

Teacher Conceptualization of Interactional Competence in L2 English Speaking Assessment

Liliann Byman Friséen (Karlstad University)

Abstract

Interactional competence (IC) is central for human communication, but has received little focus in standardized second language (L2) speaking tests. Given that raters' understandings of assessment are reflected in scores, the study examines how raters conceptualize IC assessment. The data consists of focus group interviews with six groups of teachers ($N=24$) who act as raters of the Swedish National English Speaking Test; a test administered annually to all students in grades 6 and 9 (12–13 and 15–16 years). The study adopts a theoretical and empirical based model of IC, and data was analyzed using thematic analysis. Findings show that *interlocutor support* was particularly salient in teachers' conceptualizations of IC, and that teachers acknowledged students' social roles in interaction, which goes beyond previous IC definitions. Moreover, findings indicate that teachers conflated IC with linguistic competence. The study thus furthers our knowledge of how IC is understood and defined by teachers for assessment.

Keywords: interactional competence; L2 speaking assessment; standardized tests; rater conceptualization; teacher conceptualization; peer interaction

1. Introduction

Speaking proficiency in a second language (L2¹) can be divided into two separate abilities; the ability to *talk*, and the ability to *speak* (Roever and Kasper 2018). When people *talk*, they simultaneously speak and listen. *Talking* is a purposeful activity that can be equated with interaction, where body language and spatial arrangements are used to get things done. *Speaking*, on the other hand, can be equated with monologic speech, where

¹ In this paper, L2 is used for all languages acquired after the first language/mother tongue

the situational and social context of language use is of less importance (Roever and Kasper 2018: 333). The difference between *talking* and *speaking* is key when measuring an individual's L2 speaking proficiency: Does the test mainly measure an individual's ability to *speak* the L2 fluently, using vocabulary and grammatical structures in a correct and relevant way (i.e., measuring the individual's *linguistic competence*, LC), or is an individual's ability to *talk* with interlocutors (i.e., an individual's *interactional competence*, IC) mainly in focus in the test?

Roever and Kasper (2018) and Dai (2024) argue that although interaction between test-taker(s) and interlocutor(s) is included in some standardized L2 speaking tests (e.g., IELTS Speaking or the ACTFL OPI), interaction is mainly used as a means of eliciting spoken language, measured with LC-specific criteria. This presents a problem for the real-world use of test scores, as they may give a false impression of test-takers' ability to interact 'in a variety of settings and with a variety of interlocutors and not just to produce context-free fluent, accurate and complex language' (Roever and Kasper 2018: 331).

Although studies of rating scales indicate that LC rather than IC is in focus in standardized, high-stakes speaking tests (Dai 2024), raters' understanding and operationalization of scales for assessment is, ultimately, what is reflected in test scores (Ducasse and Brown 2009). Therefore, scholars (e.g., Ducasse and Brown 2009), call for empirical studies of how raters of L2 IC conceptualize interaction. Likewise, Dai (2024: 337) calls for studies that focus on cognitive processes of IC rating, in order to further our understanding of raters' meta-IC knowledge. Previous studies (Ducasse and Brown 2009; May 2011; Borger 2019) have shown that certain IC features are salient to raters of IC, but a number of questions remain. One pertains to which specific IC aspects raters realistically can score/attend to in real-life assessment situations (Galaczi and Taylor 2018; Lam et al. 2023). Studies drawing on conversation analysis have shown that IC can be distinguished on different proficiency levels (Pekarek Doehler and Pochon-Berger 2015; Roever and Kasper 2018). However, a majority of the work on raters' conceptualizations of IC have focused on upper secondary school or adults. Consequently, there is a knowledge gap as to how raters conceptualize younger students' IC. Likewise, whether differences in IC due to different proficiency levels are also reflected in raters' conceptualizations is, to my knowledge, still unexplored. The present study addresses these gaps in knowledge.

IC is multifaceted and known to be challenging to rate, and therefore often excluded from high-stakes speaking assessment (Lam et al. 2023). In Sweden, the empirical context for the study, all students in grades 6 and 9 (12–13 and 15–16 years old) take the National English Speaking Test (NEST), in which assessment of interaction is included. The NEST is a standardized high-stakes test administered by the Swedish Agency for Education. Typically, raters of the NEST are the students' own teacher, who receives no specific rater training for the assessment task, but is provided with assessment guidelines from the Agency (outlined in section 2). Assessment criteria are similar for both grades 6 and 9, and are *high-inference*, i.e., expressed in broad terms that leave considerable room for interpretation (Lam et al. 2023). Therefore, there are reasons to believe that conceptualizations of IC diverge, both within and across raters of the tests in grades 6 and 9.

The present study aims to provide new knowledge about how teachers conceptualize L2 interaction when operationalized for assessment in a standardized test. The following research questions guided the study:

1. What aspects of L2 IC do teachers report noticing and attending to when assessing interaction?
2. What similarities and differences are there between teachers' conceptualizations of L2 IC across proficiency levels in grades 6 and 9, respectively?

Findings from the study are expected to shed light on ways in which teachers relate to the assessment of IC in a high-stakes test. In addition, findings aim to demonstrate how teachers conceptualize this multifaceted construct in light of high-inference criteria. Thus, the study can provide insights on teachers' tacit knowledge of how assessment of *talking* can be operationalized; insights that are of essence for equitable assessment. As such, findings from the study may contribute to both teacher education, rater training and training of in-service teachers.

In what follows, theoretical and empirical approaches to L2 IC are outlined, including the model adopted for the present study. Next, the data and methods of analysis are presented. The findings are detailed in the subsequent section, and finally, the contribution of the present study is discussed through an expansion of the model adopted.

1.1 The construct of interactional competence

Several terms have been applied throughout the years to the skill of speaking and listening. *Oracy*, for instance, was introduced alongside *literacy* and *numeracy* in order to give spoken language skills comparable status (Mercer et al. 2017). In Hymes' seminal paper (Hymes 1972), the concept of *communicative competence* was proposed to refer to a person's competence regarding 'when to speak, when not, and as to what to talk about with whom, when, where, in what manner' (1972: 277). When discussing the 'dynamic process of communication', Kramsch (1986: 368) initiated a discussion of whether interaction extends communicative competence, in which competence is conceptualized as a cognitive property of the individual, independent from the situational context and interlocutors' use of language (Pekarek Doepler 2021). Kramsch (1986), He and Young (1998) and Young (2008) argued for the inclusion of the co-constructedness of interaction, which goes beyond individual ability. As such, *interactional competence* (IC) has come to highlight an individual's ability to act conjointly with others, and can fundamentally be understood as the ways in which participants interact in turns and sequences (Pekarek Doepler 2021). Galaczi and Taylor (2018: 221) exemplify the manifestation of this ability as encompassing role awareness and interactional resources, e.g., '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'. According to Dai (2024), IC can theoretically be defined in two ways. The first definition is that competence is co-constructed by speakers in interaction, and distributed between them. The second definition applies to an assessment context, where IC can be viewed as a competence that an individual can claim and display.

Research on the IC construct comes mainly from two fields: Conversation Analysis (CA) and Language Assessment (Galaczi and Taylor 2018). In an attempt at defining IC to encompass existing knowledge from both fields, Galaczi and Taylor (2018) use a tree as a metaphor. The roots of the tree represent 'the macrolevel context of the speech situation' and 'the microlevel context of the speech event and speech act' (2018: 226), whereas each of the five branches of the tree represent interactional skills (*topic management*, *turn management*, *interactive listening*, *breakdown repair*, and *non-verbal or visual behaviour*). The five skills identified include more fine-grained microlevel

features (e.g., the skill *Interactive listening* where *backchanneling*, *comprehension check*, and *continuers* are identified as microlevel features). Galaczi and Taylor (2018) state that the decision to use the image of a tree to define IC is to avoid hierarchical structures, and to acknowledge the interdependency and interrelation between different IC skills and their respective microlevel features.

Lam et al. (2023) also combined insights from CA and Language Assessment in a list of IC features with which they examined 23 scoring rubrics used for the assessment of speaking. The IC features listed were, in essence, identical to IC skills defined by Galaczi and Taylor (2018), but in addition Lam et al. (2023: 210) included *manner of interaction* ('how confident and interactionally pro-active test-takers present themselves'). When analyzing the scoring rubrics in the study, the authors found that *understanding of situation and context* was prominent in the rubrics, and added it to their list. Thus, Lam et al.'s list (2023: 220) consisted of seven IC features, to be compared with Galaczi and Taylor's (2018) five. In addition, Lam et al. (2023) divided *Topic management* into four sub-features: *topic development*, *responding to partner*, *topic maintenance*, and *topic initiation*. Lam et al.'s (2023) list therefore add to Galaczi and Taylor's (2018) definition of IC, but does not replace it since microlevel features associated with each skill, as well as the macro- and microcontext of the speech situation are not included. For this reason, I have combined Galaczi and Taylor's (2018) tree model with Lam et al.'s (2023) list in a new tree model (see Figure 1), and use it as the analytical lens in the present study.

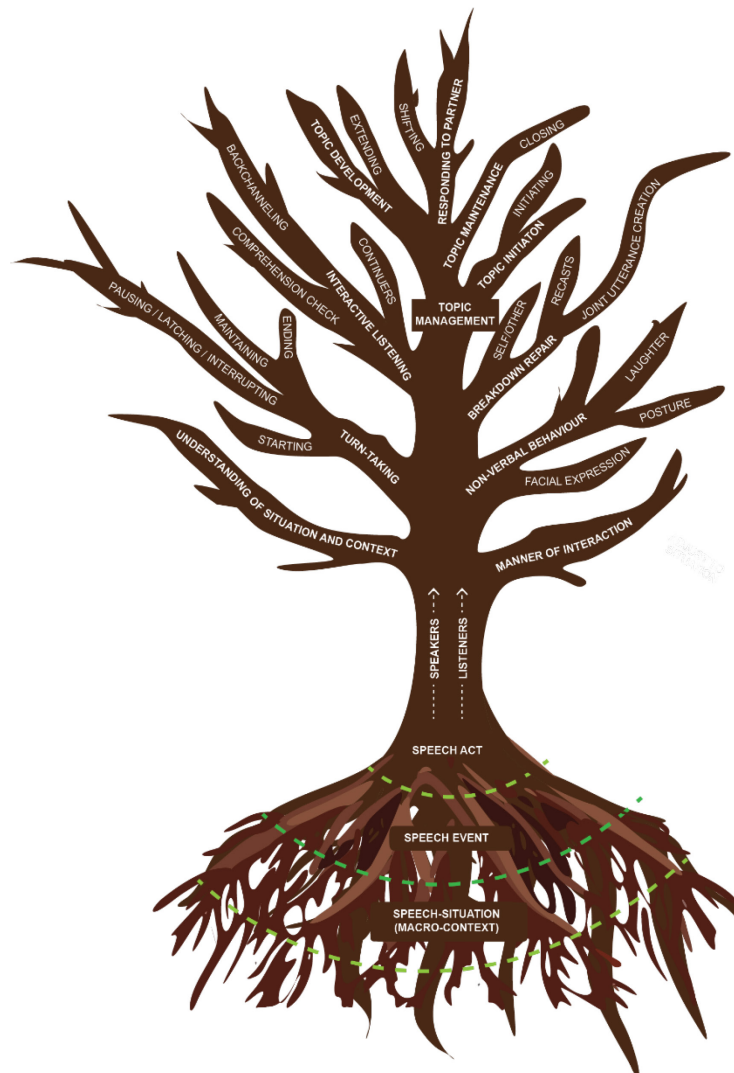


Figure 1: Tree model combining definitions of IC by Galaczi and Taylor (2018), and Lam et al. (2023)

The roots of the tree in Figure 1 are identical to the roots of Galaczi and Taylor's tree model, in which *speech situation*, *speech event* and *speech act* are seen as the context that interaction emanates from. Also, like Galaczi and Taylor's tree model, the tree trunk shows how interaction is

co-constructed by speaker and listener. The seven branches in Figure 1 represent the five IC skills defined by Galaczi and Taylor (2018) (*Turn-taking*, *Interactive listening*, *Topic management*, *Breakdown repair*, *Non-verbal behaviour*) and the two additions made by Lam et al. (2023) (*Understanding of situation and context*, *Manner of interaction*). Also, Figure 1 includes Lam et al.'s division of *Topic management* into four sub-features (*topic development*, *responding to partner*, *topic maintenance*, and *topic initiation*), represented by four smaller branches at the top of the tree. In Figure 1, several branches (i.e., IC features) have twigs that represent the microlevel features identified by Galaczi and Taylor (2018). In the present study, 'aspects of IC' will hereafter be used interchangeably with 'microlevel features of IC', when discussing previous research.

1.2 What raters orient to when assessing L2 interactional competence

Studies of raters' conceptualizations of IC are few, but have provided important insights. Borger (2019) asked one group of Swedish teachers ($N=17$), and one group of international teachers ($N=14$) to provide written verbal accounts of IC features they paid heed to when listening to audio-recordings of Swedish upper secondary school students' L2 English conversations. She found that both groups of teachers conceptualized IC in a similar way, and that three main interactional strategies were in focus in raters' conceptualizations: *topic development moves*, *turn-taking management*, and *interactive listening*. The latter (*interactive listening*) reiterates findings from Ducasse and Brown (2009), who studied rater conceptualizations of peer interaction in L2 Spanish, and May (2009, 2011), who analyzed trained raters' notes, stimulated recalls, rater discussions, and scores in a L2 English speaking test. In addition, both Ducasse and Brown (2009), and May (2009, 2011) found that *non-verbal behaviour* was salient to raters. Likewise, May et al. (2020) asked six trained examiners of a high-stakes paired L2 English speaking test to verbally report on interactional features salient in assessment. She found that raters paid attention to three non-criterion relevant categories: *body language*, *interactive listening*, and *manner of interaction*.

Both Ducasse and Brown (2009) and Borger (2019) found that raters noticed and considered *interactional management* in the peer pair, test-takers' *interactional roles* in the assessment situation and *the interrelation* of their performances. Similarly, May (2009, 2011) found that certain features of IC were viewed as co-constructed by interactants, which lead

May to suggest that shared scores for IC should be awarded in order to acknowledge the co-constructedness of interaction.

Additionally, valuable insights on L2 IC are provided by CA, which focuses on microlevel features in interaction. Kasper (2018) showed that the use of preliminaries (speech acts that lead up to a main action, such as a request) could be distinguished on three different proficiency levels, where low-level users did not use preliminaries, while more advanced students did. Likewise, Pekarek Doehler and Pochon-Berger (2015) found that low-level L2 speakers deployed a limited set of devices to achieve different actions, while high-level speakers made use of gradually diversified linguistic resources. However, Dai (2024) notes that L2 IC development is difficult to tease apart from an increase in L2 LC/general proficiency. Hence, L2 IC and L2 LC do not necessarily develop simultaneously.

2 Data and methods

2.1 The National English Speaking Test

Due to space limitations, all NEST guidelines from the Agency of Education to teachers cannot be reiterated (but see Sundqvist and Sandlund 2024 for an overview). Of essence for this study are guidelines stating that students should conduct the test in pairs², and that pairing of students should be based on similar proficiency levels. In addition, teachers are instructed to stay in the background, but may provide support if perceived as necessary.

The underlying construct for NEST in grades 6 and 9 is the same (hereafter, NEST6/NEST9): *oral production and interaction*. The test task usually starts with a warm-up descriptive part, followed by a part aimed at generating peer discussion prompted by topic cards. Students take turns reading a card and starting to speak on each statement or question³ (see Sundqvist and Sandlund 2024: 14–17, for more examples of the test task

² However, small groups of 3–4 students have previously been allowed.

³ An example of a NEST6 topic card with a question is ‘What do you like best? • Cooking • Singing • Drawing – Explain why! What don’t you like? Why? Ask your friend. What about grown-ups (parents/grandparents)?’ (University of Gothenburg 2025a). An example of a NEST9 topic card with a statement is ‘Money makes people happy. Agree? Disagree? Explain why and give examples. What does your friend think?’ (University of Gothenburg 2025b).

and a detailed description of the task procedure). Included in the guidelines is a one-page assessment document (see University of Gothenburg 2025c, a scoring rubric, and benchmark examples (audio files of students conducting the NEST along with commentary and scores from expert raters)).

2.2 Design and participants

The present study is part of a larger project on teachers' conceptualizations of oral proficiency assessment, for which ethics approval was obtained from the researcher's local university. The internal ethical review process concluded that a full ethical review with the Swedish Ethical Review Authority was not necessary, as no sensitive personal data was collected. Data consist of focus group interviews with teachers of L2 English who regularly act as raters of NEST6 or NEST9. All participants submitted their informed consent for participation in the study. Participants were colleagues working in the same school. Six groups of teachers from six schools were interviewed; three groups teaching grade levels 4–6 (rating NEST6), and three groups teaching grade levels 7–9 (rating NEST9). Henceforth, the teachers will be referred to as NEST6 or NEST9 teachers. The focus group interviews were conducted on site, during scheduled weekly teacher meetings. Therefore, the allotted time for an interview was generally limited to 60 minutes.

To be included in the study, teachers needed to have experience from acting as raters of NEST6/NEST9 at least once. The number of teachers in each group varied ($N=2-6$) depending on school size (see Table 1). Each school name is a pseudonym, and likewise, all teachers have been given a pseudonym that corresponds to the first letter of the name of the school (e.g., *L* for *Lime School*), the grade level for which they act as rater of the NEST (e.g., Grade 6=G6), followed by T for *Teacher* and a number (e.g., L-G6-T1). In total, 24 teachers participated in the focus group interviews, 12 teachers per grade level, rating NEST6 and NEST9 respectively.

Table 1: Focus groups, schools, participating teachers and experience of rating the NEST

Grade level, school and number of teachers in focus group	Teachers participating in the interview	Estimated number of times acting as rater for the NEST	Length of interview, h:min
4–6 Lime School 4 teachers	L-G6-T1 L-G6-T2 L-G6-T3 L-G6-T4	4 4 1 1	00:50
4–6 Hornbeam School 6 teachers	H-G6-T5 H-G6-T6 H-G6-T7 H-G6-T8 H-G6-T9 H-G6-T10	7 9 1 1 8 1	00:48
4–6 Oak School 2 teachers	O-G6-T11 O-G6-T12	1 2	01:02
7–9 Birch School 5 teachers	B-G9-T1 B-G9-T2 B-G9-T3 B-G9-T4 B-G9-T5	15 3 5 5 5	00:48
7–9 Maple School 4 teachers	M-G9-T6 M-G9-T7 M-G9-T8 M-G9-T9	17 17 1 5	00:52
7–9 Elm School 3 teachers	E-G9-T10 E-G9-T11 E-G9-T12	15 15 15	01:01

The interview guide was structured according to five stages of focus group interviews (*Opening, Introductory, Transition, Key, Ending*) as proposed by Krueger (1998). A brainstorming task was used as a prompt. The

subsequent discussion started from perspectives written down during the brainstorming task, followed by a discussion of the key questions, that focused teachers' conceptualizations of *interactional competence* in the test (see the Appendix).

2.3 Data analysis

Focus group interview data were analyzed qualitatively using reflexive thematic analysis (RTA, Braun and Clarke 2022), as this analytic method has the potential of illuminating as many perspectives of teachers' conceptualizations of IC as possible. I followed the six phases of Braun and Clarke's (2022: 35–36) analytic approach: *Familiarising yourself with the dataset*, *Coding*, *Generating initial themes*, *Developing and reviewing themes*, *Refining, defining and naming themes*, and *Writing up*. Data were transcribed using the AI-software Amberscript. For each phase, analysis was conducted for NEST6 and NEST9 groups separately, to examine conceptualizations of L2 IC for each proficiency level, as well as differences and similarities between these conceptualizations. In the first phase, *Familiarising yourself with the data*, I studied the transcripts meticulously in relation to the recordings, and edited them for accuracy. Initial thoughts and short summaries of each interview were written down. I subsequently imported the transcripts into the software program NVivo 14 (Lumivero 2023) for continued analysis. In the RTA analysis, I coded *data extracts* (i.e., anything potentially relevant in the transcripts) from which I generated *codes*. In turn, *themes* were generated on the basis of codes. Analysis was not linear, but followed an iterative process between phases 2 to 4, where data extracts (and consequently, codes) were moved, dismissed or added. Thus, I generated the first candidate themes. Due to the inductive approach adopted, several of the candidate themes were deemed irrelevant for the aim of the present study. In the next step, therefore, I adopted a more deductive approach to the data, where I selected candidate themes on the basis of their relevance for my research questions. The selected themes reflected what teachers reported as indicative of interactional conduct as well as conditions that may affect interaction. The themes were subjected to a new round of phase 4 analysis (*Developing and reviewing themes*), in which I reanalyzed semantically coded data extracts. In addition to using the software program in this phase, I drew thematic maps by hand for theme development. Maps made it possible to visually represent relationships between data extracts, codes

and themes. In phase 5 (*Refining, defining and naming themes*), a coherent narrative for each theme was written, which led to further changes, additions and refinement. The process resulted in the generation of two final themes for NEST6 and NEST9 respectively: *Signs of students' IC* and *Interactional Roles*. The theme *Signs of students' IC* encompasses codes that were generated by analyzing patterns of data extracts coded on a semantic level. I found the theme to be at the core of both the aim and research questions, as codes reflected different aspects of what teachers viewed as salient for IC. It is thus henceforth presented as the main theme. The theme *Interactional Roles* was generated from a more latent analysis of data. I found that it added to teachers' conceptualizations of interaction and is included as an additional finding. In the following section, findings will be presented in relation to the two themes generated through data analysis.

3. Findings

3.1 *Signs of students' IC*

The main theme generated for both NEST6 and NEST9 was *Signs of students' IC*. The theme consisted of several codes (see Figure 2), where six were the same for NEST6 and NEST9 raters. In Figure 2, these codes make up the inner circle, marked in blue. One code was found to be specific only to NEST6 teachers, and two codes were specific to NEST9 teachers. These codes are placed in the outer circle of Figure 2, marked in grey.

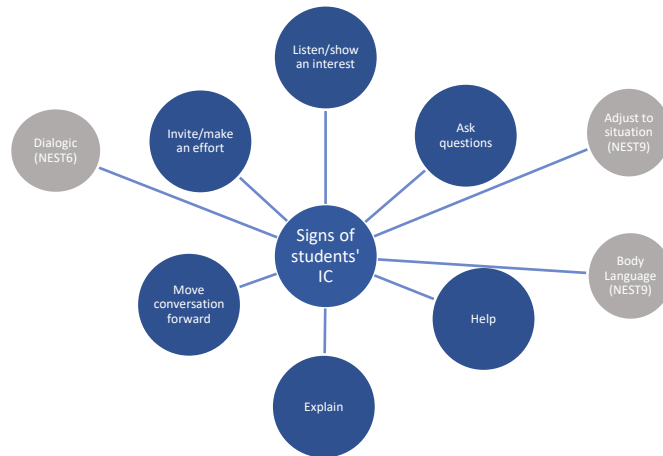


Figure 2: Codes generated from teachers' reports of *Signs of students' IC* during NEST

As previously stated, codes inherent in *Signs of students' IC* were generated by analyzing patterns across semantically coded data extracts. Extracts were in essence microlevel features mentioned as indicative of students' IC (or lack thereof). Although six codes were identical across NEST6 and NEST9 interviews, the inherent microlevel features were not necessarily the same. In order to show how microlevel features differed, Tables 2 and 3 show the codes generated for NEST6 and NEST9 teachers respectively, along with the microlevel features, and in some cases even more fine-grained microlevel features.

Table 2: Codes generated for the group of NEST6-teachers and the respective IC microlevel features they attended to/noticed

Code	Microlevel feature	Fine-grained microlevel feature
Listen/Show an interest	Listen Be interested in what peer is saying Be able to follow what peer is saying Listen attentively to try to understand peer	

Ask questions	Come up with questions Become interested and choose to ask follow-up question Wait for answer	
Help	Ask for help	By using body language From teacher or peer
	Offer help to peer	Explain Get your peer started Ask question Propose words and/or reformulate Be able to recognize peer's need of help
Explain	By using gestures	
Move conversation forward	Be able to continue and develop peer's utterances Have sufficient vocabulary to feel confident in speaking English Keep speaking even though words are lacking Dare to speak	
Dialogic	Answer questions together Not answer the question by yourself (i.e., a monologue) Give and take – time is evenly distributed Help each other to move the conversation forward	
Invite/make an effort	Both peers are equally engaged There is a flow in the conversation Leave the floor to each other (i.e., do not talk too much, show off or take over)	

Table 3: Codes generated for the group of NEST9-teachers and the respective IC microlevel features they attended to/noticed

Code	Microlevel feature	Fine-grained microlevel feature
Listen/Show an interest	Pick up what the peer has said and continue talking about it Confirm that you listen/understand Be attentive Not be preoccupied by thoughts about what you are going to say next	
Ask questions	Ask for clarification from peer if you have not understood Ask follow-up questions Ask your peer for help (NOT the teacher) Ask questions in return Dare to wait for an answer	
Adjust to situation	Engage in the task	Understand the task Support your peer in understanding the task
	Understand the 'rules of the game'	Keep to the topic Provide input and make an effort Play along
	Decode (i.e., understand what is going on and adjust accordingly)	
Body language	Look at each other (NOT at teacher) Show by the help of gestures what you mean Direction of gaze What they do with the hands	
Help	Seek confirmation from your peer (NOT from the teacher) Help and support each other	Notice whether the peer is stuck

	Help your peer	Adjusted to your peer's level of English Suggest words when peer is searching for them
	By the use of strategies	Reformulate Provide examples
Explain	Reformulate Use gestures Make oneself understood	
Move conversation forward	Discuss the topic in depth	
	Fill in for each other	
	There is a flow in the conversation	Vocabulary and idiomatic expressions (e.g., to make a joke in Eng.). Not use Swedish. Develop the topic (and thus use few cards) Keep the conversation going for a long time
Invite/make an effort	Take the initiative	Dare to be the one to start, e.g., by taking the lead and reading the instructions for one's peer
	Get each other to talk/on board Invite each other Leave the floor to your peer (i.e., do not talk too much, do not show off)	

When generating codes from microlevel features discussed by teachers in the interviews, the interrelatedness of IC features (Galaczi and Taylor 2018) became evident, as well as the fact that any model of IC, be it in the form of a tree or in the form of a table, is likely to misrepresent this interrelatedness. The categorization decision was based on the context in which each microlevel feature was discussed in the interviews, where I

made an estimation as to what for instance *ask follow-up question* seemed to mean for the teachers.

The code *Dialogic* was specific to NEST6-teachers' interviews. I found that the microlevel features inherent in the code illuminate NEST6-teachers' understanding that interaction is co-constructed (see Table 2). According to NEST6-teachers, the test is taken together and the conversation, therefore, moves forward as a result of a joint action. Co-constructedness of conversation was not addressed as specifically by NEST9-teachers, although an awareness that the test conversation is a joint accomplishment was discerned in some NEST9-codes, with microlevel features such as *help and support each other*, *fill in for each other*, and a *flow in the conversation* (see Table 3).

Analysis of several NEST6- and NEST9-codes showed that teachers' understanding of the test task included students' ability to ask and answer questions (specifically in focus in NEST6-teachers' interviews), and students' ability to discuss topics in depth (particularly attended to by NEST9-teachers). However, only NEST9-teachers reported expecting students to show an understanding of what the test situation entailed, and to perform accordingly. The code *Adjust to situation* was therefore only generated from NEST9-teachers, with microlevel features such as *understand the rules of the game*, *play along*, and *understand the task*. One explanation could be that NEST9-teachers can expect their students to be familiar with test expectations because of NEST6 experience.

Both NEST6- and NEST9-teachers reported that the use of body language was helpful in signaling a need for help and for explaining. However, the code *Body Language* was only generated from the NEST9-interviews, since it seemed to play a prominent role in NEST9-teachers' conceptualizations of IC. Gaze was specifically mentioned as an important interactional feature:

B-G9-T1: We talk a lot about this before the test period starts. We really go through what interaction is. Like, what should you do? Also this, to look at each other and confirm that I listen to what you are saying.

However, gaze was particularly important in signaling the addressee (the teacher or the other student), and as such, the extent to which students were in need of teacher support. In the Maple school interview, the teachers

discussed that students are supposed to understand that they are to leave the teacher out of the conversation, and only speak with each other. One of the teachers concluded:

M-G9-T9: [To speak with each other and not the teacher] says something about both the ability to express themselves and the ability to understand what the task is. And gaze can say something about that too, because sometimes you can see how they just stare. The eyes are screaming: Help us!

Thus, gaze was recognized as an important interactional resource, but mainly for indicating students' need of teacher support. In addition, gaze and gestures could help in discerning students' engagement in the task, as well as students' feeling of ease with the situation.

In the following, codes generated for the theme *Signs of students' IC* will be used as umbrella categories for the microlevel features presented in Tables 2 and 3, and related to the IC features presented in Figure 1. Due to lack of space, codes (instead of microlevel features) will be transformed into leaves that are added to the branches (see Figure 3).

In Figure 3, placement of codes (i.e., leaves) in relation to IC features (i.e., branches) was decided by the microlevel features inherent in each code. For instance, the NEST6-code *Move conversation forward* has been placed as a yellow leaf in the branch *Manner of interaction* because of the microlevel features *dare to speak* and *keep speaking even though words are lacking*. These microlevel features described manners in which interaction preferably should be conducted by NEST6-students according to their teachers. I therefore found them to reflect the *Manner of interaction*-branch. Because of this, NEST6- and NEST9-codes are occasionally placed in relation to different branches, and codes are also placed in relation to several branches. Had it been possible to place all microlevel features as leaves, they would have been placed in relation to one branch only.

Figure 3 shows that all codes can be categorized to one or several of the IC features presented in Figure 1. Findings thus indicate that teachers' conceptualizations of interaction relate well to Figure 1 (i.e., Galaczi and Taylor's (2018) as well as Lam et al.'s (2023) definitions of IC based on previous research). However, there were a few microlevel features that were difficult to place in the original model.

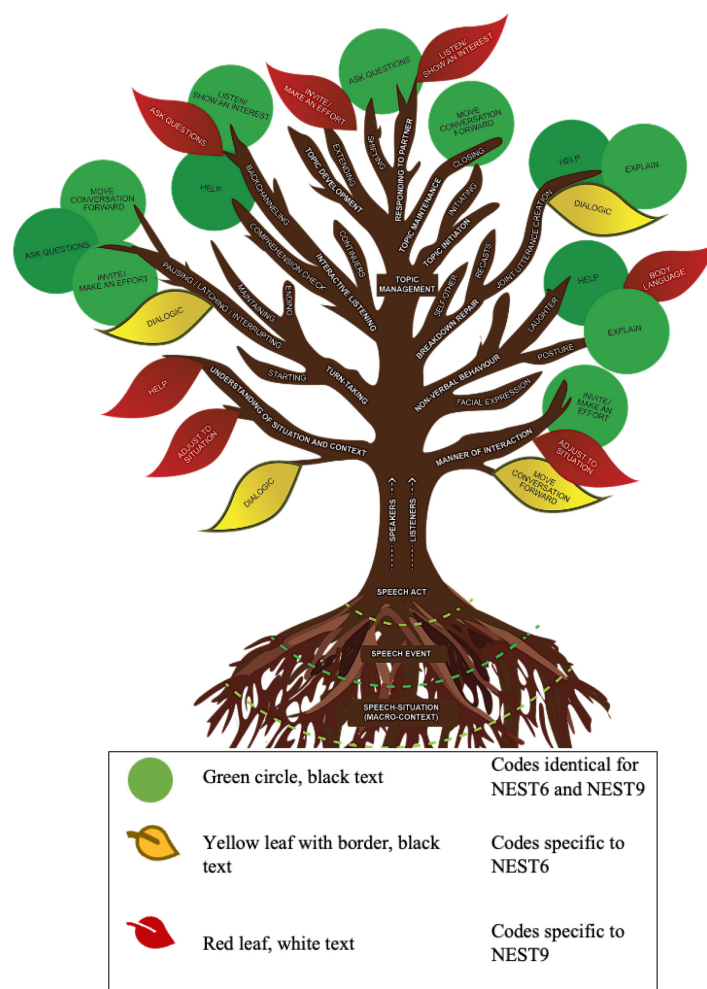


Figure 3: Codes generated for interaction in relation to Figure 1

For NEST6 teacher conceptualizations (Table 2), there were two more fine-grained microlevel features (*Get your peer started* and *Be able to recognize peer's need for help*) that were difficult to place. Likewise, in Table 3 (i.e., NEST9-teachers' conceptualizations) the microlevel feature *Get each other to talk/on board* was found not to fit. Although these

features were inherent in the codes *Help* (in Table 2), and *Invite/make an effort* (in Table 3), they also carry a meaning of being responsible for the peer's participation in the conversation, as well as having the ability to know when to step in to provide help, which was not present in Figure 1. A plausible reason for why this was not captured in Figure 1, is the fact that ways in which test-takers manage shifts from one speaker to the next in group assessment is under-examined (Al Abbas 2019).

In Figure 3, the code *Help* (for both NEST6- and NEST9-teachers) was linked to several IC features (*Non-verbal behaviour*, *Breakdown repair*, *Interactive listening* and *Understanding of situation and context*). Thus, the role of the attentive and helpful peer was viewed as prominent for IC. When asked what a student who is good at providing help does, teachers at Maple School had the following conversation:

M-G9-G6: Ask help-questions so that the other student gets started. Make sure to speak in a simple way. Ask questions so that the other student can answer in a simple way so that they also get into the conversation. They can—if their peer tries to tell [something] and describe a word, maybe they can say the word: 'is this what you mean?'
 M-G9-G9: Make space. That they don't take over, but that they invite [the peer] and make sure to make space for them
 M-G9-G8: Ask a question now and then. 'Bla, bla, bla', like, 'what do you think?'. And then get their peer to 'bla, bla, bla'.
 M-G9-G9: Sometimes they also dare to wait for an answer. To not be too quick. Some students ask a question but they are perhaps paired with a peer that needs a bit of time to think things through. Because students vary a lot when it comes to how long time they need to think. If the peer does not dare to wait for an answer but starts pushing too early, then they have not really helped their friend. That is also part of being helpful.

Being a helpful peer was thus conceptualized as being attentive to signals of discomfort and/or inability and to provide support accordingly. But it was also to adjust to the peer's level of English, and to give the peer enough time to think.

When teachers were asked to define what good interaction looked like, it was common to describe what it was *not*: it was not a monologue, that is, to hold the floor for much too long without inviting the peer into the

conversation, nor was it to ‘show off’ to the teacher. Rather, it was the ability to adapt to a co-participant and to be good at providing (the right kind of) support (at the right time and level) that was salient to teachers. Thus, students needed to balance the need to showcase their proficiency with providing support to a co-participant.

3.2 *Interactional roles*

As mentioned, the theme *Interactional Roles* was found to be salient in teachers’ IC conceptualizations. According to the teachers, pairing of students was vital, as the two students will take on different interactional roles during the NEST, which will affect students’ interaction. As the analysis revealed, the teachers attributed qualities to students that defined the interactive roles they took/were assigned in the test. The following qualities constitute the codes for the theme Interactional Roles:

- The weak or silent student
- The good student
- The support-student
- The ‘eager-to-get-a-good-score’-student

These roles affected students in different ways. One was that students’ confidence in speaking English could be both positively and negatively affected by the role they were given or took. Not surprisingly, *weak or silent students’* confidence was often seen as being negatively impacted by being paired with somebody of higher proficiency. For some, however, being paired with a more proficient student could act as a positive trigger.

Good students’ confidence could also be affected both positively and negatively by being paired up with a lower proficiency partner:

O-G6-T11: Well, the result was that the good student tried to help. Almost too much, so that the other student barely dared say anything

O-G6-T12: On the other side—the good student perhaps became even more brilliant in their-. Because they felt that they were so much better than the other person. If one had paired them with someone on their own level, maybe they wouldn’t have said as much.

However, teachers predominantly discussed negative effects on students paired up with a partner who was (very) weak or silent. One effect was stress, which could lead the *good student* to speak more than intended. Another was the fact that the lack of response from their peer led to a missed opportunity to show one's proficiency and IC:

L-G6-T1: Even that they [*good students*] try to pull those who nervously sit there. They try to get them to cooperate in one way or another, because a good student understands that if I don't get the other person going then there will be no conversation, and then I don't get to show [my proficiency].

Hence, teachers reported that asymmetric pairing of students often led to an uneven balance in terms of turn-taking and topic development, as well as an unfair situation for both students. The 'weaker' students could also find themselves in a role that hindered their opportunities to show IC:

O-G6-T12: In a conversation like that someone often becomes the dominant one, and that means that it can silence a person because the dominant one often steps in and like 'I know what you mean'. And then they say 'a dog' or something like that. Perhaps their peer would have been able to come up with that had they just gotten more time.

Asymmetric pairing could be unfair also to the 'good' student role, since they occasionally need to take on the role as the *support-student*, whose job it is to get the other person to perform. According to the teachers, however, symmetric pairing of students is not necessarily preferable in all cases, particularly not for *weak or silent* students. Two such students together might not be able to help each other at all, and conversation comes to a halt. Pairing of students was therefore reported to be given high priority, since it was seen as a prerequisite for interaction to develop successfully.

One interactional role discussed by all focus groups was the *role of the teacher*. This is not surprising, as several studies (e.g., Leyland et al. 2016) show how rater involvement in EFL group discussion tests impact interactional practices. Teachers are instructed to stay in the background, and thus not act as an interactional partner, but occasionally, teachers were invited to participate. Invitation was always in the form of directed gaze.

Teachers reported that the gaze of an *eager-to-get-a-good-score* student directed at the teacher was considered to be a sign of ‘showing off’. *Weak or silent* students could also direct their gaze towards the teacher, but this was regarded as seeking confirmation or help. Both ‘showing off’ and ‘seeking teacher confirmation’ were signs of insufficient IC.

4. Discussion

Galaczi and Taylor (2018) pose an essential question when asking what aspects of interactional competence it is sensible or realistic for raters to attend to when assessing interaction in real-life assessment contexts. Findings show that teachers recognized numerous interactional resources students used when acting conjointly with others. These resources align with Galaczi and Taylor’s (2018) and Lam et al.’s (2023) definitions of IC (see Figure 3). In addition, teachers acknowledged how different interactional roles affected students’ output. Hence, the teachers’ conceptualizations reflect findings from previous research, and the context-sensitive, socially embedded definition of IC (e.g., Pekarek Doehler 2021). It is therefore realistic to assume that raters can attend not only to aspects such as *turn-taking* and *topic development*, but also to *understanding of situation and context*, *manner of interaction*, and *non-verbal behaviour* that mirror the multi-semiotic nature of IC (Lam et al. 2023).

When assigning different interactional roles to students in the peer-pair, teachers acknowledged the interlocutor effect (O’Sullivan 2002): the fact that interlocutors will bring different characteristics into the situation which will affect the interaction, and consequently, the assessable output. In the study, the theme *Interactional roles* is presented as an additional finding, in part since it was beyond the definitions of IC in the analytical framework of the study (i.e., Figure 1). A possible reason why this perspective is not captured in Figure 1 is that the social roles participants have in interaction, for example as *teacher* or *student*, has received little attention in IC assessment (Dai 2024). When explaining the relevance of social roles for interaction, Dai (2024: 80) argues that if a person for instance considers themselves to speak from the social role of *student*, ‘the interaction itself between the interactants will further specify what [...] “student” means *in* the interaction, *to* the interactants’. The teachers in the present study particularly paid attention to whether students were able to

maintain their role as student interlocutor in the test situation, and not invite the teacher into the conversation for various reasons. A case in point is that if a ‘*eager-to-get-a-good-score*’ student directed their attention to the teacher, they stepped out of their role as student interlocutor and ‘showed off’. Or, in Roever and Kasper’s (2018) terms: the student stopped *talking*, and started *speaking*.

In terms of IC across proficiency levels, NEST6- and NEST9-teachers conceptualized IC in similar ways. The main difference between the two groups of teachers is in relation to *Understanding of situation and context*. Grade 9-students’ familiarity with the speech situation, their increased maturity, and their higher level of L2 English proficiency can partially account for why teachers had different expectations on NEST6- and NEST9-students. A plausible explanation could also be that topic cards might be expressed differently for NEST6 and NEST9, so that NEST9 topics may invite more extended discussion (see Sundqvist and Sandlund 2024). Nevertheless, NEST9-teachers expected their students to conduct more developed, authentic-sounding conversations. These findings are supported by previous research, where the ability to sustain a topic distinguishes low- and high IC level (e.g., Galaczi 2014). Also, according to Pekarek Doehler (2019: 37), insights from CA show that IC development ‘can be usefully understood as a progressive diversification of methods for action, comprising an increased sensitivity for such issues as recipient design, context sensitivity and ... preference organization’. Although the teachers in the present study did not use such terminology, NEST9-teachers’ expectations of more authentic conversations could stem from a tacit understanding of their students having a greater toolbox regarding methods for action than NEST6-students. However, both teacher groups expected their students to be able to participate in and take responsibility for the interaction without assistance.

A challenge for raters is the fact that some IC features can more easily be assessed on an individual level, whereas others are inherently co-constructed (May 2009). In the present study, teachers conceptualized IC features to be both individual and shared; the latter particularly mirrored in the NEST6-specific code *Dialogic*. In addition, in line with Sundqvist et al. (2018), the teachers found that pairing of students was vital for successful interaction. When dialogue in an initial pair failed, teachers occasionally gave students a second opportunity to take the test with another student. Thus, a valid individual score depended on the successful

interplay of both students' interactional conduct. However, IC features were for the most part conceptualized as individual competencies, for example, how a student invited their peer, the kind and level of support provided, and the ability to yield the floor when necessary. The product, that is, the conversation between the students, was not explicitly described as a joint construction, and none of the teachers expressed any difficulties in awarding individual scores, *provided that both peers contributed to the conversation and played their part*. Thus, although students' performances relied on their peer's contribution, teachers generally conceptualized IC as an individual ability. One possible reason is that teachers' conceptualizations were filtered through the specific macro-context in which the NEST is set: a standardized and high-stakes test of L2 speaking proficiency, where an individual score is required. When scrutinizing the codes generated for both NEST6 and NEST9 for the theme *Signs of students' IC* (*Listen/Show an interest, Ask questions, Help, Explain, Move conversation forward, Invite/make an effort*), five out of six codes reflect productive skills, whereas only one focuses on receptive skills. Thus, IC is conceptualized mainly as something students show or do in the test situation. This finding is in line with the definition of IC in assessment contexts as a competence that individual test-takers can claim and display (Dai 2024).

In the teacher interviews, being a supportive interlocutor stood out as indicative of students' IC. Lam et al. (2023) placed *interlocutor support* and *breakdown repair* in the same category and defined it accordingly: 'how [students] deal with conversational trouble, and ask for or provide interlocutor support' (Lam et al. 2023: 210). Teachers in the present study also conceptualized *interlocutor support* to encompass ways in which breakdown of conversation could be avoided, or repaired. However, interlocutor support was also related to *interactive listening* where students were expected to react not only to verbal contributions, but also to non-verbal conduct indicative of communication problems. For instance, students were expected to notice peer nervousness and be prepared to provide emotional support if needed. In addition, in line with previous studies of what raters view as salient in IC assessment (e.g., May 2011; Borger 2019), teachers conceptualized turn-taking as important, where students needed to be sensitive to the dialogic nature of the situation and not hold the floor for too long. Also, when paired with a peer on a lower proficiency level, students were expected to adjust their

contributions accordingly. As such, being a supportive interactant was seen as more important than displaying one's speaking skills (compare Roever and Kasper 2018); a priority that might be difficult for students to both recognize and abide by in a standardized, high-stakes test. Moreover, for students with special needs these expectations might be particularly difficult to live up to (which is an issue beyond the scope of the present study, but see e.g., Mamas et al. 2021). Brooks (2009), when asking adult test-takers about their experiences of conducting a paired speaking test or the format of an oral proficiency interview with an examiner, test-takers preferred the latter over the former. The main reason given was that the paired format put additional pressure on being responsible for interactional management. Although speculative, the pressure felt by the adult students in Brooks's (2009) study is likely felt also by the young NEST students. Thus, there is a risk that too much responsibility for providing interactional support in an L2 is put on these young students.

As mentioned, the construct assessed in NEST is *oral production and interaction*. Hence, teachers are asked to simultaneously assess *speaking/LC* and *talking/IC*. This study disregards *speaking/LC*-considerations for the assessment of the NEST, which is a limitation. Nonetheless, teachers' conceptualizations of IC reflect an acknowledgment of the difference between *talking* and *speaking* (Roever and Kasper 2018). Another limitation is that the teachers were not asked to report if and to what extent interactional features were considered in scoring. It is therefore difficult to know how interactional aspects attended to by teachers are also mirrored in scoring decisions, and how teachers' conceptualizations of IC affect final scores. However, since the rating task is to provide one score for both *speaking* and *talking*, it is plausible that both are considered. Findings indicate, on the one hand, that interaction in peer dyads in the NEST is mainly seen as a way of eliciting spoken production to be considered from a *speaking*-perspective, since there were few differences between how teachers conceptualized NEST6- and NEST9-students' IC, and both groups of students were expected to provide interlocutor support in similar ways. Thus, few L2 IC-specific considerations seemed to be taken. In addition, teachers relied on the students to *get each other to speak*. When they did, conversations ran smoothly, and awarding an individual score to students appeared unproblematic. On the other hand, students were expected to adjust to the situation, and when needed, prioritize *talking* over showing their ability to

speaking. Therefore, the teachers seemed to conflate *speaking* and *talking* in assessment. Possibly, this poses a dilemma for both teachers and students in their understandings of what is at stake in the test (and thus, what impacts scoring decisions); the ability to *talk* in English, the ability to *speaking* English well, or the ability to do both simultaneously?

5. Conclusion

The present study aimed at providing new knowledge about how teachers conceptualize L2 interaction when operationalized for assessment in a standardized test. The study demonstrates how participating teachers attended to numerous interactional aspects in line with previous research, and that interlocutor support was conceptualized as particularly salient across proficiency levels. In addition, teachers acknowledged how students' social roles vis-à-vis their peer and the testing context affected interaction. The study thus furthers our understanding of how IC is defined by teachers-as-raters in an assessment context. In future L2 IC assessment research, the Tree Model combining definitions of IC (Figure 1) can be used to analyze data in a similar way as in the present study. Additional empirical work can contribute to developing the current model, for instance by including test-takers' social roles. Moreover, the study illuminates that interaction in peer dyads has the potential of eliciting a number of ratable interactional features (Brooks 2009), while also pointing to challenges that the test situation poses for teachers and students alike. Future studies might provide insights into how raters attend to such challenges, and how they translate into scoring decisions. In addition, future research as well as CA informed findings in relation to how L2 IC develops can further support raters' level-appropriate expectations on students' IC.

The study also highlights how the teachers conflated *talking* and *speaking*. These concepts are closely interrelated (Dai 2024), making them hard to separate. More guidance on whether and how they should be distinguished would help both raters and test-takers understand which abilities matter most in paired speaking tests. In other words, is the purpose of interaction in the test first and foremost to produce ratable speaking/LC-output or is it to assess test-takers' ability to talk/interact in English? Since test scores reflect raters' conceptualizations of rating criteria (Ducasse and Brown 2009), implications of the study include a need for discussions of L2 IC-specific perspectives in both teacher

training, rater training, and among in-service teachers. In recent years, studies implementing authentic interaction data for teacher training have shown promising results (e.g., Sandlund et al. 2016; Sert and Waring 2025). Along similar lines, language teacher education could include practical tasks where student teachers verbalize their assessment of L2 interaction, and consequently, gain raised awareness of their own and others' conceptualizations of IC. As such, teaching and assessment of L2 IC might provide future L2 students with interactional resources with which they can use their L2 competently in social interaction.

References

- Al Abbas, Reem Hassanali. 2019. The enabler: facilitating next-speaker selection in L2 group oral assessments. PhD thesis, Newcastle University. <http://theses.ncl.ac.uk/jspui/handle/10443/4715>.
- Borger, Linda. 2019. Assessing interactional skills in a paired speaking test: Raters' interpretation of the construct. *Apples – Journal of Applied Language Studies* 13(1): 151–174. <https://doi.org/10.17011/apples/urn.201903011694>.
- Braun, Virginia, and Victoria Clarke. 2022. *Thematic analysis. A practical guide*. London: Sage.
- Brooks, Lindsay. 2009. Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing* 26(3): 341–366. <https://doi.org/10.1177/0265532209104666>.
- Dai, David W. 2024. *Assessing Interactional Competence*. Bern: Peter Lang. <https://doi.org/10.3726/b21295>.
- Ducasse, Ana Maria, and Annie Brown. 2009. Assessing paired orals: Raters' orientation to interaction. *Language Testing* 26(3): 423–443. <https://doi.org/10.1177/0265532209104669>.
- Galaczi, Evelina. 2014. Interactional competence across proficiency levels: How do learner manage interaction in paired speaking tests? *Applied Linguistics* 35(5): 553–574. <https://doi.org/10.1093/applin/amt017>.
- Galaczi, Evelina, and Lynda Taylor. 2018. Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly* 15(3): 219–236. <https://dx.doi.org/10.1080/15434303.2018.1453816>.

- He, Agnes W., and Richard Young. 1998. Language proficiency interviews: A discourse approach. In *Talking and testing. Discourse approaches to the assessment of oral proficiency*, edited by Richard Young and Agnes Weiyun He, 1–24. Philadelphia: John Benjamins.
- Hymes, Dell. 1972. On communicative competence. In *Sociolinguistics: Selected readings*, edited by J. B. Pride and Janet Holmes, 269–293. Penguin.
- Kramsch, Claire. 1986. From language proficiency to interactional competence. *The Modern Language Journal* 70(4): 366–372. <http://dx.doi.org/10.1111/j.1540-4781.1986.tb05291.x>.
- Krueger, Richard A. 1998. *Focus group kit*. Thousand Oaks: Sage.
- Lam, Daniel, Evelina Galaczi, Fumiyo Nakatsuhara, and Lyn May. 2023. Assessing interactional competence. Exploring ratability challenges. *Applied Pragmatics* 5(2): 208–233. <https://doi.org/10.58379/DAGH1483>.
- Leyland, Christopher, Tim Greer, and Ellen Rettig-Miki. 2016. Dropping the devil’s advocate: One novice language tester’s shifting interactional practices across a series of speaking tests. *Classroom Discourse* 7(1): 85–107. <https://doi.org/10.1080/19463014.2015.1100129>.
- Lumivero. 2023. *NVivo*, Version 14. www.lumivero.com.
- Mamas, Christophoros, Alan J. Daly, Shana R. Cohen, and Gabrielle Jones. Social participation of students with autism spectrum disorder in general education settings. *Learning, Culture and Social Interaction* 28: 1–14. <https://doi.org/10.1016/j.lcsi.2020.100467>.
- May, Lyn. 2009. Co-constructed interaction in a paired speaking test: The rater’s perspective. *Language Testing* 26(3): 397–421. <https://doi.org/10.1177/0265532209104668>.
- May, Lyn. 2011. Interactional competence in a paired speaking test: Features salient to raters. *Language Assessment Quarterly* 8(2): 127–145. <https://doi.org/10.1080/15434303.2011.565845>.
- May, Lyn, Fumiyo Nakatsuhara, Daniel Lam, and Evelina Galaczi. 2020. Developing tools for learning oriented assessment of interactional competence: Bridging theory and practice. *Language Testing* 37(2): 165–188. <https://doi.org/10.1177/0265532219879044>.

- Mercer, Neil, Paul Warwick, and Ayesha Ahmed. 2017. An oracy assessment toolkit: Linking research and development in the assessment of students' spoken language skills at age 11–12. *Learning and Instruction* 48: 51–60.
- O'Sullivan, Barry. 2002. Learner acquaintanceship and oral proficiency pair-task performance. *Language Testing* 19(3): 277–295. <https://doi.org/10.1191/0265532202lt205oa>.
- Pekarek Doehler, Simona, and Evelyne Pochon-Berger. 2015. The development of L2 interactional competence: Evidence from turn-taking organization, sequence organization, repair organization and preference organization. In *Usage-based perspectives on second language learning*, edited by Teresa Cadierno and Sören W. Eskildsen, 233–270. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110378528-012>.
- Pekarek Doehler, Simona. 2019. On the nature and the development of L2 interactional competence. In *Teaching and testing L2 interactional competence: Bridging theory and practice*, edited by M. Rafael Salaberry and Silvia Kunitz, 25–59. New York: Routledge.
- Pekarek Doehler, Simona. 2021. L2 Interactional competence and L2 education. In *Classroom-based conversation analytic research. Theoretical and applied perspectives on pedagogy*, edited by Silvia Kunitz, Numa Markee, and Olcay Sert, 417–426. Cham: Springer.
- Roever, Carsten, and Gabriele Kasper. 2018. Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Language Testing* 35(3): 331–355. <https://doi.org/10.1177/0265532218758128>.
- Sandlund, Erica, Pia Sundqvist, and Lina Nyroos. 2016. Testing L2 talk: A review of empirical studies on second-language oral proficiency testing. *Linguistics and Language Compass* 10(1): 14–29. <https://doi.org/10.1111/lnc3.12174>.
- Sert, Olcay, and Hansun Zhang Waring. 2025. *Conversation analysis and language teacher education. Intervention studies*. Cham: Springer.
- Sundqvist, Pia, Peter Wikström, Erica Sandlund, and Lina Nyroos. 2018. The teacher as examiner of L2 oral tests: A challenge to standardization. *Language Testing* 35(2): 217–238. <https://doi.org/10.1177/0265532217690782>.
- Sundqvist, Pia, and Erica Sandlund. 2024. *Testing talk. Ways to assess second language oral proficiency*. London: Bloomsbury.

- University of Gothenburg. 2025a. *Exempel på uppgiftstyper för årskurs 6. Focus: Speaking. Likes and dislikes – Elevkort*. Accessed June 2025. <https://www.gu.se/nationella-prov-frammande-sprak/prov-och-bedomningsstod-i-engelska/engelska-arskurs-1-6/exempel-pa-uppgiftstyper-for-arskurs-6>.
- University of Gothenburg. 2025b. *Exempel på uppgiftstyper för årskurs 9. Focus: Speaking. Elevkort – The world around us*. Accessed June 2025. <https://www.gu.se/nationella-prov-frammande-sprak/prov-och-bedomningsstod-i-engelska/engelska-arskurs-7-9/exempel-pa-uppgiftstyper-for-arskurs-9>.
- University of Gothenburg. 2025c. *Noteringsunderlag till delprov A. Nationellt prov i engelska för årskurs 9*. Accessed January 2025. <https://www.gu.se/nationella-prov-frammande-sprak/prov-och-bedomningsstod-i-engelska/engelska-arskurs-7-9/nationellt-prov-i-engelska-for-arskurs-9>.
- Young, Richard. 2008. *Language and interaction: An advanced research book*. New York: Routledge.

Appendix: Interview Guide

Question type	Questions
Opening Purpose: Background	What is your name and for how long have you worked as an English teacher? For how long have you worked at this school? Make an estimation of how many times you have rated the national English speaking test.
Introductory Purpose: get the participants to recollect and to picture themselves in the assessment situation	Tell me how the test usually is conducted at this school. You are going to do a brainstorm activity (hand out post-it notes and pens/pencils). Think back to your role as rater for the national English speaking test. What do you usually notice and attend to when students speak and interact? Write one thing per post-it note. (Give them time to write and wrap up when you see that they are out of ideas)

Transition Purpose: Sorting their different perspectives and leading the rest of the interview to aspects of <i>interaction</i>	Take turns reading what it says on your post-it notes. (To the one who starts:) Put the note down in the middle of the table and give a brief explanation to what you mean by what you have written. Does anybody else have a similar note/perspective? In that case, put it down on top of the first note and give a brief explanation of what you mean by it. (Continue the same way until all notes are put in different piles on the table) (Researcher chooses some of the piles of notes and says that we are going to zoom in on these perspectives. Make sure to include perspectives such as ‘adaption to purpose, recipient and situation’, ‘co-operation’, ‘interaction’, ‘engagement’, ‘will to communicate’, ‘communicative strategies’, etc.) (Possible steps ahead in case such perspectives are lacking: Go directly to key-questions and start with one of the aspects that is related to ‘interactional strategies’ Stay in the transition-question type and ask them to expand on their thoughts. Then lead the discussion to the key questions from here)
Key Purpose: illuminate participants’ collective knowledge about interactional competence and how students’ abilities manifest during the test	Let us start with X – what did you think about when you wrote it? What does a student who is good at X do? What does a student who is good at X say? How can you see or hear X in a student? How can you see if a student is good at X? Do you give feedback on X to your students? What do you usually say or write then? For a couple of years now, students are supposed to conduct the test in pairs (two and two). What are your thoughts about that?

The two last questions are only asked if teachers do not talk about it themselves	What do you think about when you hear ‘adaption to purpose, recipient and situation’? What does a student who is good at that do or say? What does ‘communicative strategies’ mean? What does a student who is good at that do or say?
Ending Purpose: finish off and make sure that nothing important has been missed	We have now talked about perspectives X, X, and X. You have said that Is that correct? Are there any other important aspects related to assessment of speaking proficiency that we have not talked about? Which ones?