Authentic interaction and examiner accommodation in the IELTS speaking test: A discussion

Anna Filipi

Monash University

Speakers naturally adjust their speech in interactions with others and will use accommodative strategies if their co-speaker is having difficulty understanding. These same adjustments have also been found in examiner accommodation in second language speaking tests (Cafarella, 1997; Ross, 1992).

In the training of examiners in the IELTS speaking test, there is an attempt to control the degree of examiner accommodation in the interests of consistency. Examiners are explicitly instructed to avoid the use of response tokens or to repeat a question only once without rephrasing it in the face of repair (Seedhouse & Egbert, 2006). This specific attempt to remove aspects of what is deemed to be authentic[1] interactional behaviour runs counter to what speakers do 'in the wild' as the growing body of research in conversation analysis shows (see for example Hutchby & Wooffit, 2008). We believe that it is timely to discuss the issue of examiner accommodation within a language-testing context against a backdrop of what is now known about naturally occurring interaction. We initiate such a discussion by reviewing the scholarly literature on interaction, and on the IELTS speaking test and examiner accommodation.

Keywords: IELTS speaking test, examiner accommodation, conversation analysis, authentic interaction

Introduction

In part, the issues presented in this paper have been driven by the anecdotal accounts of international students who sit the IELTS test in order to pursue study at Australian universities. Students may need to achieve 7.5 or 8 as a score in the IELTS speaking test in order to pursue a Master of Teaching for example, so they might sit the test as many as three times in a single year to improve their scores. As a result of this repeated experience, these students report a

Anna Filipi, Faculty of Education, Monash University; Email: anna.filipi@monash.edu.

discrepancy in IELTS examiner behaviour in the speaking test with respect to their perceived accommodation practices.

There is evidence from studies around the world that validates these anecdotal accounts. In a study conducted by Seedhouse and Egbert (2006) on the IELTS speaking test, for example, they found that examiners did indeed behave differentially with respect to accommodation to facilitate interlocutor understanding. Some examiners adhered strictly to the instructions in the training scripts that urged repeating a question exactly and not more than once when there were signs that a test-taker had not understood a question. Others accommodated by rephrasing their question or repeating it more than once. These different actions raise questions about the *practice* of language testing and the potential for inconsistency in examiner behaviour which are directly linked to the reliability of the test.

Reliability is an ongoing issue of concern in language testing. We need to feel confident in the judgements made about a test-taker's language ability and performance so that the information and inferences drawn can be used in ways that are accountable and justifiable (Messick, 1996). The discrepancy in rater judgements of performance is one area that can strongly impact reliability (Brown, Hudson, Norris, & Bonk, 2002; Sydorenko, Maynard & Guntly, 2014). In a recent paper that highlights this problem, Sydorenko, Maynard and Guntly (2014) maintain that in tests that are designed to assess students' pragmatic competence, the variability between examiners about what might be deemed to be appropriate pragmatic norms is highly contestable. This poses a significant challenge for test developers with respect to the training of examiners.

Equally problematic is examiner behaviour particularly if we are to take seriously the need to incorporate test designs that allow us to make judgements or predict test-takers' competence and abilities in being able to communicate in authentic ways in the real world. Indeed, a principal concern of proficiency tests is the extent to which they provide a representative sample of actual language use as per Bachman's (1990) Target Language Use (TLU) so that valid inferences can be made about a test-taker's ability to communicate in a range of real life situations. Thus in designing a test, attention must be given to its construct definition of language ability (and its component parts expressed in scales) as well as to the tasks that will elicit the performance sought against these scales (Bachman, 2002).

The issue about how test-takers' performance and results are affected by the divergence of examiners from the advice in the examination training manuals have been concerns held by researchers of a range of speaking tests including

the Certificate of Competency in English (ECCE) (Lu, 2005), the IELTS speaking test (O'Sullivan & Lu, 2006) and the Cambridge Assessment of Spoken English (CASE) (Lazaraton, 1996a). These investigations invite further exploration of the tension between the need for reliability and validity necessitating the restriction and control of natural or authentic interactional behaviour on the one hand, and natural or authentic interactional behaviour that emerges during speaking tests when there is mishearing or misunderstanding resulting in possible deviations from the training on the other.

This paper argues that differential examiner behaviour is inextricably linked to authentic interactional behaviour, which, as findings about examiner divergence show, appears to be difficult to control. What is deemed authentic behaviour (both as it emerges through the test interactions, and as it is conceived in the construct of the IELTS speaking test and the criteria for assessment) deserves consideration with respect to both the design of the test and the training of examiners. The paper begins the exploration of these issues by examining research findings about interaction drawing from conversation analysis, with specific reference to the practice of repair and how it is related to accommodation.

Interaction and speaking tests

In interaction participants strive for intersubjectivity or mutual understanding. The locus for uncovering this achievement is in the next turn at talk and what participants do next (Schegloff & Sacks, 1973). This analytic perspective is at the heart of conversation analysis with its focus on methods that are inductive and data-driven.

Originally developed by Harvey Sacks, Emanuel Schegloff and Gail Jefferson in the late 1960s and early 1970s, conversation analysis is premised on a view of interaction as socially situated action (Friedman, 2012). That is, language is used not only to convey meaning but to also enable people to do things in interaction (Schegloff, Koshik, Jacoby & Olsher, 2002). The interaction itself is the collaborative achievement of the speakers and a display of their intersubjectivity, of how they come to a joint understanding, and how this is achieved turn by turn.

One example of how speakers display their intersubjectivity is in repair initiation, a practice that involves speakers looking for sources of trouble that impede understanding, which might be the result of mishearing or misunderstanding (Schegloff, Jefferson & Sacks, 1977). Seedhouse and Egbert

(2006) maintain that intersubjectivity is not a goal of the IELTS speaking test. However, if this is true, then to what extent can the IELTS speaking test inspire confidence in being able to generalise about test-takers' abilities to apply their skills in the real world? A test-taker's capacity to initiate repair successfully (rather than to give an incorrect answer or to provide a minimal response as shown in some of the samples in Seedhouse & Egbert, 2006) ought to be judged as an ability to deal with trouble that arises in talk as proficient speakers of English do in any given situation. It is a crucial skill that language learners need to acquire and demonstrate as part of their interactional competence in their second language (Filipi & Barraja-Rohan, 2015).

Interaction demands that such skills be achieved collaboratively. With respect to repair, it is "natural" that when a speaker initiates repair, her or his co-speaker will have a range of options to assist in reaching mutual or intersubjective understanding so that the talk can then continue. This may involve repeating the preceding turn or reformulating it, actions that can be repeated if necessary to achieve a satisfactory outcome. The training scripts in IELTS however, demand that a question only be repeated once and exactly. So already we have a deviation from what can be construed as natural, authentic interactional behaviour.

Indeed, even within the "community of examiners", opposing views about repair have been reported. In a study of IELTS examiner perceptions about student self-repair (linked to the coherence and fluency scale), Brown (2006) found that some examiners viewed these abilities as mirroring the strategies of competent speakers. There were also conflicting views about hesitation (also linked to the same scale). Again this was considered by some examiners to be a behaviour that one would expect to see in competent speaker talk.

If we take the stance that coherence and fluency are jointly achieved and as much the products of the examiner's talk, then this raises important questions about examiners' interactional styles, how they perceive these practices, and their own actions. Thus how an examiner frames a question, what words s/he chooses, and how her/his utterances lead to expanded student turns have a co-dependency with how candidates respond to an examiner and structure their next turns. These actions will determine how successful candidates are in demonstrating oral competence in the speaking test. Importantly too, an examiner's interactional behaviour contributes in important ways to the overall authenticity of interactions in the speaking test, and is therefore an important component in a speaking test's construct.

The construct of the IELTS speaking test

The IELTS speaking test purports to be "interactive and as close to a real life situation as a test can get." (British Council, 2015). In terms of the test purpose, this means that the test measures constructs that define the ability to use language authentically. It follows then that in the speaking test examiners look for performances that demonstrate this ability. However, research on language testing has shown that a test-taker's performance is affected by a range of factors (Kim, 2010; McNamara, 1996), one of which is the examiner's interactional behaviour. An important part of this behaviour is the interactional work that speakers (including examiners) engage in to be understood and to understand. This leads to the establishment of intersubjectivity referred to above.

The IELTS speaking test comprises three sections. Section 1 is an introduction and interview about general topics. In section 2 test-takers are asked to speak for one to two minutes on a topic they are given one minute to prepare. In section 3, examiner and test-takers engage in discussion of the topic. The rubric for assessing performance comprises four constructs: fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation.

In terms of interactive ability, research on talk-in-interaction that draws on conversation analysis has shown that this ability revolves around turn-taking abilities and sequence organising abilities. These include for example the ability to monitor when it is appropriate to take the floor, how to hold the floor, how to ask and answer questions, how to open and close a conversation, how to close a sequence through third turn assessments or acknowledgement tokens, how to initiate and perform repair as stated above, and how to collaboratively complete utterances using a range of features that include verbal and non-verbal resources (Wong & Waring, 2010).

In past research that has been conducted on the IELTS speaking test from the perspective of conversation analysis, it has emerged that given the design of the test, there is minimal scope for students to engage in a variety of sequence types and turn-taking behaviours. This includes restrictions on asking questions (Seedhouse & Egbert, 2006), managing topics (Seedhouse & Harris, 2011) and opening and closing a conversation (Seedhouse & Egbert, 2006). Seedhouse and Egbert (2006) maintain that the absence of these characteristics contributes to the institutional quality of the interaction or the ways in which constraints operate on turn-taking as speakers orient to the institutional context and their identities (Heritage, 1998) which may lead to possible asymmetry between participants. Institutional interactions display features that have been identified as belonging to the work that is being conducted in the particular context (Drew & Heritage,

1992). These findings mirror those of Filipi (1994, 1998a, 1998b) who also used conversation analysis to investigate the interactions in a state oral examination in Italian as a second language.

In terms of interaction for academic purposes, Ducasse and Brown (2011) maintain that interviewer led tests are also restrictive because they curtail the range of interactions that might be possible and required for academic contexts. In their study, students were identified as needing skills to offer opinions, sustain arguments and give information in reaction to questions, to manage a range of interactions and to engage in team-work. In section 3 of the IELTS speaking test, Ducasse and Brown (2011) maintain that interaction (related specifically to opportunities for test-takers to develop topics and expand or control the direction of the discussion necessary for successful engagement in academic contexts) is largely missing. This suggests a problem in being able to reach a conclusion that a student's performance in the IELTS speaking test is predictive of her or his future academic performance.

The interactional nature of testing is one that should be fronted, irrespective of the institutional character of the interaction and the asymmetry of the participants with respect to speaking rights and turn allocation. The performance of the test-taker does not exist in a vacuum which is why performance and how to extrapolate a test-taker's ability to use language to wider contexts is fraught. Among the factors that affect performance are the interactions themselves and examiner talk.

Examiner talk

Examiner talk has been of pivotal importance in testing research because it goes to the heart of reliability. In order to make valid judgements about a student's speaking ability, an examiner needs to elicit behaviours that demonstrate key features of the ability through content that is consistent. However, they must also adopt consistent approaches. Using the microanalytic approach of conversation analysis, Filipi (1994, 1998a) showed that examiners in a final Italian oral state exam displayed styles that were either more interaction like or more interrogation like. Such differential behaviours have been shown to seriously affect reliability and they call into question judgements about ability (Brown, 2007).

With respect to the IELTS speaking test, Brown and Hill (1998) also found that examiners behaved in differential ways and, like Filipi's (1994, 1998a) findings for the Italian oral exam, were either more supportive or less supportive in their

styles. As a consequence, some test-takers could be more disadvantaged than others.

Research conducted on oral proficiency tests has identified a number of examiner accommodation strategies including rephrasing, elaboration, tag-questions, yes/no questions, either/or questions, repetition, modified articulation, lexical and grammatical simplification, display questions and fronting (Cafarella, 1997; Ross, 1992). It has also shown that by constructing questions in a particular way examiners can accommodate to prior interaction troubles in the interview. Such accommodative questions may indicate how interviewers form attributions of test-takers' speaking proficiency (Berwick & Ross, 1996; Lazaraton, 1996b; Ross, 1992; Ross & Berwick, 1992). Results of Ross and Berwick's (1992) study suggested that in the interview process interviewers accommodated to their interlocutor in much the same way that speakers in real life settings have been shown to accommodate through 'foreigner talk' to speakers whose first language is not English using actions such as rephrasing, incrementing or expanding a turn, and modifying (Gardner, 2004; Pallotti & Wagner, 2011). Ross (1992) suggested that in assigning final ratings in oral proficiency tests, the amount of accommodation that occurs should be taken into account so that the role of the examiner in the interaction is included.

Recurrent patterns of assessor accommodation have also been found in a state oral examination in Italian as a second language in Australia (Cafarella, 1997). These accommodations occurred as part of the interactive nature of the oral examination and were mainly displayed in repair sequences as exemplified in the extracts to be examined below. The immediate success of the accommodations was dependent on which form was used. Cafarella's (1997) research showed that repetition (or the 'repeat') was not immediately successful because the source of trouble was not identified. In the majority of cases, the trouble sources were lexical items or question ambiguity that had the potential to cause non-understanding or misunderstanding. It was through subsequent rephrasing and elaboration that assessors were able to succeed in overcoming such sources of trouble and in establishing student understanding as demonstrated through an appropriate next turn response.

In further research on examiner questions, Kasper and Ross (2007) found that questions contained in the same turn (which they termed horizontal multiple questions) were more accommodative. It was the examiners' extra interactional work that expedited the goal of getting the test-taker to produce a response projected by the question or request. Kasper and Ross (2007) concluded that for speaking tests to be valid, it is necessary to take into consideration interviewer question formulation styles as part of the rating.

These studies highlight the fact that accommodation is a natural feature of interaction, and that as speakers we do indeed assist each other in reaching a successful interactional outcome. The question for test developers and examiner trainers is not only whether, as Ross (1992) urged, these features should be taken into consideration in examiner behaviour if the purpose of the oral examination is to elicit the best performance with respect to the most natural behaviour, as defined above, but how, if reliability is to be secured and maintained. We will come back to this point shortly.

Examiner accommodation in the IELTS speaking test

Despite specific training of examiners, and changes to the rating scales for the revised interview scripts or Interlocutor Frames in the IELTS speaking test introduced in 2001 (see Brown, 2007), more recently Seedhouse and Egbert (2006) also reported that different examiner behaviours persisted. Some examiners were very supportive, mirroring situations in real life where speakers display accommodation when faced with co-participants who may be struggling to make themselves understood, while others were not and adhered strictly to the training scripts and Interlocutor Frames. This is in spite of specific training guidelines and tighter interview scripts where examiners are instructed very clearly to refrain from explaining key vocabulary for example.

The following transcripts taken from Seedhouse and Egbert (2006) provide examples of the kinds of examiner deviation from the training scripts that occurred in the IELTS test, which Seedhouse and Egbert (2006) maintain can advantage or disadvantage test-takers.

"Extract 15

63 E: what qualifications or certificates do you hope to get? (0.4)

64 C: sorry? (0.4)

65 E: what qualifications or (.) certificates (0.3) do you hope to get (2.2)

66 C: → could you ask me another way (.) I'm not quite sure (.) quite sure about

67 this (1.3)

68 E: it's alright (0.3) thank you (0.5) uh:: can we talk about your childhood?

69 (0.7)

(Part1)"

"Extract 16

40 E: uh so (0.5) ↑how would you improve (0.4) the city you live in (1.8)

41 C: I:: (0.8) how do I pro::ve? (0.2)

42 E: how would you impro:ve (.) the city (0.3)

43 C: sorry I don't know (.)

44 E: improve? (0.3)

45 C: yeah (.)

46 E: → how would you make the city better? (0.3)

47 C: → o::h yes (0.5)"

(Seedhouse & Egbert, 2006, p17.)

Both extracts show examples of trouble that emerge for the test-takers who then initiate repair. In Schegloff, Jefferson and Sacks' (1977) work on repair they found that it is not trouble that leads to repair in the talk but that repair initiation locates something in the preceding talk as a source of trouble. Repair or correction as a result of the initiation can thus be seen as an accommodative action.

In the two extracts above, we have exactly this kind of trouble emerging. We note that the repeat in line 65 in extract 15 and in line 42 in extract 16 (a repair strategy that orients to a problem of mishearing in the prior turn), is used by both examiners. As stated above, as an accommodation strategy in a testing situation the repeat has been found to be unsuccessful because the source of trouble is usually a lack of vocabulary knowledge rather than mishearing (Cafarella, 1997). However, in the IELTS training script, the repeat is the only form of accommodation that is permitted, and then only once. In extract 15, we can see that the examiner adhered to these instructions in the training script. Her/his line 65 *what qualifications or (.) certificates (0.3) do you hope to get (2.2)* is an exact verbal repetition of the initial question, although we do note that there are some pauses which suggests possible accommodation through other non-verbal and paralinguistic features.[2] In extract 16, however, the examiner accommodated twice. Initially s/he oriented to the possibility that the test-taker's problem was indeed one of mishearing thereby adhering to the training script, although

improve (the word causing trouble) is slightly stretched prosodically (*impro:ve*) again indicating a slight paralinguistic accommodation. Subsequently, in response to the test-taker's explicit statement of the problem, s/he reformulated by rephrasing the initial question where the problematic *improve* was replaced with *make the city better*. The result was the test-taker's ability to understand the question as indicated by the change of state token *oh* (Heritage, 1984) in line 47, and presumably her/his ability to then proceed to answer the question successfully. Clearly, intersubjective or mutual understanding, the aim of repair in everyday conversation, has been achieved in extract 16 and importantly, the student would have been advantaged.

Such findings about differential accommodation continue to raise questions about the fairness of the test if for example the result is that some candidates are more disadvantaged than others by a less accommodative examiner, who it must be stressed is strictly adhering to the test protocols. These examples also prompt us to ask whether an examiner's interactional actions should more closely mirror competent speaker accommodation behaviour when interacting with non-English speakers of English even in a test situation. If it is not possible to take the "natural" or "authentic" speaker out of *all* examiners as appears to be the case, then the implications for the IELTS speaking test (and indeed for standardised, high stakes speaking tests more generally) may be to dispense with the effort expended in training examiners to remove levels of supportive behaviour and features of authentic conversation, and to focus instead more directly on what kinds of support and at which levels these supportive behaviours may be appropriate.

The implications of the above findings point to the need for further research on the IELTS test to identify which features of accommodation from those identified above (derived from both natural and examination settings) will lead to more successful outcomes for the test-takers. Test-taker behaviours elicited through the interactions with the examiner that more closely mirror actions in authentic, naturally occurring contexts, and that are appropriate to each scale or group of scales, would need to be identified and established empirically from both test trials and actual test samples. By way of illustration, if we were to take repair, features we might seek to look for include:

- the extent of other-initiated repair required by the examiner or the degree of pursuit of test-taker understanding
- the types of other-initiated repair deployed by the examiner – e.g. rephrasing, modified articulation, grammatical simplification – to achieve test-taker understanding

- the extent of withholding or wait time by the examiner before producing a next turn.

Scoring could be based on the extent of pursuit, on the identification of more complex actions deployed in accommodation, and on the length of pauses between examiners and test-takers before responding.

Clearly, such changes that seek to describe the finer details of examiner accommodation and attempt to link them to scoring would have implications for test practicality and for test design. For example, a second examiner would be needed to assess the performance against a checklist rather than take part in the actual interaction. The time allocated to the test itself would also need to be reconsidered. Nonetheless, there would be considerable positive washback in the improvements to the interactional quality of the speaking test itself as it more closely mirrors interaction as it happens in the wild. Future research is strongly urged to elucidate exactly how such fine-grained attention to examiner accommodation might be related to student performance and to the creation of a more authentic speaking test.

Conclusion

The above discussion has explored the practice of controlling accommodative practices and examiner variability through training in the IELTS speaking test, and through the use of Interlocutor Frames and scripts that examiners must follow (O'Sullivan & Lu, 2006; Seedhouse & Egberts, 2006). The discussion acknowledges the wider practice of "examiner control" in the interests of reliability in other tests such as CASE (Lazaraton, 1996a) and ECCE (Lu, 2005). It concludes that despite these efforts, deviations from the scripts do indeed persist, and that accommodation actually continues to occur even if only through paralinguistic adjustment that escapes explicit attention. This draws into question the on-going practice of controlling such "natural" actions that are so persistent in examiner behaviour.

In drawing such conclusions, our discussion urges the ongoing need for examination boards to monitor test content, test design and examiner training against developments in language acquisition and use, in recognition of the dynamic nature of the relationship between them (Filipi, 2012). Indeed, the only accommodation feature that IELTS allows, the repeat, has been shown to be ineffectual (Cafarella, 1997), raising concerns about the extent to which test developers are keeping abreast of research. Changes made to test content, test design and examiner training that take into account developments in our

understanding of interaction would no doubt have a positive washback effect on the teaching of speaking in the classroom as well.

We conclude with five questions that we believe should be addressed by developers of the high stakes speaking tests such as IELTS where so much is at stake for the test-takers.

- Firstly, to what extent should the IELTS speaking test reflect authentic language use?
- Secondly and relatedly, what aspects of authentic interaction should be elicited in a testing situation now that such rich findings in CA have been amassed through research in both first (eg. Hutchby & Wooffitt, 2008) and second language interaction (eg. Gardner & Wagner, 2004; Pallotti & Wagner, 2011)?
- Thirdly, how are behaviours and performance a product of examiner training?
- Fourthly, how can training be developed to raise examiner awareness of authentic behaviour? Such awareness is important to understanding how examiner's actions as speakers in interaction can affect examination outcomes.
- Finally, how can natural, authentic interactional behaviour be given "space" to emerge without compromising the reliability and validity of a speaking test?

With the continuing global spread of the English language, international students will continue to be drawn to English speaking countries such as Australia, Britain, Canada and the USA for study and for work. This paper has attempted to make some initial observations about the constructs of authentic interaction, of accommodation as a practice both 'in the wild' and in controlled testing contexts, and of the problems that the latter creates when attempts are made to curtail authentic interaction in the interests of test/examiner consistency.

This paper does not downplay the need for consistency. On the contrary it hopes to have made visible how natural speaker behaviour is difficult to control. In the name of test accountability and fairness (as it relates both to examiner consistency and to the need for authentic, natural examiner behaviour), we invite further consideration about the extent to which accommodation that emerges in repair, itself a deeply human practice, should be better understood for its role in eliciting the best performance of students, and how it might be best managed in test administration and examiner training.

Acknowledgments

Endnotes

[1] The notion of the 'native speaker', and the dichotomy it creates with its opposite and often more deficient 'non-native speaker' notion, has long been contested in Applied Linguistics (eg Davies, 1991; Firth & Wagner, 2007; Hall & Cook, 2012; Paikeday, 1985). With the spread of English as an international language and lingua franca, this notion is further questioned (Firth & Wagner, 2007; Seidlhofer, 2001) because all speakers regardless of their proficiency in English bring their linguistic resources to bear on the need to create successful communication. For the purposes of this paper, "native" speaker is avoided in preference to "natural" or "authentic" speaker and/or interaction.

[2] Communication accommodation theory (eg. Giles, Coupland & Coupland, 1991) provides a case for the importance of non-verbal features in talk such as pauses and speech rate, which are indicative of how speakers positively accommodate (or converge) to one another. The multi-modal or non-verbal organisation of interaction (Stivers, & Sidnell, 2005; Streeck, 2009) adds another layer of complexity to the issue of examiner training. As far as we have been able to ascertain, this aspect of examiner behaviour has not been researched. However, it is an area that needs elucidation.

References

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. (2002). Some reflections on task-based language proficiency assessment. *Language Testing*, *19*, 453-476. DOI: 10.1191/0265532202lt240oa.

Berwick, R., & Ross, S. (1996). Cross-cultural pragmatics in oral proficiency interview strategies. In M. Milanovic & N. Saville, N. (Eds.), *Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium*, (pp. 34–54). Cambridge: Cambridge University Press.

Brown, A. (2006). An examination of the rating process in the revised IELTS speaking test. *IELTS Research Reports*, *6*, 41-69. Canberra, Australia: IELTS/ UK: British Council.

Brown, A. (2007). An investigation of the rating process in the IELTS oral interview. In L. Taylor & P. Falvey (Eds.), *Studies in language testing, 19: IELTS collected papers, 2: Research in speaking and writing assessment,* (pp. 98-141). Cambridge, UK: Cambridge University Press.

Brown, A., & Hill, K. (1998). Interviewer style and candidate performance in the IELTS oral interview. *IELTS Research Reports*, *1*, 1-19. IELTS Australia, Canberra and British Council, London.

Brown, J. D., Hudson, T. D., Norris, J. M., & Bonk, W. (2002). *Investigating task-based second language performance assessment.* Honolulu, HI: University of Hawai'i Press.

Cafarella, C. (1997). Assessor accommodation in the V.C.E. Italian oral test. *Australian Review of Applied Linguistics*, *20*, 21-41.

Davies, A. (1991). *The native speaker in applied linguistics.* Edinburgh: Edinburgh University Press.

Drew, P., & Heritage, J. (1992). *Talk at work: Interaction in institutional settings.* Cambridge England, New York: Cambridge University Press.

Ducasse, A., & Brown, A. (2011). The role of interactive communication in IELTS speaking and its relationship to candidates' preparedness for study or training contexts. *IELTS Research Reports*, *12*, 125-150.

Filipi, A. (1994). Interaction in an Italian oral test: The role of some expansion sequences. *Australian Review of Applied Linguistics*, *11*, 119-136.

Filipi, A. (1998a). Interaction or interrogation? A study of talk occurring in a sample of the 1992 VCE Italian oral common assessment task (Cat 2). *Australian Review of Applied Linguistics*, *21*, 123-137.

Filipi, A. (1998b). Managing roles in the openings and closings of an Italian oral test. *Monash University Linguistics Papers*, *1*, 3-10.

Filipi, A. (2012). Do questions written in the target language make foreign language listening comprehension tests more difficult? *Language Testing, 29*(4), 511-532. DOI: 10.1177/0265532212441329.

Filipi, A., & Barraja-Rohan, A-M. (2015). An interaction-focused pedagogy based on conversation analysis to develop L2 pragmatic competence. In S. Gesuato, F. Bianchi & W. Cheng (Eds.), *Teaching and learning about pragmatics: Principles, methods and practices,* (pp. 231-252). UK: Cambridge Scholars Press.

Firth, A. & Wagner, J. (2007). Second/foreign language learning as a social accomplishment: Elaborations on a reconceptualized SLA. *The Modern Language Journal, 91*(1), 800-819. DOI: 10.1111/j.1540-4781.2007.00670.x.

Friedman, D. A. (2012). How to collect and analyze qualitative data. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide, first edition,* (pp. 180-200). Oxford: Wiley-Blackwell.

Gardner, R. (2004). On delaying the answer: question sequences extended after the question. In R. Gardner & J. Wagner (Eds.), *Second language conversations*, (pp. 246–266). London: Continuum.

Gardner, R. & Wagner, J. (Eds.), (2004). *Second language conversations*. London: Continuum.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics*, (pp. 1-68). New York, NY: Cambridge University Press.

Hall, G., & Cook, G. (2012). Own-language use in language teaching and learning. *Language Teaching, 45*(3), 271-308. DOI:10.1017/S0261444812000067.

Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. Maxwell, A. & J. Heritage (Eds.), *Structures of social action*, (pp. 299-347). Cambridge: Cambridge University Press.

Heritage, J. (1998). Conversation analysis and institutional talk: Analyzing distinctive turn-taking systems. In S. Cmejrková, J. Hoffmannová, O. Mullerov & J. Svetlá (Eds.), *Proceedings of the 6th International Congress of IADA (International Association for Dialog Analysis)*, Tubingen: Niemeyer, pp.3-17.

Hutchby, I. and Wooffitt, R. (2008). *Conversation Analysis, second edition*. Cambridge: Polity.

Kasper, G., & Ross, S. (2007). Multiple questions in oral proficiency interviews. *Journal of Pragmatics, 39*, 2045-2070. DOI: 10.1016/j.pragma.2007.07.011

Kim, H. J. (2010). Investigating the construct validity of a speaking performance test. *Spaan Fellow Working Papers in Second and Foreign Language Assessment, 8*, 1-30.

Lazaraton, A. (1996a). A qualitative approach to monitoring examiner conduct in the Cambridge assessment of spoken English (CASE). In M. Milanovic & N. Saville (Eds.), *Performance testing, cognition and assessment: Selected papers from the 15th Language Testing Research Colloquium,* (pp. 18–33). Cambridge: Cambridge University Press.

Lazaraton, A. (1996b). Interlocutor support in oral proficiency interviews: the case of CASE. *Language Testing*, *13*, 151–72. DOI: 10.1177/026553229601300202.

Lu, Y. (2005). A validation study of the ECCE NNS and NS examiners' conversational Styles from a discourse analytic perspective. *Spaan Fellow Working Papers in Second and Foreign Language Assessment, 3*, 73-99.

McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256. DOI: 10.1177/026553229601300302.

O'Sullivan, B., & Lu, Y. (2006). The impact on candidate language of examiner deviation from a set interlocutor frame in the IELTS Speaking Test. *IELTS Research Reports, 6*, 91-118. IELTS Australia, Canberra and British Council, London.

Paikeday, T. M. (1985). May I Kill the Native Speaker? *TESOL Quarterly*, *19*(2), 390-395.

Pallotti, G. & Wagner, J. (Eds.), (2011). *L2 learning as social practice: Conversation-analytic perspectives.* Honolulu: University of Hawai'i at Mānoa, National Foreign Language Resource Center.

Ross, S. (1992). Accommodative questions in oral proficiency interviews. *Language Testing, 9*, 173–186. DOI: 10.1177/ 026553229200900205.

Ross, S., & Berwick, R. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition, 14*, 159–76.

Schegloff, E.A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*, 361-82. DOI:10.2307/413107.

Schegloff, E. A., Koshik, I., Jacoby, S., & Olsher, D. (2002). Conversation analysis and applied linguistics. *Annual Review of Applied Linguistics*, *22*, 3-31. DOI: 10.1017/S0267190502000016.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 8*, 289-327. DOI:10.1515/semi.1973.8.4.289.

Seedhouse, P., & Egbert, M. (2006). The interactional organisation of the IELTS speaking test. *IELTS Research Reports*, *6*, 161-206. IELTS Australia, Canberra and British Council, London.

Seedhouse, P., & Harris, A. (2011). Topic development in the IELTS speaking test. *IELTS Research Reports*, *12*, 55-110. IELTS Australia, Canberra and British Council, London.

Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics, 11*(2), 133-158. DOI: 10.1111/1473-4192.00011.

Stivers, T. and Sidnell, J. (2005). Introduction: Multimodal interaction. *Semiotica. 156*(1/4), 1–20. DOI:10.1515/semi.2005.2005.156.1.

Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning*. The Netherlands: John Benjamins Publishing.

Sydorenko, T., Maynard, C. & Guntly, E. (2014). Rater behaviour when judging language learners' pragmatic appropriateness in extended discourse. *TESL Canada Journal, 32*(1), 19-41.

Take IELTS, British Council. Retrieved from http://takeielts.britishcouncil.org/prepare-test/understand-test-format/speaking-test, September 2015.

Wong, J., & Waring, H. Z. (2010). *Conversation analysis and second language pedagogy: A guide for ESL/EFL teachers*. New York: Routledge.