

The Interactive Demonstration of Interactional Competence in the ITA TEACH Test

Santoi Wagner

University of Pennsylvania

This paper reports on a work-in-progress that examines a high-stakes teaching and oral language performance assessment known as the TEACH test, in which international teaching assistant (ITA) candidates are required to demonstrate their competency in a panel-rated teaching simulation. This study draws upon the analytic tools of Conversation Analysis to examine 29 videotaped TEACH tests recorded at a large Research 1 (R1) university in the eastern United States with the objective of providing a more in-depth understanding of the demands placed on test-takers and raters. The initial analysis shows how prospective ITAs manage questions that engender problematic responses, and how audience participation can be supportive of the demonstration of competence. In the conclusion, directions for future work are discussed.

Introduction

Many graduate students who come to the United States to pursue their studies teach introductory undergraduate courses in their field of scholarship as part of their graduate funding. The 1980s and 1990s saw a wave of concern expressed by undergraduate students and their parents regarding what they considered to be the lack of English language ability of these international teaching assistants (ITAs). Lippi-Green (2012) offers a common student perspective: "Of course it's hard to understand them, and of course I resent it. Why can't I get what I pay for, which is a teacher like me who talks to me in my own language that I can understand?" (p. 91). The response in many states was legislation that established language requirements for instructors. In Pennsylvania, the English Fluency in Higher Education Act of 1990 requires institutions of higher education to certify annually that all instructors (who teach undergraduate credit courses, including graduate teaching assistants) are "fluent in the English language in the classroom" (English Fluency in Higher Education Act, 1990). This fluency is to be assessed using "varied and appropriate criteria, such as personal interviews, peer, alumni, and student observations and evaluations, publications, professional presentations, tests, or any other appropriate criteria which effectively evaluates fluency" (ibid.) This paper reports on a work-in-progress that examines the TEACH test, a panel-rated teaching simulation that seeks to assess the oral language and teaching skills of prospective ITAs.

What Is the TEACH Test?

In assessing the English language ability of ITAs, many U.S. universities require a special assessment of spoken language (Briggs, 1994). There are various instruments employed across institutions, but many of the oral testing procedures include a form of teaching simulation. For the university in this study, if prospective ITAs fail to meet institutionally mandated standardized test cut-off scores, they must take a credit-bearing ITA ESL training course. The course focuses on English language skills, pedagogical skills, and American university classroom culture and norms. The culminating assessment for the course is the TEACH test. This is an audience-rated teaching simulation with the prospective ITA teaching a mini-lesson consisting of a seven-minute presentation in their field of expertise, followed by a two to three minute Q&A session with the audience. The audience generally consists of faculty, representatives from the ITA Program, and ESL teachers from the university's English Language Program, with an average of seven panelists for each TEACH test session. The skills that are assessed are language presentation skills (e.g., comprehensibility, grammar, and word choice) and teaching skills (e.g., lesson organization, interaction with students, and non-verbal communication).

The attraction of the TEACH test is that it offers face validity as it seemingly meets the criteria of situational and interactional authenticity (Bachman, 1990). In other words, it purportedly provides an environment, and elicits the type of language, that approximates a real classroom. However, to my knowledge, there are few micro-analytic studies of the discourse produced in this type of teaching simulation performance assessment. Without a greater understanding of what interactional work is being accomplished during a TEACH test, it seems difficult to make claims regarding authenticity. For the purposes of this paper, I will focus on the type of interactional work that is being done when prospective ITAs are required to demonstrate linguistic and teaching competency. My objective is to provide a more in-depth understanding of the demands placed on test-takers and raters in this assessment. I begin with a background to the study that outlines the analytic framework and relevant literature. I then briefly describe the data and method. In the analysis section, I show how the prospective ITAs and the audience cope with problematic responses to audience questions, and how audience participation can be supportive or not to the prospective ITA.

Background

In this study, I take a practice-oriented view of interaction and competence (Hall, 1995; Young, 2003), in which competence resides not in the individual but within a social and jointly constructed context. Interactional competence is a further elaboration of second language (L2) knowledge (Young, 2003) and is co-constructed by all participants through interactive practices, that is, recurring episodes of talk that are of sociocultural significance to a community of speakers. Individuals do not acquire a general, practice-independent communicative competence, but rather a practice-specific interactional competence. The development and demonstration of interactional competence in the professions is a growing area of research (Nguyen, 2006) in which, being an expert means performing knowledge and skills effectively in social practices (Lave & Wenger,

1991) rather than simply possessing knowledge. Expertise is constituted via the performance of knowledge and skills in social practices, and so can be renewed, negotiated, and resisted during social interaction (Jacoby & Gonzales, 1991). Hence, novice professionals need to develop the interactional competence to construct themselves as experts in interaction.

The TEACH test is a performance assessment structured as a teaching simulation, in which a prospective ITA takes on the role of the teacher and the audience take on the role of undergraduate students. To my knowledge, the only conversational-analytic study that has looked at a comparable assessment is Theodoropoulos (2012). However, in that assessment, the audience consisted of actual undergraduate students who did not evaluate the prospective ITA's performance. The evaluation was the responsibility of a separate rating panel. Theodoropoulos (2012) claimed that the discourse in the test resembled interviewer-interviewee talk. Hoekje and Linnell (1994) evaluate the authenticity of three instruments used to evaluate the spoken language proficiency of non-native English speaking teaching assistants: the SPEAK (Spoken Proficiency English Assessment Kit) test, the ACTFL OPI (American Council on the Teaching of Foreign Languages, Oral Proficiency Interview), and the authors' own IP (Interactive Performance) test which consisted of a ten-minute mock teaching presentation. In comparing the test method facets (Bachman, 1990) and the type of discourse elicited by the three tests, Hoekje and Linnell (1994) conclude that the IP is preferable "because its tasks engage the speaker's language competence in ways similar to the tasks of the target-use context" (p. 122).

There is a larger body of discourse-analytic work on the interaction that takes place within a closely related assessment: the oral proficiency interview (OPI) (e.g., Brown, 2003; Lazarton, 1996; Young & He, 1998). This OPI work has cast a critical eye on the assumption that such interviews are representative of natural conversation. Another important thread in the literature examines variation in examiner behavior and its possible impact on candidate performance through accommodation (e.g., Ross, 1992; Ross & Berwick, 1992), the provision of feedback and the formulation of questions (Brown, 2003), and non-verbal and paralinguistic variables (Jenkins & Parra, 2003; Plough & Bogart, 2008). The work on OPIs, however, has tended to focus on the dyadic interview task in which an examiner interviews the examinee.

There are fewer studies that examine the role-play tasks found in some types of OPIs. Van Lier (1989) aligned role-playing ability with acting ability, and thus posited that role-play was not an appropriate vehicle to assess conversational ability, as the skills needed for role-play were not necessarily congruent with those present in conversation. In contrast, Kormos (1999), using data from the Hungarian English Oral Proficiency Exam, argued that role-play was a suitable means by which to assess conversational ability based on more equal nature of candidates' participation vis-à-vis the examiner in role-play tasks compared to interview tasks. More recently, Okada (2010) examined the construct validity of role-play activities in OPIs, and argued that despite the asymmetric relationship between the candidate and the examiner, the role play is an appropriate assessment task because the interactional competencies required by candidates to participate in the role-play are similar to those employed in ordinary conversation.

In looking at role-plays outside of the second language assessment field, there are a number of informative studies that move beyond the dyadic context to multiple-person role-plays. Sharrock and Watson (1985) investigate the use of role-play (they use the terms “simulations” and “games”) in an EFL classroom in France, and show how members’ activities give rise to the “realities” of the frame (p. 198). Although roles are assigned, there is no scripted talk of what precisely should be said. Instead, players use their cultural knowledge of membership categories to produce category-bound activities (Sacks, 1972). In a follow up study, Watson and Sharrock (1988) came to a similar finding regarding the centrality of category-bound activities in the use of role-play as a teacher-training device for English teachers. Francis (1989) recognized the interactional complexities of role-playing, and how it “trades upon everyday social life” (p. 68) in his examination of an extended business negotiation game for sales personnel of a national newspaper in a business negotiation course. He observes how the participants must manage two layers of identities: the “identities within the game” being the roles assigned to the participants as part of the role-play, and the “game context identities” being the training context of the role-play (Francis, 1989, p. 58). He terms this the “double-setting character of simulation/gaming” and suggests it “could cause specific problems of interactional and communicative adjustments” (ibid.). Stokoe (2013) specifically addresses the under-researched question of whether role-played interaction is authentic, that is, whether it mimics what happens in real life. She compares two datasets: the opening sequences of actual police investigative interviews and role-played simulations that are used for training. She found that on a gross level, the formulation of the same actions occurred (e.g., the identification of present parties) but there were differences in their design and organization. In particular, some role-play actions were more exaggerated or elaborate to be “interactionally visible” (p. 183), displaying the police officers’ orientation to the training assessment context.

Given that there is little direct work on the talk produced in TEACH test type performance assessment, my initial guiding questions in this study were conceived to be wide-ranging and to capture the complexity of the speech event: (1) In what ways do ITAs construct themselves as competent in the TEACH test? (2) What is the nature of the role played by the audience in the TEACH test?

Data and Method

The dataset currently consists of 29 TEACH tests from an R1 university in a major city on the east coast of the United States. The prospective ITAs come from a variety of disciplines (e.g., computer & information science, engineering, physics, chemistry, math, and economics being the most common), and L1 backgrounds (e.g., Chinese, Hindi, Korean), though STEM fields and Asian L1s are the most common. The TEACH tests are video-recorded as part of the standard procedure for the test. In the recordings, only the prospective ITA is visible, so the audience can only be heard and not seen. The tests were transcribed in their entirety using a common Conversation Analysis system (see Appendix A). I have tried to balance the detail necessary for the presentation of my analysis with readability. Because the audience is not visible in the recording, I have identified audience members using the notation AF1 (audience female 1), AM1 (audience male 1), and so on.

However, these designations are not constant across the different TEACH tests, that is, AF1 in the TEACH test for ITA1 is not necessarily the same rater as AF1 in the TEACH test for ITA2.

In this study, I primarily employ a conversation analytic approach. Conversation Analysis (CA) is concerned with the study of the order of human interaction and seeks to explicate how participants, through a set of tacit common procedures, produce, understand and deal with talk-in-interaction. CA works with naturally occurring interaction that is audio and/or video recorded and rendered into transcripts that attempt to preserve the interactional details of the talk as it is produced. Many analysts subscribe to the ideal of *unmotivated looking* (Psathas, 1995), that is, to approach the data without a specific agenda or preformulated hypothesis in mind. The analytical focus concentrates on what social action is being done in the interaction, and how it is accomplished. It is grounded in the displayed understandings and orientations of the participants themselves, through the close examination of each turn's production and receipt, with attention to interactional details such as sequential ordering, word choice, syntax, timing, and intonation (for an introduction, see Hutchby & Wooffitt, 2008).

Analysis

The TEACH test is a role-play in which reality is suspended, and the prospective ITA and audience are expected to take on their "identities within the game" (Francis, 1989), of teacher and students. Although the roles in the TEACH test are preassigned, there is no predetermined script. Thus, role-plays are locally managed interactional orders, where the contingency of interaction and its locally oriented character are just as significant as in any authentic social interaction (Francis, 1989). Even in the seven-minute presentation phase, for which many of the prospective ITAs have prepared a rehearsed lesson, this talk is rarely realized as an uninterrupted monologue because questions by the audience do occur. I will consider first how prospective ITAs and the audience cope with problematic responses to audience questions, and then I move onto how different forms of audience participation can be either more or less supportive.

Managing Problematic Responses

As a performance assessment, the TEACH test seeks to assess the oral proficiency and teaching skills of prospective ITAs. Obviously, the candidate's prime concern is to demonstrate the skills that are valued by the raters. According to the TEACH test rubric, an important element of demonstrating competent interaction skills with students is that candidates *encourage questions*¹, and provide *clear, concise, and relevant* answers. In addition, they should provide *appropriate feedback* and responses should be *friendly and non-judgmental*. In this section, the particular interactional task I will consider is how prospective ITAs attempt to demonstrate competency in line with the TEACH test rubric when they encounter questions that engender problematic responses. The two salient discourse features found in the data set that I will describe here are: (1) the use of the positive

¹ Descriptors (in italics) are taken from the TEACH test rubric.

assessment to an audience question, such as “It’s a great question,” or “that’s a good question”; (2) the yes-no interrogative, “Do I answer your question” or “did I answer your question.” These features are present throughout the dataset in the presentation and Q&A phases. They can appear independently of each other in both phases or occur in the same question-answer extended sequence.

In excerpt 1 below, the prospective ITA, identified as ITA7, employs both features. Her TEACH test topic is the tools of monetary policy in the United States. The time limit for the presentation phase of the TEACH test has ended, and the timekeeper, AF3, has called time (line 1).

Excerpt 1: ITA7 Central Bank

01 AF3: Time,
 02 (2.0) ((ITA finishes writing on blackboard))
 03 ITA7: Okay (.) any: questions?
 04 (1.0)
 05 AF?: [()] ((ITA7 looks at AF? but then turns to AF2
 06 as AF? stops talking))
 07 AF2: [Is this] (.) type of thing happening in other
 08 countries?= >Or you you< just mentioned the United
 09 States,
 10 ITA7: Mm [mm
 11 AF2: [Is this () happening in your country? or::
 12 (0.6)
 13 ITA7: → Mmmm, ah:: it’s a great question. And uh un-until now
 14 we talk about the the general conditions in the United
 15 States. And the Fed is the central bank in Unite-
 16 central bank system in the United States. In my
 17 country (.) I:: am from China and there is a similar
 18 system too. And uh the but we don’t have Fed. We have
 19 central bank. Central bank. That is the People’s Re-
 20 the Chinese the People’s (.) Re- public pe- the
 21 People’s (.) bank of (.) China. That is ou:r central
 22 bank
 23 (0.2)
 24 in China.
 25 AF2: °Mmmm°
 26 ITA7: → Do I answer your question?
 27 AF2: Yea:h sure did

In line 3, ITA7 produces a transitional “Okay” (Beach, 1993) that acknowledges what precedes it—the call for time—while prefacing what follows, “any: questions?” This type of yes-no question formulation is found in almost every case in the data set in the sequential environment at the boundary between the presentation and Q&A phases. This activity boundary, where the “time” call is made by the timekeeper, marks the presentation phase of the TEACH test as being over, and that the Q&A phase should begin. In contrast to how “any questions” is commonly used in classrooms as an understanding check and possible activity closing (Waring, 2012), here “any questions?” functions as an activity opening, indicating that the prospective

ITA is ready for the Q&A phase and opens the floor to the audience members to ask questions.

In lines 7-9 and 11-12, AF2's questioning turn relates to the use of the three tools of monetary policy by the Federal Reserve that ITA7 has been describing in the presentation phase. ITA7 begins her response with the vocalized hesitation markers "Mmm, ah::", followed by a positive assessment of AF2's question, "it's a great question" (line 13). This precedes the substantive content of her answer (lines 13-24) in which, rather than directly answering the question of whether the tools of monetary policy described earlier are also employed in her home country, she provides the name of the central bank in China. The answer is not produced fluently; there are numerous cut-offs (lines 15, 19, 20), hesitation markers (lines 13, 18), and disjunctions (line 18). One may speculate on the reason for the difficulties in providing the answer. The self-initiated self-repair (Sacks, Schegloff, & Jefferson 1974) in line 15 suggests that ITA7 is attempting to draw a nuanced difference between the U.S. central bank system and the Chinese central bank. In any case, the latter half of the extended turn is taken up by the effort to proffer the name of the Chinese central bank.

Before going on to examine how AF2 responds to ITA7's answer, I want to consider what the positive assessment, "It's a great question," (line 13) accomplishes. Firstly, it is overtly doing positive evaluation or praising of the audience question. In the particular context of the TEACH test, by using "It's a good question," the prospective ITA is demonstrating certain skills that are valued on the assessment rubric, in which effective interaction is conceptualized as *encouraging questions from students, providing appropriate feedback, and being friendly in their response*.² However, in light of the most common sequential environment in which it is found, that is, preceding a response that is problematic in content and production, it seems that additional interactional work is being tackled (cf. Waring, 2008). As the sequence unfolds, we see that while ITA7 does produce a response to the question, the positive assessment delays the content segment of the response by pushing it back further into ITA7's turn. It allows ITA7 to continue after her initial receipt tokens "Mmm, ah::" (line 13) and thus attends to the progressivity of the talk.

Another brief example will further illustrate this interactional work. In excerpt 2 below, ITA9 is in the presentation phase of her lesson on injective functions. Audience member AM1, asks a question in lines 1-3 to clarify his understanding of an injective function using the example the ITA has discussed earlier. Here, despite what seems to be a clear agreement, "Yes right" (line 4), ITA9 then encounters difficulty in proceeding with her explanation, displayed by cut-offs "er we-" and "only-" (line 4). She then delivers the positive assessment "that's a- that's a good question" (lines 4-5), before continuing to formulate an answer, "we sit different seats in airplane but seat doesn't make us (0.8) sit down uh on seat" (lines 5-8), that is quite opaque in meaning.

Excerpt 2: ITA9 Airplane seats

```
01 AM1: >Does this mean< that you and your friends (.)
02 sit in seats in the airplane the airplane
03 doesn't (.) put its seats on top of you.
```

² Descriptors (in italics) are taken from the TEACH test rubric.

04 ITA9: → Yes right. (Er we-) Only- that's a- that's a good
 05 question but (1.0) this means (.) we sit different
 06 seats in airplane but seat doesn't makes us (0.8) sit
 07 down uh on seat.
 08 [Okay? Okay. Then]
 09 [(looking at blackboard)]
 ((turn continues))

Similar to excerpt 1, the positive assessment attends to progressivity by enabling the prospective ITA to coherently proceed with her talk, and project that there is indeed a forthcoming response to the audience question. Again, we see that what follows the positive assessment is problematic as a substantive answer to the question.

How does the audience respond to answers such as this? Unfortunately, since there is no visual data of the audience, only verbalized audience turns are available. In excerpt 2 above, there is no audible acknowledgment from AM1. Of course, it is possible that he made a physical gesture of acknowledgement, but it is interesting to note that line 8, as ITA9 delivers “Okay? Okay,” she is not looking at AM1 but at the blackboard, and thus is not overtly inviting AM1 to acknowledge receipt or adequacy of the answer. In other cases, though, the problematic nature of the answer is made demonstrably relevant by the audience member who asked the question. In returning to excerpt 1, recall that the answer provided by ITA7 provides the name of the central bank in China rather than whether the tools of monetary policy are the same in China as they are in the United States. The relevant segment is produced below.

Excerpt 1a: ITA7 Central Bank

13 ITA7: → Mmmm, ah:: it's a great question. And uh un-until now
 14 we talk about the the general conditions in the United
 15 States. And the Fed is the central bank in Unite-
 16 central bank system in the United States. In my
 17 country (.) I:: am from China and there is a similar
 18 system too. And uh the but we don't have Fed. We have
 19 central bank. Central bank. That is the People's Re-
 20 the Chinese the People's (.) Re- public pe- the
 21 People's (.) bank of (.) China. That is ou:r central
 22 bank
 23 (0.2)
 24 in China.
 25 AF2: °Mmmm°
 26 ITA7: → Do I answer your question?
 27 AF2: Yea:h sure did

AF2 orients to ITA7's answer as inadequate. The first sign of inadequacy is seen in lines 22-24. ITA7 reaches a possible completion point in her turn, and this is followed by a 0.2 second acoustic silence, before she adds a grammatically fitted increment, “in China.” There may be some analytic ambiguity to whether the 0.2 second pause is hearable by ITA7 as a relevantly missing acknowledgement by AF2 of ITA7's answer. On the one hand, this seems to be related to the practice of using the repair of an indexical reference in transition space to pursue a response (Bolden & Mandelbaum, 2012). There seems to be little room for not recognizing

the referent for “ou:r” (line 21), especially since ITA7 has spent considerable effort in naming the bank accurately, and that the name of the bank includes the name of the country “China.” Bolden and Mandelbaum (2012) argue that indexical repair is a means for the speaker to treat the missing response “delicately” (p. 144), by masking the lack of an immediate response, and shaping the problem as one of referent recognition rather than “more interactionally charged issues” (p. 153).

In line 25, AF2 displays minimal uptake of ITA7’s answer, with “Mmm” produced at low volume and continuing intonation, communicating a sense of non-committment. This precipitates, in line 26, a yes-no interrogative from ITA7, “Do I answer your question?” This, together with the grammatically correct variant, “Did I answer your question?” are often paired, as in this case, with “It’s a great question,” but can appear independently. They tend to occur after relatively extended ITA responses that are problematic in some way and to which there has been little audience uptake or reaction. “Do I answer your question?” is a yes-no interrogative (YNI), and is responded to positively in line 27 by AF2, “Yea:h sure did.”

What, we might ask, is gained from venturing the YNI in this position? Based on the non-committal receipt token by AF2, there is a sense that the original question was not answered satisfactorily, and given that this is a teaching assessment, prospective ITAs are likely to be conscious of the need to pursue student understanding. The YNI makes relevant an additional response from AF2 (possible responses could be: “No, I still don’t understand” or “So, does the People’s Bank of China use the same tools as the Fed?”) and thus offers ITA7 another opportunity to provide a more appropriate or extended answer. Alternatively, it also provides an opportunity for AF2 to upgrade the minimal response initially given. This is what occurs when AF2 responds with “Yea:h sure did” (line 27). The YNI enables this particular Q&A sequence to be closed in a manner that may be more likely to be evaluated positively by the audience because AF2 has stated that her question was answered to her satisfaction. In the current data set, this type of YNI is overwhelming answered positively with no instances of type-conforming negative responses (Raymond, 2003) such as “No, you didn’t.” However, there are a few instances in which the audience member pursues a further response from the ITA.

An example is given in excerpt 3 below. ITA4’s topic is superconductors, and he had begun his talk by drawing an analogy between friction and electrical resistance, explaining how a blackboard eraser being pushed along a table will eventually stop because of friction, and how electrons moving along a wire will also meet friction, that is, resistance. During the Q&A phase, AF2 asks a question relating to the difference between electrical resistance and friction (lines 1-2).

Excerpt 3: ITA4 Resistance

01 AF2: What’s the difference between the electrical
02 resistance and friction?
03 ITA4: Oh it’s a very good question and I I uh I I just
04 just make a comparison between the friction and
05 between the electrical resistance because the
06 electrical resistance uh is is a little li:ke um
07 li:ke the friction uh when object move in in
08 here ((moves the eraser along the table))
09 because what- (0.4) as electricity is: is

10 produced by electron=by moving electron in in
 11 the wire. So when- as that you can imagine that
 12 uh when electron moves in the in the wire it
 13 will also meet er hav- hav uh the friction.
 14 It's it's not friction it's (.) resistance.
 15 So we we call this kind of resistance as (.)
 16 sup- uh electrical resistance.
 17 (1.8)

18 ITA4: Do you- do I answer your question?

19 AF2: → °Yeah I think so°. So the >electrical resistance is
 20 always for the flow of electricity< (.) the friction
 21 is for the flow- movement of objects?

22 ITA4: >Yuh yuh yes<. It has uh similar meaning but it's not
 23 same.

24 AF2: Okay. °Thanks°.

ITA4 provides an extended response (lines 3-16) that arguably does not clearly distinguish between the two concepts (although electrical resistance shares some conceptual parallels with the mechanical notion of friction, they are not completely analogous). His answer is met with a 1.8 second silence. In lines 19-21, AF2 responds positively to ITA4's "Do you- do I answer your question?" but "Yeah" is produced at low volume, and is modified by "I think so," also at low volume, that displays a degree of uncertainty. AF2 continues at louder volume, and partly at a faster pace, to produce a yes-no declarative (YND) that makes relevant a confirmation (Raymond, 2010) by ITA4 that her understanding of resistance vis-a-vis friction is correct. AF2's pursuit of a further response through the vehicle of YND requires less work from ITA4 than the original, open-ended question in lines 1-2 because it only requires confirmation or disconfirmation by ITA4. Confirmation is forthcoming in lines 22-23, and AF2 closes the sequence. Even in this case when there is a pursuit of a more adequate answer from an audience member, it is formulated in a way that is less demanding than the original question.

In this section, I have provided some preliminary analysis with regard to how prospective ITAs manage problematic responses to audience questions through the use of "It's a good question" and "Do I answer your question?" What is noticeable is that this management is not accomplished by ITA actions alone. *Managing* in this way necessitates the collaboration of the audience in general and the original questioner in particular in that they do not engage in a sustained pursuit for a further or more adequate response. Thus, the ability to successfully cope with questions from the audience is not a skill that can be entirely traced to the individual ITA, a finding clearly noted from work in OPIs (e.g., Brown, 2003; Lazarton, 1996).

Audience Participation

In discussing the audience member's work in excerpt 3 above, I noted that the formulation of the follow up questioning turn gives rise to a confirmation, a response that requires less interactional effort than the open-ended question. In this section, I will focus more closely on how the participation of the audience

may support the prospective ITA in their performance of competency, first in terms of the questions that they ask, and second, in demonstrations of their own knowledge. To begin, I return to excerpt 1, with the relevant segment reproduced below as excerpt 1b, to consider the questioning turn taken by AF2 in lines 7-8. Recall that ITA7's topic is the tools through which monetary policy is carried out in the United States. AF2's questioning turn consists of two contiguous questions on the same topical point within the same sequence-initiating turn.

Excerpt 1b: ITA7 Country question

```

07 AF2:    →  [Is this ] (.) type of thing happening in other
08           countries?=>Or you you< just mentioned the United
09           States,
10 ITA7:    Mm [mm]
11 AF2:           [Is this ( ) happening in your country? or::
12           (0.6)

```

Here, in line 7, AF2 produces a YNI first pair part of the *adjacency pair*, “[Is this] (.) type of thing happening in other countries?” An adjacency pair is a sequence of two turns produced by different speakers and ordered such that a particular first pair part requires or makes conditionally relevant a particular second pair part (Schegloff, 2007). In this case, the adjacency pair is a question-answer sequence. The first pair part, “[Is this] (.) type of thing happening in other countries?” makes relevant a yes or no answer from ITA7 as the second pair part. However, even though a possible completion point has been reached (Sacks, Schegloff, & Jefferson, 1974), AF2 does not relinquish the floor, but builds a multi-unit turn via use of the rush-through (Schegloff, 1996). She produces the next three lexical items at a faster pace in line 8, “>Or you you<”, and that are latched to the previous lexical item “countries?” She overlaps ITA7’s “Mmmm” to produce another YNI in lines 11-12 “Is this () happening in your country?” and then the turn is completed with “or::,” produced as a trail-off (Local & Kelly, 1986) in that it is delivered with a slower speed, and designed to be pragmatically complete. There is an ensuing 0.6 second silence signaling that from AF2’s perspective her turn is complete and hence ITA7 can respond. It is also perhaps an indication that ITA7 may have difficulty in understanding that AF2 is yielding the floor. In sum, by designedly constructing two contiguous questions on the same topical point, AF2 does not relinquish the floor after the first question, and thus ITA7 is not given the opportunity to answer until both questions have been asked.

Another example is from ITA8, whose topic is algorithms. Excerpt 4 below begins just after time has been called at the end of the presentation phase.

Excerpt 4: ITA8 Disciplines

```

01 AUD:           ((sparse [ applause]))
02 ITA8:           [((applauds))]
03                (2.0)
04 ITA8:           Uhm, any (quick) $questions?$
05                (3.0)

```

06 AF4: → Uhm ((clearing throat)) algorithms, in which um
 07 disciplines are they used=is it only in math? or are
 08 they used in other disciplines.
 09 ITA8: Uhohm you mean the [algorithms.=
 10 [((underlines the word "algorithms"
 11 on BB))
 12 AF4: =Ye::s

13 ITA8: This algorithm is actually an abstract idea (1.0)
 14 That means if in computer science (.) you can c-
 15 consider it as uh (0.2) a solution to some problem.
 16 Actually this this idea can be also applied to other
 17 uh area like economics? [()]

18 AF4: [Oh=]

19 ITA8: =Yeah.

20 AF4: → [Thank you].

21 ITA8: [(You got it)] Okay (you're) welcome.

In lines 6-8, AF4 initiates her turn with a hesitation marker (and possibly attention-getter), "Uhm," and then topic fronting "algorithms," followed by an open question first pair part "in which um disciplines are they used." Similar to AF2 in excerpt 3, she reaches a possible completion point after "used" in line 7, but rather than allowing ITA8 to respond, she holds the floor with the latch of "used" with "is" to produce an immediate reformulation that is an alternative choice question "only in math? or are they used in other disciplines." Again, two contiguous questions are produced in a way that suggests they are proactively designed as being contiguous, rather than being generated by a delay in response, any observable repair initiation, or a problematic initial response (cf. expanded question sequences in Gardner, 2004; vertical multiple questions in Kasper & Ross, 2007).

The multiple questions in the two excerpts above occur at the very beginning of the Q&A phase, when the audience is tasked with asking questions related to the topic that has presented in the previous seven minutes. Multiple questions do occur in the presentation phase itself, and in these cases, they tend to relate to topics that could be perceived as tangential to the preceding content. In both the presentation and Q&A phases, then, they occur in environments where the interaction can be topically disjunctive, and in which there may be few contextual cues available in the immediately preceding talk. As such, they bear a great deal of similarity to the horizontal multiple questions in OPIs, described by Kasper and Ross (2007), which are used as a preemptive technique to enable relevant responses in interactional spaces that are particularly vulnerable to trouble such as topic changes or shifts.

In the TEACH test context where the appropriate handling of audience questions is important, multiple questions seem to provide greater opportunities for prospective ITAs to generate an acceptable response. They are a possible proactive means by which the audience is able to decrease the likelihood of the ITA displaying a lack of requisite knowledge. In excerpt 1, the second question "Is this () happening in your country?" has a narrower focus in terms of the

knowledge required to answer, and additionally, the ITA is likely to be familiar with the situation in her own country, compared to the first question, “[Is this] (.) type of thing happening in other countries?”. The trail-off “or” that completes the turn (in line 11) has a generalizing impact in that it re-opens the possibility of a wider response range, retroactively mitigating the need to specifically respond to the previous question about ITA7’s own country. Its overall impact is to give ITA7 a broad choice in how she might answer. Linell, Hofvendahl and Lindholm (2003) find similar practices in examining multiple unit questions across various institutional settings, and also argue that, “narrowing (particularizing) and broadening (generalizing) components of [multiple unit questions] serve the purpose to secure a response” (p. 564).

However, it is not just narrowing or broadening the components of the questioning turn that help the prospective ITA. The additional information provided as part of the questioning turn is also a way of guiding the ITA to an appropriate answer. In excerpt 4, AF4 moves from an open question in lines 6-8 “in which um disciplines are they used” to an alternative choice question “only in math?” or “in other disciplines.” The choices given offer candidate answers (Pomerantz, 1988), which *model* for the respondent what sort of answer would be satisfactory, and thus, increases the probability that the ITA will provide an acceptable response. Svennevig (2013) also demonstrates that the reformulation of questions with candidate answers by the institutional representative in social welfare consultations promotes understanding and participation by their clients who are non-native speakers of Norwegian and guide the clients to a relevant response.

In the above, I have argued that the type of multiple questions asked by the audience is a means through which the audience can guide and assist the prospective ITA in producing talk that can be favorably assessed. The second type of audience participation that I will consider can sometimes engender less beneficial interactional environments for the demonstration of ITA competency. Excerpt 5 below is again taken from the TEACH test for ITA7 whose topic is the Federal Reserve’s tools of monetary policy. Here I will show two instances in which audience members demonstrate their own knowledge of the topic under discussion and how these can vary in their interactional consequences for the ITA.

Excerpt 5: ITA7 The Fed

01 AM1: Have these policies always been in place for as long
 02 as the Fed has been around?=or (.) are they more
 03 recent.
 04 ITA7: Hmm::: (.) they:: usually use the three tools are
 05 actually the uh (.) uh usually along the time are used
 06 along the time (0.2) not just recently.
 07 AM1: → They’ve been used since the like (0.2) seventeen
 08 eighties when the [Fed] started=
 09 ITA7: [Ye:s]
 10 AM1: =Oh okay.=
 11 ITA7: =Actually (0.2) it is right.

12 AM1: °Okay°.
 13 ITA7: °Okay°.((nodding))
 14 AM2: → That's what Alexander Hamilton'ss (.) trying to do.
 15 AM1: °Mhmm°.
 16 (1.8)

In lines 1-3, AM1 produces a question-answer sequence using a multiple question format similar to those described earlier, relating to how long monetary policies have been used. If we consider the production of the question sequence as it unfolds, we see that adequately answering “Have these policies always been in place for as long as the Fed has been around?” presupposes knowledge of how long the Federal Reserve has been in existence. This historical knowledge may or may not be within the purview of the prospective ITA, but AM1 seems to orient to the potential unknowing by latching “or (.) are they more recent” (lines 2-3), a much less specific time period, allowing a greater latitude in what would be perceived as an accurate answer. Although ITA7 displays some hesitation markers (“Hmm:” in line 4, micro pauses in lines 4 and 5, and a 0.2 second pause in line 6), she displays her knowledge that the tools have been used for a while, “not just recently” (line 6). Interestingly, in lines 7-8, AM1’s third position turn demonstrates that he has some content knowledge of the topic. The tentativeness signaled in line 7 through the use of the approximation marker “like” and the 0.2 second pause, of how long the Fed has been around and thus how long the tools have been used, “since the like (0.2) seventeen eighties,” contributes to the utterance being hearable not as an assertion of knowledge but as seeking confirmation, which ITA7 does in partial overlap, and, eventually, the sequence seems to come to a close in line 13.

Audience knowledge is again demonstrated in line 14, with AM2’s turn, “That’s what Alexander Hamilton’s (.) trying to do.” However, the way this is done seems rather different. The declarative is produced with falling intonation, contributing to it being hearable as an assertion rather than as a question or a request for confirmation. In addition, lexical meaning is not transparent as there is no clear antecedent for “That’s.” Therefore the connection to the preceding talk is dependent on one’s content knowledge (that Alexander Hamilton was the first Secretary of the Treasury and created the government-owned Bank of the United States). Although the intended addressee of AM2’s utterance cannot be determined as AM2 is not in the video frame and so no visual indication is available, what is notable is that there is a minimal response in line 15 by AM1, the audience member last to have spoken, with the receipt token “Mhmm,” that minimally acknowledges the prior utterance, produced at low volume and falling intonation. There is no response by ITA7. After a 1.8 second pause, a question on a different topic is raised by a different audience member, and the topic of Alexander Hamilton is not revisited.

In both instances, the demonstration of audience knowledge, in displaying familiarity, interest, and involvement in the topic serves as a means to establish common ground with the prospective ITA, and offers, for example, the possibility of extending or furthering the discussion on this point, or even creating affiliation between the two parties (Svennevig, 2013). There does, however, seem to be a potential downside, as illustrated in the second instance. Across the data set, in certain situations in which audience members engage

in sequence-initiating turns, there is no response by the prospective ITA or there are responses made by other audience members. In the latter cases, the prospective ITAs often display difficulties in reestablishing a foothold in the on-going talk. The parameters of these situations remain to be examined more closely in future work. However, for this specific example, although one might argue that the issue at hand may be a lack of content knowledge (i.e., about Alexander Hamilton), I posit that ITA7 struggles with understanding the action that AM2's utterance about Alexander Hamilton is accomplishing. In the context of a teaching and language assessment, the framing of the utterance may be opaque for the prospective ITA. Ross (1998) argues that frame, as understood in the Goffmanian sense, impacts how questions in OPIs are understood: if an examinee takes an information value frame, then questions are taken to necessitate answering literally. In contrast, if an examinee takes an assessment frame, then questions are taken to be elicitation for talk about a relevant topic. Within the setting of a typical classroom, student-initiated contributions and student-student discussions may be valued and positive occurrences, as they enable students to demonstrate and co-construct knowledge. In allowing this type of participation, classroom teachers reserve the right, however, to self-select and take a turn at any time, and thus regain control of the floor. In the TEACH test context, this ability is complicated by the dual roles played by the participants: the prospective ITA as teacher and examinee, and the audience members as students and examiners.

Discussion and Conclusion

In this work-in-progress report, I have taken a micro-analytical perspective to examine the talk-in-interaction in a teaching and language performance assessment. I have considered how prospective ITAs and the audience cope with problematic responses to audience questions, and how different forms of audience participation can be, to varying degrees, supportive of the ITA's demonstration of competency. At its heart then, the TEACH test is a collaborative enterprise undertaken by the prospective ITA and the audience. In this section, I will discuss three issues raised by the initial analysis and point to areas that call out for further research: first, there is a tension between the demonstration of pedagogic language for assessment purposes and the pursuit of audience understanding of content; second, that the nature of audience participation raises questions with regard to issues of power, congruency of role-play and real world behaviors, and variation across TEACH tests; and third, that the time constraint is a factor that deserves more attention.

Throughout the data set, there are clear attempts by prospective ITAs to display the teaching and interactional behaviors that are to be assessed by the audience according to the TEACH test rubric. Prospective ITAs practice teaching simulations during the ITA course and have access to the grading rubric in order to prepare for the TEACH test. Specific elements in the descriptors, such as making reference to materials previously learned, connecting content to contexts that are familiar to students along with providing personal examples that students can relate to, checking that students understand, being friendly, and using effective gestures and facial expressions, are evident

throughout prospective ITA performance. This washback (Bachman, 1990) is advantageous: by exposing prospective ITAs to skills that are valued in the American classroom and giving practice in their use, it is more likely that candidates will have a successful ITA experience.

However, it is noticeable throughout the dataset that some of the displayed behaviors are overt and very frequent, contributing to a sense of exaggeration. Positive assessments, such as “It’s a good question,” are commonplace and at times seem to appear indiscriminately as a response to audience questions that are patently not “good.” The understanding check, “Do I answer your question?” is often posed, but despite a lack of audience uptake, prospective ITAs do not undertake further explication. Stokoe (2013) suggests that more exaggerated or elaborate behaviors in police investigative interview role-plays are performed so as to be “interactionally visible” (p. 183) and hence available for assessment. Luk (2010) describes how in L2 oral proficiency assessments, conducted via peer group interactions, discourse features were “ritualized, contrived, and colluded” and this behavior was driven by students wanting to appear as “effective interlocutors for scoring purposes” rather than in the pursuit of authentic communication (p. 25). Although a similar phenomenon may be borne out by the data here, its occurrence should not be taken as a criticism of ITA behavior: it is a reasonable strategy for prospective ITAs given that their futures as ITAs are dependent on demonstrating that they are aware and capable of such behaviors. However, it may be problematic from a teaching as well as assessment perspective if a candidate displays a desirable strategy mechanically or indiscriminately, even if it is in the test rubric, without taking into account where it is situationally appropriate (Luk, 2010) or understanding its pedagogic value. It is in this sense then that there may be an underlying tension between displaying overt pedagogic language or behavior, and the pursuit of appropriate, sound pedagogical objectives, such as the student understanding of content.

The complex nature of audience participation in the TEACH test stems from the two normative frameworks that participants were required to orient to: their “identities within the game” of ITA teacher and students, and the “game context identities” of ITA examinee and examiners (Francis, 1989, p. 68). One clear concern is the potential difference in relative power relationships between roles in the two frameworks, and how this might impact the interaction that occurs within the role-play. At certain moments within the role-play, all participants are concurrently orienting to a particular framework. For example, when time is called at the end of the presentation or Q&A phase, participants are orienting to being examinee and examiners. This congruency in orientation is not static or given. When a faculty member asks a question that the prospective ITA considers irrelevant or tangential to the topic at hand, what normative framework is made demonstrably relevant? The impact of asymmetries of power within testing situations is an underexplored area (Plough & Bogart, 2008) ripe for future research.

Another related concern is audience behavior in the role-play vis-à-vis the real world. Briggs (1994) claims that the limited background knowledge of the audience is beneficial as they are able to realistically play the role of the undergraduate student in an introductory class, and that they are “quite adept in asking the type of question a student might ask about the material” (p. 75). However, she continues that sometimes they are “too naïve” but, unfortunately,

without further explication of exactly what this entails. This is an important issue because congruence between the type of questions and general behavior that examiners produce and that of real-world students would support face validity and task authenticity (Bachman, 1990). Although there may be cases in my dataset in which questions from the audience do seem to fall into Briggs's category of "too naïve" in terms of discipline content knowledge, perhaps more consequential are differences in the design and format of the questions from an audience of examiners compared to those from actual undergraduates. This is an issue that needs to be investigated empirically. Some insight can be gleaned from Theodoropoulos (2012). As described earlier, this study investigates a prospective ITA interactive performance assessment with undergraduates as audience members. He found that the interaction between a student and the ITA was analogous to an interview, comprising of two-part question-answer sequences, with the student taking the role of the interviewer and the prospective ITA the role of the interviewee.

As shown by the analysis, the role of the audience is key: first, the management of problematic answers involves the collaboration of the audience in not engaging in a sustained pursuit for a further or more adequate response, and second, the design of multiple questions is such that the audience can guide and assist the ITA in producing talk that can be favorably assessed. In other words, prospective ITAs' display of the type of talk that fulfills the test rubric criteria is an interactional product of prospective ITAs' and the audience's participation. If audience participation varies across different TEACH test sessions (and it appears to do so), then this obviously has consequences in the consideration of test validity. Another observation that calls for further investigation is that more questions tend to be asked of prospective ITAs who appear more proficient in their language and teaching ability. Weaker candidates tend to be asked fewer questions. This may be due to the audience attempting to avoid asking questions that the prospective ITA may have difficulty in answering, but it seems problematic from an assessment standpoint, as fewer questions may reduce the opportunities for the prospective ITA to demonstrate their competency.

Finally, I turn my attention to an underlying factor that seems to deserve greater attention: the limited time that is available for the prospective ITA to demonstrate competency in the TEACH test. In contrast to simulated encounters that exist for training where the mock interviews are often longer than typical real world interviews (Stokoe, 2013), the TEACH test affords only ten minutes, a greatly reduced amount of time compared to real world teaching. The impact of this limited time is seen throughout the test. For example, many teaching behaviors, even if present, are abbreviated forms of what would appear in a real teaching environment. Although similar actions might be present, there may be a difference in their design and organization, and it is through these details in which competency may be displayed. Time concerns may play heavily in how audience members ask questions and how prospective ITAs answer them. If the audience asks too many questions in the presentation phase, the prospective ITA may not finish the prepared lesson. Hence it would be difficult to assess lesson organization and content development. Prospective ITAs may avoid extended answers, tangential topics, or engaging in long repair sequences when responding to questions. Again, the consequences of the time factor is an

empirical question for further research. However, despite the obvious practical considerations, it seems that extended opportunities to observe language use and teaching would be helpful.

Although much of the analysis is in its early stages, it is hoped that this report of a work-in-progress demonstrates the interactional complexity and challenges of the TEACH test. In addition, through highlighting some of the key issues to be explored, I seek to underscore the need for further research in this area. By deepening our understanding of the demands placed on prospective ITAs, we can better serve and support international graduate students who enrich the linguistic and cultural diversity of many American campuses.

Acknowledgements

I presented an earlier version of this paper at the annual meeting of the Language and Social Interaction (LANSI) Working Group, in New York, NY, in 2013. I am grateful to the members of LANSI for their thoughtful insights in our data analysis sessions. I also thank the WPEL reviewers and editors for their helpful comments during the revision process.

Santoi Wagner is Lecturer in Educational Linguistics at the University of Pennsylvania. Her research interests lie at the interface of language and social interaction. In particular, she is concerned with how interactional competence is enacted in the institutional orders of mediation and education.

References

- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.
- Beach, W. (1993). Transitional regularities for “casual” “okay” usages. *Journal of Pragmatics*, 19, 325-352.
- Bolden, G., & Mandelbaum, J. (2012). Pursuing a response by repairing an indexical reference. *Research on Language and Social Interaction*, 45, 137-155.
- Briggs, S. (1994). Using performance assessment methods to screen ITAs. In C. G. Madden & C. L. Myers (Eds.), *Discourse and performance of International Teaching Assistants* (pp. 63-80). Alexandria, VA: TESOL.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20, 1-25.
- English Fluency in Higher Education Act of 1990. Retrieved from <http://www2.legis.state.pa.us/WU01/LI/BI/BT/1989/0/SB0539P1913.pdf>
- Francis, D. (1989). Game identities and activities: Some ethnomethodological observations. In D. Saunders & D. Crookall (Eds.), *Communication and simulation: From two fields to one theme* (pp. 53-68). London, UK: Multilingual Matters.
- 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, UK: Continuum.
- Hall, J. K. (1995). (Re)creating our world with words: A sociohistorical perspective of face-to-face interaction. *Applied Linguistics*, 16(2), 206-232.

- Hoekje, B., & Linnell, K. (1994). "Authenticity" in language testing: Evaluating spoken language tests for international teaching assistants. *TESOL Quarterly*, 28, 103-126.
- Hutchby, I., & Wooffitt, R. (2008). *Conversation analysis*. Cambridge, UK: Polity Press.
- Jenkins, S., & Parra, I. (2003). Multiple layers of meaning in an oral proficiency test: The complementary roles of nonverbal, paralinguistic, and verbal behaviors in assessment decisions. *The Modern Language Journal*, 87, 90-107.
- Jacoby, S., & Gonzales, P. (1991). The constitution of expert-novice in scientific discourse. *Issues in Applied Linguistics*, 2(2), 149-181.
- Kasper, G., & Ross, S. J. (2007). Multiple questions in oral proficiency interviews. *Journal of Pragmatics*, 40, 2045-2070.
- Kormos, J. (1999). Simulating conversations in oral-proficiency assessment: A conversation analysis of role-plays and non-scripted interviews in language exams. *Language Testing*, 16, 163-188.
- Lazarton, A. (1996). Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing*, 13, 151-172.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Linell, P., Hofvendahl, J., & Lindholm, C. (2003). Multi-unit questions in institutional interactions: Sequential organization and communicative functions. *Text*, 23, 539-571.
- Lippi-Green, R. (2012). *English with an accent: Language, ideology, and discrimination in the United States*. London, UK and New York, NY: Routledge.
- Local, J., & Kelly, J. (1986). Projection and "silences": Notes on phonetic and conversational structure. *Human Studies*, 9, 185-204.
- Luk, J. (2010). Talking to score: Impression management in L2 oral assessment and the co-construction of a test discourse genre. *Language Assessment Quarterly*, 7, 25-53.
- Nguyen, H. (2006). Constructing 'expertness': A novice pharmacist's development of interactional competence in patient consultations. *Communication & Medicine*, 3, 147-160.
- Okada, Y. (2010). Role-play in oral proficiency interviews: Interactive footing and interactional competencies. *Journal of Pragmatics*, 42, 1647-1668.
- Plough, I. C., & Bogart, P. S. H. (2008). Perceptions of examiner behavior modulate power relations in oral performance testing. *Language Assessment Quarterly*, 5, 195-217.
- Pomerantz, A. (1988). Offering a candidate answer: An information seeking strategy. *Communication Monographs*, 55, 360-373.
- Psathas, G. (1995). *Conversation analysis: The study of talk-in-interaction*. Thousand Oaks, CA: Sage.
- Raymond, G. (2003). Grammar and social organization: Yes/no interrogatives and the structure of responding. *American Sociological Review*, 68, 939-967.
- Raymond, G. (2010). Grammar and social relations: Alternative forms of yes-no-type initiating actions in health visitor interactions. In A.F. Freed & S. Ehrlich (Eds.), *'Why do you ask?': The function of questions in institutional discourse* (pp. 87-107). Oxford, UK: Oxford University Press.

- Ross, S. J. (1992). Accommodative questions in oral proficiency interviews. *Language Testing*, 9, 173-186.
- Ross, S. J. (1998). Divergent frame interpretations in oral proficiency interview interaction. In R. Young & A. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 333-354). Amsterdam, Netherlands: John Benjamins.
- Ross, S., & Berwick, R. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14, 159-176.
- Sacks, H. (1972). On the analyzability of stories by children. In J.J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication* (pp. 325-345). New York, NY: Rinehart and Winston.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696-735.
- Schegloff, E. A. (1996). Turn organization: One intersection of grammar and interaction. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 52-133). Cambridge, UK: Cambridge University Press.
- Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge, UK: Cambridge University Press.
- Sharrock, W. W., & Watson, R. D. (1985). 'Reality construction' in second-language-learning (L2) simulations. *System*, 13, 195-206.
- Stokoe, E. (2013). The (in)authenticity of simulated talk: Comparing role-played and actual interaction and the implications for communication training. *Research on Language and Social Interaction*, 46, 165-185.
- Svennevig, J. (2013). Reformulation of questions with candidate answers. *International Journal of Bilingualism*, 17, 189-204.
- Theodoropoulos, C. (2012, April). *Investigating the "interactive" in the Interactive Performance Test (IPT): A conversational analysis approach*. Paper presented at the annual conference of the Language Testing Research Colloquium, Princeton.
- Van Lier, L. (1989). Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly*, 23, 489-508.
- Waring, H. Z. (2008). Using explicit positive assessment in the language classroom: IRF, feedback, and learning opportunities. *The Modern Language Journal*, 92, 577-594.
- Waring, H. Z. (2012). "Any questions?": Investigating the nature of understanding-checks in the language classroom. *TESOL Quarterly*, 46, 722-752.
- Watson, R. D., & Sharrock, W. W. (1988). Some social-interactional aspects of a business game for special purposes in the (L2) teaching of English. In D. Crookall (Ed.), *Simulation-gaming in education and training* (pp. 177-186). Oxford, UK: Pergamon.
- Young, R. (2003). Learning to talk the talk and walk the walk: Interactional competence in academic spoken language. *Working Papers in Linguistics*, 2, 26-44.
- Young, R. F., & He, A. W. (1998). *Talking and testing: Discourse approaches to the assessment of oral proficiency*. Amsterdam, Netherlands: John Benjamins.

Appendix A

Transcription convention (modified from Hutchby & Wooffitt, 2008)

.	falling intonation
?	rising intonation
,	continuing intonation
-	abrupt cut-off
::	prolonging of sound
<u>no</u>	stressed syllable or word
↑	marked rising intonational shift
↓	marked falling intonational shift
WORD	loud speech
° word°	quiet speech
> word <	quicker speech
< word >	slower speech
hhh	speaker out-breath
.hhh	speaker in-breath
ha, heh, huh	laughter
[]	overlapping speech
=	latching of utterances
(1.4)	length of a time gap in seconds
(.)	micro-pause (< 0.2 seconds)
(word)	doubt over transcribed speech
((tapping))	description of non-speech sound or transcriber comments on contextual or other features
→	indicates analytic focus