is calling her uncle to talk about the American tradition of having turkeys at Thanksgiving. It is also worth noting that in this turn Larissa evokes family categories 'husband' and 'uncle' which, combined with other linguistic and paralinguistic resources, pave the way for inferences such as 'husband is living with Larissa in USA', 'husband is calling Larissa's uncle to tell the story of American Thanksgiving' and 'uncle is most likely still living in Russia'. These inferences are crucial to Larissa's storytelling in subsequent turns. Therefore, despite Larissa's linguistic inadequacies, she is still interactionally competent in securing a storytelling sequence through the mobilisation of turn design, body language and category evocation. Regrettably, these markers of interactional resourcefulness are not covered in any existing speaking rubric. In summary, developmental IC studies have offered fruitful insight into how L2 speakers differ in different facets of talking, but we need more evidence to prove that such differences can add variance to our existing speaking tests.

Different from the qualitative evidence in Lee and Hellermann (2014), Ockey et al. (2015) is a rare testing study that explicitly compares IC measures with proficiency measures. The researchers compare 222 Japanese university students' TOEFL iBT performances with their performance on three language tasks (group discussion, picture description and oral presentation). One component score for the three language tasks is IC, which is defined as 'participation and smoothness of interaction (e.g. turn-taking, responding to others, asking questions, introducing new gambits, paraphrasing, and hedging)' (Ockey et al., 2015: 46).

Pearson correlations reveal high correlations between test-takers' TOEFL scores and their component scores on pronunciation, fluency and vocabulary in the three tasks. However, only moderate correlation is achieved between TOEFL scores and their IC component scores. Such a finding offers evidence that TOEFL, a psycholinguistically grounded speaking test, can offer good prediction on real-life performances in terms of pronunciation and fluency but not IC. It also shows that the asynchronous monologic speaking tasks employed by TOEFL can barely cover basic IC variance such as turn-turning or hedging. We speculate that higher-order talking/IC sub-traits such as action formation and social role enactment are even harder to predict from existing speaking constructs and rubrics, but such speculations are only tentative until corroborated by empirical evidence.

When piloting his L2 Chinese IC test, Dai also garnered emergent quantitative evidence supporting the separation between talking competence/IC and speaking competence/proficiency. Dai piloted his nine-item test on 22 test-takers, comprising 11 native speakers (NSs) and 11 non-native speakers (NNSs) of differing proficiency in Chinese. Three L1-Chinese raters provided intuitive IC rating on all 198 performances (9 items*22 test-takers), with 'successful interaction' coded as 3, 'average interaction' coded as 2 and 'unsuccessful interaction' coded as 1 in the IC ability measure.

Rating results were analysed via many-facet Rasch measurement, and fair scores (which correct for rater severity and item difficulty) are pre-presented in Table 2.1. Although NSs dominate the higher end of the scale, many NSs did not outperform NNSs in the middle range of the scale and one NS (ID 13) even scored below all 11 NNSs. Within the middle range, it is also telling to note that test-taker 12 is a beginner-level NNS, overtaking not only five NSs but also eight NNSs of much higher Chinese proficiency. Such findings challenge our conventional wisdom that NSs and highly proficient NNSs are invariably better at interaction than lower-proficiency NNSs, which lends preliminary support to IC being a distinctive construct from the traditional proficiency construct.

3.3 IC marker selection

The third concern is the selection of IC markers, which can offer backing to both scoring and explanation inferences in Kane's framework. Drawing on copious findings from CA, developmental IC studies have offered a wide array of potential IC markers such as progressivity (Balaman & Sert, 2017), alignment (Dings, 2014), post-expansion (Greer, 2016), dispreference structure (Al-Gahtani & Roever, 2012) and recipient design (Al-Gahtani & Roever, 2018). However, the choice of IC markers, just like the choice of social actions, should be a purposeful one from a test operationalisation perspective. The markers selected should be

Table 2.1 NS and NNS test-takers in Dai's (2019) study

ID	Group	Score
5	NS	2.99
4	NS	2.97
8	NS	2.90
7	NS	2.73
18	NS	2.59
2	NS	2.55
3	NNS	2.55
6	NNS	2.47
12	NNS	2.43
14	NNS	2.36
16	NS	2.36
11	NNS	2.28
20	NS	2.28
19	NNS	2.20
9	NNS	2.16
10	NS	2.16
15	NNS	2.08
21	NNS	2.04
1	NNS	1.95
22	NS	1.91
17	NNS	1.60
13	NS	1.34

consistently observable, ratable and scalable in test-taker performances. They should be those aspects of human interaction that make a speaker like Wes (Schmidt, 1983) interactionally competent despite their linguistic deficiencies. In terms of the scoring inference, IC markers and their rubrics are what translate test-taker performances into numbers, and a poor choice of markers can potentially mislead raters, directing their attention to aspects of talk not related to the IC construct. In terms of the explanation inference, IC markers carry the responsibility of construct validation explicated in the previous section as it is these markers that cover the added variance that existing speaking tests might fail to capture. Finally, due to the constraints of assessment, only a limited number of markers can be selected and incorporated into assessment rubrics. Therefore, the privileging of, for example, topic management over turn-taking, needs to have empirical support. Test designers should hence select markers that cover greater IC variance and that make a more tangible impact on the performance of talking.

One approach to purposeful marker selection is to combine etic, researcher-based judgment with emic, data-based evidence. Researchers can develop indigenous criteria through eliciting target group members' assessment on sample test-taker performances (Dai, 2020; Elder & McNamara, 2016; Knoch & Macqueen, 2019; Pill, 2016). Let us take the two added criteria, 'clinician engagement' and 'management of interaction' in OET, as an example. Pill (2016) developed both criteria by first obtaining medical educators' and clinical supervisors' judgments on simulated clinical performances from trainee healthcare professionals. He then conducted a thematic analysis on the data and generated these two new criteria, which were not covered by the previous OET rubric. Although Pill did not employ CA, it is evident that CA can be productively applied to the analysis of the 'management of interaction' criterion to collect micro-level evidence of how interaction is co-constructed. 'Clinical engagement', on the other hand, requires the ability to talk in a manner befitting the role of a doctor, a physiotherapist or a nurse. In other words, medical professionals' language should evoke their respective professional categories in relation to their interlocutors. Sequential CA, with its predominant concern for sequential properties, is clearly limited in collecting such information. This brings us to our last concern in IC assessment: social role enactment and MCA.

3.4 Social role enactment

As discussed earlier, there is no language use unbounded to social roles. In EAP contexts test-takers talk as a student or a classmate, whereas in OET contexts they talk as a doctor, a nurse or a physiotherapist. However, the ability to perform a particular social role is largely ignored by existing language tests, with OET being a rare example. A recent study on domain experts' indigenous criteria of L2 Chinese IC has identified role enactment as one of the five IC rating criteria that are most salient to and considered most crucial by domain experts (Dai, 2020). Here we argue that social role enactment should be foregrounded in our conceptualisation of IC and we should combine the analytic power of both CA and MCA to further our understanding of social role enactment as an IC marker.

Although MCA was developing rather slowly while CA flourished following Sack's seminal work (Sacks, 1992), we are now seeing a renaissance of interest in MCA (Stokoe, 2012), with a recent edited book on advances in MCA (Fitzgerald & Housley, 2015) and a special MCA section in the *Journal of Pragmatics* (Fitzgerald et al., 2017). This shows that CA analysts have acknowledged that, apart from the many sequential concerns of talk, there is also a categorical aspect of talk that is worthy of investigation. How speakers evoke categories such as doctors, nurses and students indexes their members' knowledge of social roles in their host society and culture. L2 developmental studies are yet to grapple with the affordances and challenges of MCA, although Lee and Hellermann (2014)

offer some nascent findings in this regard. MCA also has not made inroads into L2 IC assessment, as IC rubrics so far still focus solely on sequential markers such as turn-taking, topic development, adjacency pairs and back-channelling without situating such markers in test-takers' categorical knowledge (Galaczi, 2013; Ikeda, 2017; Youn, 2015). The last section of this chapter aims to make an exploratory attempt at demonstrating how MCA can be productively applied to the analysis of test-taker performances on the rating criterion, social role enactment.

4 MCA and Testing

Dai's (2019) needs analysis-grounded IC test of L2 Chinese explicitly assessed social role enactment as a marker of IC. Each item is delivered to test-takers in video format with three still images and a soundtrack of a scenario script recorded in Chinese. The English translation of the script for one such item is presented below and the three still images of this item are presented in Figures 2.1–2.3.

Video script in English:

You recently went to a different city for work for six months and sub-let the apartment you rented to Wang Hao's son Wang Bin. Wang Hao is your best friend and has helped you greatly over the years. You have also met Wang Bin before when you visited Wang Hao and have a good impression of him. Today the body corporation of your apartment building calls you, telling you that lately there has been a lot of noise and loud music in your apartment late at night. Sometimes the neighbours also see drunken youths coming in and out of your apartment. You want to discuss this with Wang Bin but since you are still away so you decide to talk to him via video chat.

As the prompt makes clear, Wang Bin (the interlocutor) is making too much noise late at night and causing annoyance to the test-taker's neighbours. The prompt does not prescribe what social actions the test-taker



Figure 2.1 The test-taker sub-letting his apartment to his best friend's son (interlocutor)

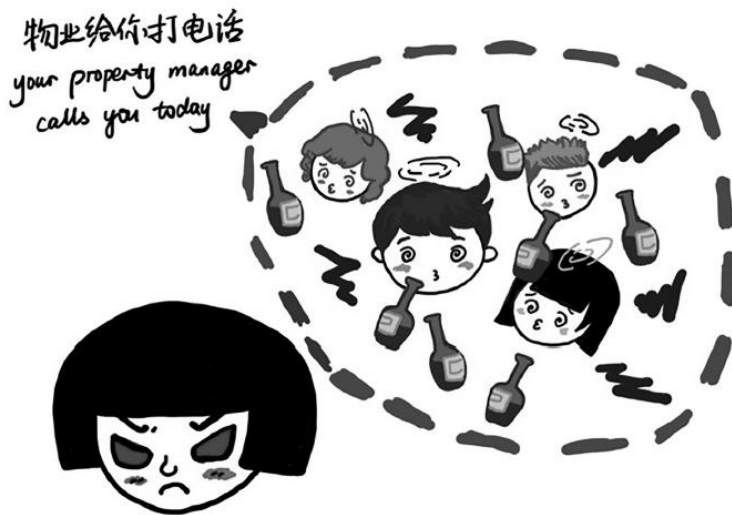


Figure 2.2 The test-taker receives a call from the building manager about noise in their apartment



Figure 2.3 The test-taker decides to have a video chat with the interlocutor about this incident

needs to implement but most test-takers orient to criticising the interlocutor. Despite unanimity on the social action, different test-takers construct, evoke and utilise different aspects of their categories vis-à-vis the categories of their interlocutor. Here we focus on two performances from intermediate L2 Chinese test-taker Brian (Test-taker ID 19 in Table 2.1) and beginner L2 Chinese test-taker Hans (Test-taker ID 12 in Table 2.1), each playing the role of the aggrieved apartment owner.

Brian: 'I want to be good to you'

Excerpt 2.1 is the start of the role play between Brian and the NS interlocutor.

Excerpt 2.1 Brian's greeting sequence

```

1  B      .tch (0.2) Wang Bin      [ni hǎo:
           (full name)      hello
           '.tch (0.2) Wang Bin  [hello'

2  I

           [Ai
           PRT
           ['Hi'

3  I → (.) Ai      Brian      shushu hao
           PRT      (first name)  uncle  good
           '(.) Hi Uncle Brian,hello'

4      (0.9)

5  B      .tch (0.2) .tch Zuijin zenme jiyang?
           Lately how condition
           '.tch (0.2) .tch how have you been? lately'

6      (0.4)

7  I      AAA: >zuijin haihao< jiu xuexi >shenme de<
           PRT lately pretty good just study what NOM
           'AHH: >Pretty good lately< just studying >and the sort<'

```

In Line 1, Brian (abbreviated as B in the transcript) launches a greeting sequence, addressing the interlocutor (I in the transcript) by the full name of his role, Wang Bin. The interlocutor overlaps with Brian's talk, produces an acknowledgement token in Line 2 and returns the greeting in Line 3, suggesting that Brian's greeting is formatted in an easily recognisable manner. What is of analytical interest here is that the interlocutor addresses Brian by using his English name, followed by a Chinese kinship term 'shushu' (uncle). On the interlocutor's prompt we do not specify how

the test-taker should be addressed, but here the NS interlocutor draws on his member's knowledge and decides to address Brian with a kinship term in situ and in vivo. 'Shushu' in Chinese is not reserved only for blood relatives but is an inference-rich address term that carries specific duties, obligations and expectations. Instead of resisting the interlocutor's attempt at categorising, Brian starts in Line 5 with a question enquiring about the interlocutor's well-being. Brian's implicit sanctioning of the category 'shushu' shows that the interlocutor's activity is permissible and pre-existent in both parties' shared category knowledge.¹ By establishing intersubjectivity, it also licenses Brian to exploit the inferences bestowed by the 'shushu' category.

Excerpt 2.2 Brian laying groundwork for criticising

- 1 B >NI ZHIDAO WO shi< (1.7) ni BA,
You know I be your father
'>YOU KNOW I am< (1.7) your DAD'
- 2 B jiu (.) >wo de hao pengyou<
just I ASSOC good friend
just (.) >is my good friend<
- 3 (0.4)
- 4 I Dui=dui=dui°
Yes yes yes
'Yes=yes=yes°'
- 5 (0.7)
- 6 B → Aaa::: (0.3) Suoyi (.) wo xiāng ↓DUI ni hao
PRT therefore I want towards you good
'Ahh::: (0.3) Therefore (.) I want to be good ↓TOWARDS you'
- 7 B (0.3)Wo ye ↑XI↑WANG ni: (0.1) ni ye:
I also hope you you also
'(0.3) I also ↑HOPE you: (0.1) you also'
- 8 (0.9)
- 9 B → dui WO: hao.
towards I good
'are good towards ME'
- 10 (0.4)
- 11 I Ye::: ↓en↑hen (0.3) ye?
PRT PRT PRT PRT
'Yes::: ↓um↑hum (0.3) Yeah?'

Excerpt 2.2 takes place a few turns after Excerpt 2.1 where Brian makes use of his 'shushu' category when he criticises the interlocutor. Brian first makes an explicit attempt at self-categorisation, abandons it, topicalises the test-taker's father, and establishes a standardised relational pair between test-taker's father and himself from Lines 1–2. A standardised relational pair is a pair of categories that carry mutual obligations and responsibilities (Stokoe, 2012), such as patient-doctor, salesperson-customer and friend-friend. By categorising the interlocutor's father as his good friend, Brian also self-categorises as a good friend to the interlocutor's father, as friendship cannot be claimed one-sided. Brian's relational pair and self-categorisation are quickly endorsed by the interlocutor with repeated acceptance tokens in Line 4. When we combine the categorisation work of 'shushu' in Excerpt 2.1 and the relational pair in Excerpt 2.2, we can see how three categories, father, son and uncle, emerge as duplicatively organised (Francis & Hart, 1997). This makes explicit a common-sensical practice in Chinese society: if a male Chinese speaker is on close terms with another male Chinese speaker of similar age, the son of one of the speakers is bound to call the other speaker 'shushu' (uncle), despite there being no blood connection between the son and the other speaker. Such knowledge is neither explicated in nor mandated by the scenario script for test-takers or the prompt for NS interlocutors. It is both parties' members' knowledge of their respective social roles that makes them co-construct this duplicative organisation.

Having examined the category-building work from both Brian and the interlocutor, now let us look at how Brian makes his laborious categorisation work for him. Categories are not void concepts. Sacks first noted that categories have activities bound to them, such as in his famous example, the baby (category) cried (activity) and the mommy (category) picked it up (activity) (Sacks, 1974). Subsequent work on MCA has further substantiated the activities (also called predicates) tied to categories, such as entitlements, duties and knowledge (Jayyusi, 1984; Payne, 1976; Watson, 1978). In Excerpt 2.2, after sanctioning his category as 'shushu' and the interlocutor's category as 'close friend's son', in Line 6 Brian makes explicit the obligations he perceives to have been engendered by their relational categories: 'I (shushu) wanted to be good towards you (close friend's son)' and 'I (shushu) hope you (close friend's son) are also good towards me (shushu)'. Note Brian's choice of verbs here. When he explicates his moral obligation as a 'shushu', he uses 'want' (xiang) in Line 6, a word indexing strong agency. When he describes the obligation of his friend's son, he chooses 'hope' (xi wang) in Line 7, a word that hedges the proposition. The word 'also' (ye) similarly offers a softening effect as it emphasises reciprocity.

In Line 11, the interlocutor first issues an affiliative token 'ye' with a downward intonation, licensing Brian's category inferences and acknowledging the mutual obligations that Brian purposely establishes. The subsequent rising and falling intonation of the particles 'en hen' is particularly

intriguing here. ‘En hen’ suggests that the interlocutor predicts that there is more to come from Brian as Brian’s exposition of mutual obligations is knowledge that already exists in their shared category knowledge. The fact that Brian has painstakingly spelt out such details of their category predicates could only implicate one thing: Brian wants to make use of such predicates. Indeed, although not reproduced here, Brian’s ‘I wanted to be good towards you’ later paves the way into his more attentive enquiries about what the interlocutor does at night, what kind of people he brings home and whether they have been drinking irresponsibly. Brian’s ‘I hope you will be good to me too’ provides grounds for Brian’s criticising of the interlocutor’s behaviour, as what the interlocutor does obviously is not ‘being good’ to ‘shushu’. Everything takes on a different hue after the meticulous category groundwork Brian has laid for his criticising action. Although what the interlocutor has done is irresponsible in itself (e.g. disturbing the neighbours), Brian adds ammunition to his criticism by evoking the moral obligations rooted in their categories. Therefore, what the interlocutor has done has thus become not just irresponsible, but also immoral.

‘Guai haizi’

Previous excerpts demonstrate that test-takers do not solely focus on launching the intended social action implied in the scenario video. They contextualise actions in a manner that is congruent to their social roles and carefully design their talk so that it is role appropriate. It should also be noted that role enactment is a joint process by both parties, as evidenced by the duplicative organisation ‘father-son-uncle’ in Brian’s example. However, as Brian has skilfully demonstrated, test-takers have the ability to utilise categorisation work to their advantage in an interaction-ally relevant fashion. The following excerpts feature a beginner-level test-taker Hans, who shows us that even low-proficiency L2 speakers have an awareness of role enactment and can mobilise category knowledge to their benefit, despite limited linguistic resources.

Excerpt 2.3 Hans (abbreviated as H in the transcript) categorising the interlocutor as a ‘guai haizi’

- 1 H Wo=wo=wo=wo xiang: ni shi
 I I I I think you be
 ‘I=I=I=I thought: you were’
- 2 H → yi ge guai hai (.) zì (.) suoyi
 one C good kid therefore
 ‘a good ki(.)d(.) therefore’
- 3 H (0.3)
- 4 H Anm:(.)yexu (0.1) e: (0.3) YOU shihou (0.3) juede (0.5)
 PRT perhaps PRT be time think
 ‘Hmm: (.)perhaps (0.1) Err: (0.3) There ARE times I think’

5 H E: gen pengyou he: (0.1) yi liang bei
PRT with friend drink one two C
'En:(.)you might have(0.1)one or two (0.3) drinks with friends'

6 H (0.3)En (0.6) suoyi (0.2) bu hui you wenti dè= buguo
PRT therefore N AUX be problem NOM but
'(0.3)Hmm (0.6)therefore (0.2) it wouldn't be any problems=but'

Line 1 in Excerpt 2.3 follows Hans's storytelling of him receiving a call from the neighbour saying that lately there has been a lot of noise in his apartment. Hans makes multiple attempts at starting the turn but gets stuck at the subject 'I' (wo). When he finally succeeds at the fourth attempt, he smoothly delivers his assessment of the interlocutor in Line 2, calling him a 'guai haizi' (good/obedient kid). It is worth noting that Hans categorises the interlocutor as a 'haizi' (kid), despite the fact that his 'friend's son' is old enough to host parties and drink recklessly. Hans calling his friend's son 'haizi' positions the interlocutor at the lower end of a hierarchical relational pair compared to Hans, as Hans is the 'grown-up' here while the interlocutor is the 'kid' (Stokoe, 2012). Therefore, when a 'haizi' misbehaves, a grown-up has sufficient moral grounds to criticise him, which is exactly what Hans's turn foreshadows.

Another interesting feature is the adjective 'guai' that Hans predicates on 'haizi'. 'Guai' is a cultural-specific description that does not lend itself well to translation. Semantically it falls between 'being good', 'being nice', 'being obedient', 'do not cause trouble' and 'do not contradict'. 'Guai' can only be used by a senior member on a junior member in Chinese society and the senior and junior members are related on an intimate personal level. For example, a younger person in Chinese society would not use 'guai' as a quality expected of a person older than them. An older person would also not randomly call out people younger than them as 'guai' as there is no entitlement for such an expectation. What licenses Hans's usage of 'guai' is the same duplicative father-son-uncle organisation explicated before. Similar to Brian's categorisation, 'guai' also carries a strong moral obligation as there is a preference for a category like 'haizi' to co-select with its bound predicates like 'guai' (Hester & Eglin, 1997). 'Guai' as a predicate is a cultural expectation of the 'haizi' category, which in this particular instance manifests as 'being nice to shushu', 'do not contradict shushu' and certainly 'do not drink irresponsibly and cause trouble for shushu'. Therefore, Hans's characterisation work is purposeful as now the interlocutor's behaviour is not just unsociable, but also flouts the moral obligations implicated in his category.

In terms of Hans's production of 'guai haizi', it is smoothly delivered in Line 2, despite a micropause between 'hai' and 'zi'. This contrasts greatly with the following Line 3–Line 6 where Hans struggles to describe what the interlocutor was doing. Considering Hans's beginner-learner proficiency level, it is understandable that he finds it challenging to

mobilise language in his recount of the interlocutor's irresponsible drinking episodes. However, Hans easily formats a predicate + category combination (*guai haizi*), which suggests that such a categorisation, its bound predicate and inferences are not constructed on the spot. This combination is locally regulated, culturally pre-packaged and contextually embedded in the very social role of Hans vis-à-vis the role of the interlocutor.

Excerpt 2.4 Hans's closing sequence where 'guai' resurfaces

1 I Wo hui=wo=>wo hui zhuyi de< (0.1)
 I AUX I I AUX pay attention NOM
 'I will=I=>I will be more careful< (0.1)'

2 I °Bu hao yisi° (0.3) °Hai shushu°
 I'm sorry (Hans's family name) uncle
 °I'm sorry° (0.3) °Uncle Hai°'

3 (1.7)

4 H → Hao a: (0.3) Guai:
 good PRT obedient
 'Ok: (0.3) Be good:'

5 (0.3)

6 I °E (0.1) hao de°
 PRT good NOM
 'Oh (0.1) Okey'

7 (0.6)

8 I Xin na wo=°na wo yihou hui zhuyi de°=
 Ok then I then I future AUX pay attention NOM
 'Okey, so I=°so I will be more careful in the future'

9 I =na xiexie shushu
 then thank uncle
 '=so thank you uncle'

10 (1.1)

11 H En
 PRT
 'Ok'

12 (0.6)

13 I Hao de
 Good NOM
 'Okey'

Lastly, the 'guai haizi' categorisation is not a one-off isolated incident, as Hans also makes it interactionally relevant in his closing sequence. Excerpt 2.4 is towards the end of the interaction and starts off with the interlocutor promising not to misbehave in the future. After issuing a positive assessment 'hao' in Line 4, Hans recycles 'guai' with emphasis and

elongation, only this time as an imperative. 'Guai' here takes on different functions as before: it is an assessment as in 'I underwrite what you just promised as guai behaviour'; it is also a veiled admonition as in 'be a guai kid and don't misbehave again'. In Line 6 the interlocutor first issues a change-of-state token 'E', which is similar to 'Oh' in English (Heritage, 1990). This token could be occasioned by surprise at the warning tone in Hans's 'guai', or incredulity at Hans's highly idiomatic use of 'guai' and his cultural knowledge of how Chinese people reprimand kids. Regardless of the mental state behind 'E', the interlocutor quickly proffers an agreement token and, after a 0.6 gap, reformats his promise from Line 1 in Line 8 and thanks Hans for his assessment/admonition in Line 9. The interlocutor's behaviour from Line 6 to Line 9 ratifies the legitimacy of Hans's use of 'guai' and Hai's implicated moral rights. It would be highly problematic if the interlocutor rebutted by saying, 'Hang on, why should I be guai or be a guai haizi?' or 'Who do you think you are to talk to me in such a patronising manner?'. This would be synonymous to a mother saying 'Why should I pick up my baby just because they are crying?', in Sack's example. It is the inherent moral order residing in categories that makes the claims from category-bound predicates irrefutable.

We hope that the above analyses have illustrated our point that MCA can be applied to L2 speakers' talk to describe their competence in social role enactment. As we have argued multiple times in this chapter, there is no language use unbounded to social roles. Hence, we maintain that the ability to enact appropriate social roles should be integrated into the IC construct. Drawing on the MCA analysis in this chapter, role enactment can entail the ability to 'assign categories to oneself and the interlocutor as appropriate in the roleplay scenario' and to 'evoke category-congruent predicates/activities to facilitate interaction'. Brian's and Hans's performances indicate that this ability is unaccounted for in existing speaking frameworks. From CA transcription we can observe numerous cases where Brian's and to a much larger extent Hans's language is linguistically inadequate, including excessively long gaps, awkward turn designs, infelicitous lexical choices and non-native prosodic features. However, inadequacies in 'speaking' do not obfuscate their ability to 'talk', to undertake categorisation or to make their categorisation work for them interaction-ally. This mismatch between 'speaking competence' and 'talking competence' can go a long way towards strengthening the construct validity of IC assessment.

5 Conclusion

This chapter makes a case for moving from the 'psycholinguistic- individualist perspective' to a 'sociolinguistic-interactional perspective' in our assessment of L2 speakers' competence in interaction (Roever & Kasper, 2018: 332). We argue that existing speaking tests fail to assess test-takers'

ability to interact, to launch social actions and to enact social roles. Such speaking tests cannot generate reliable inferences for stakeholders to make accurate prediction of test-takers' IC in real-world settings. To combat this deficiency, we survey existing attempts at IC assessment and highlight four concerns that are particularly pertinent to IC test designers. We also draw special attention to one much-overlooked sub-trait of IC – social role enactment – and propose that MCA can serve as a potential analytic candidate to unpack this trait and the categorical features of interaction. Future research can draw on the analyses and suggestions from this chapter to further our understanding and depiction of IC in assessment contexts.

Note

(1) There are also cases in the dataset where test-takers explicitly self-categorise as 'shushu' or 'ayi', which is the equivalent female category to shushu. Due to space limitations, these data are not presented but they reinforce the argument that the category knowledge of 'shushu/ayi' is shared between test-takers and NS interlocutors.

References

- Al-Gahtani, S. and Roever, C. (2012) Proficiency and sequential organization of L2 requests. *Applied Linguistics* 33 (1), 42–65.
- Al-Gahtani, S. and Roever, C. (2015) The development of requests by L2 learners of modern standard Arabic: A longitudinal and cross-sectional study. *Foreign Language Annals* 48 (4), 570–583.
- Al-Gahtani, S. and Roever, C. (2018) Proficiency and preference organization in second language refusals. *Journal of Pragmatics* 129, 140–153
- Austin, J.L. (1962) *How to Do Things with Words*. Oxford: Clarendon Press.
- Balaman, U. and Sert, O. (2017) Development of L2 interactional resources for online collaborative task accomplishment. *Computer Assisted Language Learning* 30 (7), 601–630.
- Bella, S. (2014) Developing the ability to refuse: A cross-sectional study of Greek FL refusals. *Journal of Pragmatics* 61, 35–62.
- Bernstein, J., Van Moere, A. and Cheng, J. (2010) Validating automated speaking tests. *Language Testing* 27 (3), 355–377.
- Brown, A. (2003) Interviewer variation and the co-construction of speaking proficiency. *Language Testing* 20 (1), 1–25.
- Brown, P. and Levinson, S. (1987) *Politeness: Some Universals in Language*. Cambridge: Cambridge University Press.
- Cekaite, A. (2007) A child's development of interactional competence in a Swedish L2 classroom. *The Modern Language Journal* 91 (1), 45–62.
- Chapelle, C.A., Enright, M.K. and Jamieson, J.M. (eds) (2008) *Building a Validity Argument for the Test of English as a Foreign Language*. London: Routledge.

- Chapelle, C.A., Enright, M.K. and Jamieson, J.M. (2010) Does an argument-based approach to validity make a difference? *Educational Measurement: Issues and Practice* 29 (1), 3–13.
- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Council of Europe (2018) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*. Strasbourg: Council of Europe.
- Curl, T.S. and Drew, P. (2008) Contingency and action: A comparison of two forms of requesting. *Research on Language and Social Interaction* 41 (2), 129–153.
- Dai, D.W. (2019) What do second language speakers really need for real-world interaction? A needs analysis of L2 Chinese interactional competence. *Language Teaching Research*. 1-38.
- Dai, D.W. (2020) Conversation-analytically validating domain experts' indigenous interactional competence rating criteria in test-taker exemplar responses. Manuscript submitted for publication.
- Dai, D.W. and Roever, C. (2019) *A task-based needs analysis of interactional competence in Chinese as a foreign language*. Paper presented at the 2019 American Association for Applied Linguistics conference, Atlanta, GA.
- Dings, A. (2014) Interactional competence and the development of alignment activity. *The Modern Language Journal* 98 (3), 742–756.
- Dinkins, C.S. (2005) Shared inquiry: Socratic-hermeneutic interpre-viewing. In P.M. Ironside (ed.) *Beyond Method: Philosophical Conversations in Healthcare Research and Scholarship* (pp. 111–147). Madison, WI: University of Wisconsin Press.
- Drew, P. (2013) Turn design. In J. Sidnell and T. Stivers (eds) *The Handbook of Conversation Analysis* (pp. 131–149). Hoboken, NJ: Wiley-Blackwell.
- Eggy, S., Musial, J. and Smulowitz, J. (1999) The relationship between English language proficiency and success as a medical resident. *English for Specific Purposes* 18, 201–208.
- Elder, C. and McNamara, T. (2016) The hunt for 'indigenous criteria' in assessing communication in the physiotherapy workplace. *Language Testing* 33 (2), 153–174.
- Fitzgerald, R. and Housley, W. (eds) (2015) *Advances in Membership Categorisation Analysis*. London: Sage.
- Fitzgerald, R., Housley, W. and Rintel, S. (ed.) (2017) *Membership categorisation analysis. Technologies of social action (Special issue)*. *Journal of Pragmatics* 118.
- Francis, D. and Hart, C. (1997) Narrative intelligibility and membership categorization in a television commercial. In S. Hester and P. Eglin (eds) *Culture in Action: Studies in Membership Categorization Analysis* (pp. 123–152). Washington D.C.: International Institute for Ethnomethodology and Conversation Analysis and University Press of America.
- Galaczi, E.D. (2013) Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics* 35 (5), 553–574.
- Galaczi, E. and Taylor, L. (2018) Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly* 15 (3), 219–236.

- Golato, A. (2002) German compliment responses. *Journal of Pragmatics* 34, 547–571.
- Greer, T. (2016) Learner initiative in action: Post-expansion sequences in a novice ESL survey interview task. *Linguistics and Education* 35, 78–87.
- Hall, J.K. and Pekarek Dohler, S. (2011) L2 interactional competence and development. In J.K. Hall, J. Hellermann and S. Pekarek Doehler (eds) *L2 Interactional Competence and Development* (pp. 1–15). Bristol: Multilingual Matters.
- Hall, P., Keely, E., Dojeiji, E., Byszewski, A. and Marks, M. (2004) Communication skills, cultural challenges and individual support: Challenges of international medical graduates in a Canadian healthcare environment. *Medical Teacher* 26 (2), 120–125.
- Hellermann, J. (2007) The development of practices for action in classroom dyadic interaction: Focus on task openings. *The Modern Language Journal* 91 (1), 83–96.
- Hellermann, J. (2008) *Social Actions for Classroom Language Learning*. Clevedon: Multilingual Matters.
- Heritage, J. (1984) *Garfinkel and Ethnomethodology*. Cambridge: Polity Press. Heritage, J. (1990) Intention, meaning and strategy: Observations on constraints on interaction analysis. *Research on Language and Social Interaction* 24 (1–4), 311–332.
- Hester, S. and Eglin, P. (eds) (1997) *Culture in Action: Studies in Membership Categorization Analysis*. Washington, DC: University Press of America.
- Huth, T. (2006) Negotiating structure and culture: L2 learners' realization of L2 compliment-response sequences in talk-in-interaction. *Journal of Pragmatics* 38 (12), 2025–2050.
- Hymes, D. (1974) *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia, PA: University of Pennsylvania Press.
- IDP IELTS (2021) *Understanding the IELTS Speaking Band Descriptors*. <https://ielts.idp.com/prepare/article-understanding-the-ielts-speaking-band-descriptors>
- 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 dissertation, University of Melbourne. Jacoby, S.W. (1998) *Science as performance: Socializing scientific discourse through conference talk rehearsals*. Unpublished doctoral dissertation, University of California.
- Jayyusi, L. (1984) *Categorisation and the Moral Order*. London: Routledge & Kegan Paul. Johnson, M. (2001) *The Art of Nonconversation*. New Haven, CT: Yale University Press. Kane, M.T. (2006) Validation. In R.L. Brennan (ed.) *Educational Measurement* (4th edn) (pp. 17–64). Westport, CT: Greenwood.
- Kane, M.T. (2012) Validating score interpretations and uses. *Language Testing* 29 (1), 3–17.
- Knoch, U. and Macqueen, S. (2019) *Assessing English for Professional Purposes*. London: Routledge.
- Lazaraton, A. (1992) The structural organization of a language interview: A conversation analytic perspective. *System* 20 (3), 373–386.

Lee, Y.A. and Hellermann, J. (2014) Tracing developmental changes through conversation analysis: Cross-sectional and longitudinal analysis. *TESOL Quarterly* 48 (4), 763–788. Levelt, W.J.M. (1989) *Speaking: From Intention to Articulation*. Cambridge, MA: MIT

Press.

Levelt, W.J.M. (1999) Producing spoken language: A blueprint of the speaker. In C. Brown and P. Hagoort (eds) *The Neurocognition of Language* (pp. 83–122). Oxford: Oxford University Press.

Levinson, S.C. (2006) On the human ‘interaction engine’. In N.J. Enfield and S.C. Levinson (eds) *Roots of Human Sociality* (pp. 39–69). Oxford: Berg.

Levinson, S.C. and Torreira, F. (2015) Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology* 6, 731.

Lim, G.S. (2018) Conceptualizing and operationalizing second language speaking assessment: Updating the construct for a new century. *Language Assessment Quarterly* 15 (3), 215–218.

Malone, M.E. and Montee, M. (2014) Stakeholders’ beliefs about the TOEFL iBT® test as a measure of academic language ability. *ETS Research Report Series* 2, 1–51.

Messick, S. (1989) Validity. In R.L. Linn (ed.) *Educational Measurement* (3rd edn) (pp. 13–103). New York: Macmillan.

Murray, J.C., Cross, J.L. and Cruickshank, K. (2014) Stakeholder perceptions of IELTS as a gateway to the professional workplace: The case of employers of overseas trained teachers. *IELTS Research Reports Online* 78.

Ockey, G.J., Koyama, D., Setoguchi, E. and Sun, A. (2015) The extent to which TOEFL iBT speaking scores are associated with performance on oral language tasks and oral ability components for Japanese university students. *Language Testing* 32 (1), 39–62. O’Sullivan, B. (2019) The future of language testing. In C. Roever and G. Wigglesworth (eds) *Social Perspectives on Language Testing* (pp. 197–216). Frankfurt: Peter Lang.

Payne, G.C.F. (1976) Making a lesson happen: An ethnomethodological analysis. In M. Hammersley and P. Woods (eds) *The Process of Schooling*. London and Henley: Routledge & Kegan Paul/Open University Press.

Pekarek Doehler, S. (2018) Elaborations on L2 interactional competence: The development of L2 grammar-for-interaction. *Classroom Discourse* 9 (1), 3–24.

Pekarek Doehler, S. and Pochon-Berger, E. (2011) Developing ‘methods’ for interaction: A cross-sectional study of disagreement sequences in French L2. In J.K. Hall, J. Hellermann and S. Pekarek Doehler (eds) *L2 Interactional Competence and Development* (pp. 206–243). Bristol: Multilingual Matters.

Pekarek Doehler, S. and Pochon-Berger, E. (2015) The development of L2 interactional competence: Evidence from turn-taking organization, sequence organization, repair organization and preference organization. In T. Cadierno and S.W. Eskildsen (eds) *Usage-based Perspectives on Second Language Learning* (pp. 233–268). Berlin: Mouton de Gruyter.

- Pekarek Doehler, S. and Pochon-Berger, E. (2016) L2 interactional competence as increased ability for context-sensitive conduct: A longitudinal study of story-openings. *Applied Linguistics* 39 (4), 555–578.
- Pill, J. (2016) Drawing on indigenous criteria for more authentic assessment in a specific-purpose language test: Health professionals interacting with patients. *Language Testing* 33 (2), 175–193.
- Plough, I., Banerjee, J. and Iwashita, N. (eds) (2018) Special issue on Interactional Competence. *Language Testing* 35 (3).
- Roever, C. and Kasper, G. (2018) Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Language Testing* 35 (3), 331–355.
- Sacks, H. (1974) On the analysability of stories by children. In R. Turner (ed.) *Ethnomethodology: Selected Readings* (pp. 216–232). Harmondsworth: Penguin.
- Sacks, H. (1992) *Lectures on Conversation* (2 vols). Oxford: Blackwell.
- Sato, M. (2014) Exploring the construct of interactional oral fluency: Second language acquisition and language testing approaches. *System* 45, 79–91.
- Schmidt, R. (1983) Interaction, acculturation and the acquisition of communicative competence. In N. Wolfson and E. Judd (eds) *Sociolinguistics and Second Language Acquisition* (pp. 137–174). New York: Newbury House.
- Schumann, J.H. (1978) The Pidginization Process: A Model for Second Language Acquisition (pp. 367–379). Rowley, MA: Newbury House.
- Searle, J. (1969) *Speech Acts*. Cambridge: Cambridge University Press.
- Seedhouse, P. and Egbert, M. (2006) The interactional organisation of the IELTS speaking test. *International English Language Testing System (IELTS). Research Reports 2006* (6), 1.
- Seedhouse, P. and Nakatsuhara, F. (2018) *The Discourse of the IELTS Speaking Test: Interactional Design and Practice*. Cambridge: Cambridge University Press.
- Stokoe, E. (2012) Moving forward with membership categorization analysis: Methods for systematic analysis. *Discourse Studies* 14 (3), 277–303.
- van Lier, L. (1989) Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly* 23 (3), 489–508.
- Waring, H.Z. (2013) ‘How was your weekend?’: Developing the interactional competence in managing routine inquiries. *Language Awareness* 22 (1), 1–16.
- Watanabe, A. (2016) Engaging in an interactional routine in EFL classroom: The development of L2 interactional competence over time. *Novitas-ROYAL (Research on Youth and Language)* 10 (1), 48–70.
- Watson, D.R. (1978) Categorisation, authorisation and blame-negotiation in conversation. *Sociology* 12 (1), 105–113.
- Youn, S.J. (2015) Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing* 32 (2), 199–225.

Youn, S.J. (2018) Task-based needs analysis of L2 pragmatics in an EAP context. *Journal of English for Academic Purposes* 36, 86–98.

Youn, S.J. and Burch, A.R. (eds) (2020) Where conversation analysis meets language assessment. *Papers in Language Testing and Assessment* 9 (1).