

INTERACTION IN THE ORAL PROFICIENCY INTERVIEW: PROBLEMS OF VALIDITY¹

Marysia Johnson²

Abstract

This article reports on the findings of a discourse analysis study whose purpose was to provide answers to the following research question: *What kind of speech event is the OPI? Is it more like an everyday, friendly conversation, an interview, or something else?*

To obtain some answers to this research question, a discourse analysis of 35 OPIs was conducted. Frequency counts and chi-square inferential statistics were performed on the transcribed and coded OPI data. The results of the discourse analysis were then compared with the prototypical features of the two speech events selected for the study (i.e., conversation and interview) to determine and describe the most typical discourse and linguistic features of the OPI.

The findings of the discourse analysis contradict the ETS claim that “a well structured OPI tests speaking ability in a real life context - conversation” (ETS 1989). The OPI does not test speaking ability in a real life context - conversation. It tests speaking ability in the context of an interview, or more precisely, in the context of a survey research interview, which is based on the behavioristic theory of stimulus and response (Mishler 1986). These findings raise a question about the validity of the OPI testing instrument.

Keywords: Oral proficiency interview, Conversation, Speech event

1 Introduction

Advocates of the Oral Proficiency Interview (OPI) claim that “a well-structured oral proficiency interview tests speaking ability in a real-life context - a conversation. It is almost by definition a valid measure of speaking ability” (ETS 1989). However, much of how everyday conversation works is deceptively obvious so that people studying and testing language often overlook fundamental characteristics of conversation and in the process violate them. It is precisely on these grounds that van Lier (1989) has challenged the ETS claim that it measures speaking ability in the context of a conversation. Van Lier was the first to pose the question: “Is it really a conversation?” (Van Lier 1989: 494).

¹ From a forthcoming book with Yale University Press.

² This paper is based on the findings of my Ph.D. dissertation. I would like to thank my Ph.D. dissertation committee members: Jeff Connor-Linton, Deborah Schiffrin, and Eduardo Cascallar for guidance, support, and knowledge. I would like to thank John Staczek for his valuable comments and suggestions.

This paper reports on the findings of a discourse analysis study whose purpose was to provide answers to the following research question: *What kind of speech event is the OPI? Is it more like an everyday, friendly conversation, an interview, or something else?*

To obtain some answers to this research question, a discourse analysis of 35 OPIs was conducted. Frequency counts and chi-square inferential statistics were performed on the transcribed and coded OPI data. The results of the discourse analysis were then compared with the prototypical features of the two speech events selected for the study (i.e., conversation and interview) to determine and describe the most typical discourse and linguistic features of the OPI.

The findings of the discourse analysis contradict the ETS claim. The OPI does not test speaking ability in a real life context - conversation. It tests speaking ability in the context of an interview, or more precisely, in the context of a survey research interview, which is based on the behavioristic theory of stimulus and response (Mishler, 1986). These findings raise a question about the validity of the OPI testing instrument.

2. The oral proficiency interview

2.1. Overview of the OPI

The Oral Proficiency Interview (OPI) is a widely used instrument for assessing second and foreign language speaking ability within government institutions such as the Foreign Service Institute (FSI) and Defense Language Institute (DLI), and nongovernment institutions such as the Educational Testing Service (ETS) and the American Council on the Teaching of Foreign Languages (ACTFL).

In the OPI, which is based on the ACTFL/ETS/ILR scale and speaking level descriptions, the examinee converses face to face with one or two trained testers on a variety of topics for 10 to 30 minutes. The elicited sample is then rated on a scale ranging from 0 (no functional ability) to 5 (proficiency equivalent to that of a well-educated native speaker). "Plus" ratings are also assigned when the examinee's speaking proficiency exhibits features of the next higher level. Overall, the examinee's level of speaking proficiency is measured on an 11-point scale (Lowe 1983, 1986).

It is estimated that several thousand OPI tests are administered each year. Frequently, entire professional careers, future job assignments, pay-increases, and entrance or exit from college language programs depend on the rating obtained during the oral interview.

2.2. Structure of the oral proficiency interview

The OPI has both a general and a level-specific structure. The OPI consists of four phases: *Warm-up*, *Level Checks*, *Probes*, and *Wind-Down*. The purpose of the *Warm-up* phase is to give the testers a preliminary indication of the candidate's level of speaking proficiency. This preliminary indication must be confirmed in the next phase of the interview, the *Level Check*.

The purpose of the *Level Check* is to find out the candidate's highest sustainable

level of speaking proficiency. When the candidate successfully passes the *Level Check*, his/her performance provides a floor for the rating.

The purpose of the *Probes* phase is to show the tester(s) whether the candidate has reached his/her highest level of speaking proficiency. The last phase is the *Wind-Down*. The purpose of this phase is to leave the candidate with a feeling of accomplishment.

2.3. Elicitation techniques

To obtain a ratable sample (i.e., the sample to which a rating can be assigned), testers must make sure that not only general, but also level specific requirements have been fulfilled. Level-specific requirements include a series of tasks and functions that are assigned to a given level. To elicit level-specific tasks and functions, testers use questions and role-play situations.

A variety of question types constitute the core of the OPI elicitation procedures. A given set of question types is recommended for a particular level or levels of speaking proficiency. Thus, for instance, for *Level 0+*: *Yes/No* and *Choice Questions* are recommended; for *Level 1* and *2*: *Information Questions*; for *Level 3, 4* and *5*: *Hypothetical* and *Supported Opinion* questions are required. The examples of these question types are presented below (Lowe 1988):

<i>Yes/No Questions:</i>	Do you live in Washington?
<i>Choice Questions:</i>	Would you like tea or coffee?
<i>Information Questions:</i>	What did you do last summer?
<i>Hypothetical Questions:</i>	If you were the Prime Minister, what would you do to improve the economic situation in your country?
<i>Supported Opinion Questions:</i>	Why are you against this type of policy?

3. Literature review

3.1. Critical analysis of the OPI

A group of researchers (Bachman 1988, 1990; Savignon 1985; Lantolf and Frawley 1985, 1988; Johnson 1997; Johnson and Tyler 1998) working in the field of language testing and teaching have voiced strong criticism of the OPI. Their criticism centers primarily on the issue of the OPI's validity, and the theory of proficiency the OPI claims to represent.

Another critic of the OPI, van Lier (1989), also calls for a thorough investigation of the OPI's validity. However, contrary to the authors mentioned above, he does not insist on developing an external criterion (i.e., a theoretical framework) against which the construct of the OPI should be evaluated. He calls instead for a thorough examination of the OPI from *within*. He advocates an ethnographic approach to determine what kind of speech event the OPI is, to find the answer to his original question: "Is it really a conversation?"

Van Lier finds it difficult to accept that the OPI represents instances of natural

conversation because the ultimate goal of the OPI is to elicit a ratable sample (i.e., a sample to which a rating can be assigned), and not to conduct a conversation. Agreeing with the Jones and Gerard (1967) model of dyadic interaction, van Lier points to different distributions of rights and duties in interviews and in conversations. An interview is characterized by asymmetrical contingency (i.e., the interviewer has a predefined plan and conducts the interview to execute the plan). In contrast, friendly, everyday conversation is based on mutual contingency with equal distributions of rights and duties. Van Lier, thus, finds it unclear how the OPI might accommodate these mutually exclusive types of contingency.

Van Lier (1989) also rightfully points out that “every manifestation of speaking is speaking in context; every contextual manifestation of speaking ability requires, in addition to speaking ability, context-specific skills and experiences ” (p. 500). If the OPI does not measure speaking ability in the form of conversation, as it claims to measure, then the users of the system may be misled about the candidate’s ability to actually carry on real life oral interactions.

Furthermore, it may very well be the case that the OPI as a unique speech event has its own unique norms and rules. These rules and norms may not adhere to those of any naturally occurring speech events. In that case, the ability to generalize the OPI scores beyond the OPI context could be severely undermined. Precisely for that reason, the investigation of the question: “Is it really a conversation?” is of the utmost importance.

In order to determine whether indeed the OPI represents an instance of an everyday conversation or an interview, the prototypical features of two speech events most relevant to the finding of this study (i.e., conversation and interview) need to be described.

3.2. Conversation as a speech event

As noted above, supporters of OPI have argued that it is a valid measure of speaking ability because it represents a real-life context - a face-to-face conversation. However, this assertion can only be maintained if careful analyses of Oral Proficiency Interviews show that they contain the established features of natural conversation. I turn now to a consideration of the research which has examined these features.

The various works of Schegloff and Sacks (1973), Sacks, Schegloff and Jefferson (1974, 1978) and Schegloff, Jefferson and Sacks (1977) have firmly established the highly organized and locally managed system operating within conversation. Key features of the locally managed system are systematic turn-taking mechanisms and adjacency pairs. The local management system accounts for patterns of stable and recurrent actions responsible for creating order in conversation.

Examination of everyday conversation reveals it is produced on a turn-by-turn basis. Although the turn-taking is normally accomplished smoothly, virtually no aspect of it is specified in advance. Turn size, turn order, and turn distribution are not fixed. In other words, at the beginning of a conversation the participants do not know and cannot accurately predict how much any one participant will contribute, in which order participants will talk, nor how frequently any one participant will talk. Neither is the content of a participant’s remarks specified in advance. According to Sacks et al. (1974), the turn-taking organization for natural conversation, by contrast with other speech exchange

systems such as interviews, “makes no provision for the content of any turn, nor does it constrain what is (to be) done in any turn” (p. 710). Indeed, the unplanned nature and unpredictable outcomes constitute primary characteristics of natural conversation.

Another salient feature of everyday conversation is a spontaneously created and negotiated topic. Although researchers have failed to agree upon an operational definition of topic, topic can be regarded as a “pre-theoretical notion of what is being talked about” (Brown and Yule 1983: 71) “through some series of turns at talk” (Schegloff 1979: 27). In natural conversation, topic is negotiated, and topical coherence is “constructed across turns by collaboration of participants” (Levinson 1983: 313). We expect topic to emerge spontaneously, and as Ochs points out “it is relatively unplanned and locally managed” (Ochs 1970: 58).

3.3. Interview as a speech event

An interview is considered to be a prominent research method in the social and behavioral sciences. Schiffrin (1994) and Mishler (1986) distinguish between different types of interviews such as survey research interviews and sociolinguistic interviews.

A survey research interview undergoes major scientific scrutiny in Mishler’s (1986) *Research Interviewing: Context and Narrative*, which leads to the establishment of several basic characteristics of that type of interviewing. In Mishler’s opinion, a survey research interview is strongly embedded in the behavioristic theory of stimulus and response where an interview is viewed as a *verbal exchange* rather than a *form of discourse*.

In a survey research interview, questions and answers are regarded as stimuli and responses. All “extraneous material” is suppressed in order that the findings may be generalized to a larger population. This attempt to emulate positivistic, scientific research leads to interpreting each question and response in isolation (i.e. independent of particular features of context). The context is not viewed as an important factor influencing participants’ interaction. The role of the interviewer is to become an expert in stimulus sending, so that the interviewee may become the ideal response-emitter.

Moreover, Mishler (1986) points to another typical feature of survey research interviews: The asymmetrical distribution of power. This asymmetry is evident in the interviewer’s exclusive control over who will speak, when, and for how long (i.e., turn-taking), what topics are discussed, and what is relevant and what is not relevant to the interview.

In contrast to survey research interviews, sociolinguistic interviews allow for a variety of different genres, such as narratives or descriptions outside a question-answer format (Schiffrin 1994). Interviewers are trained to avoid the question-answer format and to elicit different types of talk similar to casual conversation. The role of the interviewer and the interviewee in sociolinguistic interviews is also less rigidly defined. Although the asymmetrical distribution of power still exists in sociolinguistic interviews, the interviewee is allowed to change roles (i.e., ask questions of the interviewers), change and initiate topics, and have greater control over who holds the floor. Different speech events that are allowed to emerge contribute to a better sense of cooperation and solidarity between participants (i.e., asymmetry of power is less evident in sociolinguistic interviews).

4. Analysis

4.1. Data description

For the purpose of this research study, 35 OPIs, 10 each at levels 2, 3, 4, and 5 at level 1 were randomly selected from the pool of one of the federal institutions in Washington, DC. These 35 OPIs were conducted in English by the female and male testers. Originally, 40 OPIs were to be selected (i.e., 10 OPIs for each base level), however, because this particular agency does not keep the records of candidates who do not obtain a level 2 rating, the number of *Level 1* OPIs was reduced to five. All 35 OPIs were audiotaped. Following the agency's testing policy, the OPIs were conducted over the telephone from the testing headquarters in Washington DC. A similar policy of substituting a telephonic OPI for a face-to-face OPI is common practice in all U.S. government language institutions.³

All 35 OPIs were numbered according to the official rating given to the candidate during the OPI. That first digit indicates the level of the candidate's proficiency while the last two digits indicate the number of the tape. Thus, 105 indicates that it is a *Level 1*, tape number 5. All 35 OPI (each lasting from 35 to 50 minutes) were transcribed according to transcription conventions taken from Schiffrin (1994: 431–432) and coded according to a newly developed and validated coding system (see Appendix 2 for details). The pilot study was conducted to determine whether the selected transcription conventions and the coding system reliably captured information relevant to the study.

4.2. Coding system

The development of the coding system proceeded through two major stages. In the first stage, some major categories of the coding system were determined (such as *Turn*, *Topic*, and *Discourse Unit*), and their operational definitions were developed. These major categories were identified to capture the prototypical features of the unmarked form of interaction - conversation against which prototypical features of other speech events could be compared and contrasted. In the second stage, the subcategories of each of the major category were determined and operationally defined in order to provide answers to such questions as:

- Who takes the initiative in the distribution of turns?
- Who holds the floor most frequently?
- Who controls the floor?
- What are the typical features of OPI topics? Is topic spontaneously developed? Is topic negotiated?
- What is the OPI basic discourse unit? Is the OPI discourse unit similar to an everyday, friendly conversation, or an interview, or does the OPI have its own unique structure?

³ Although it is customary to replace a face to face OPI by a telephonic OPI, no research studies have been conducted to investigate the comparability of the tests results obtained on these two types of tests.

5. Findings of the discourse analysis

5.1. Distribution of turn

The results of the discourse analysis indicate that when the tester holds the floor a straightforward *Question* is the most typical *Turn unit*. Another most typical tester's *Turn Unit* is a variation on *Question: Utterance(s) followed by a Question* where the tester sets up a context for his/her question and then asks a question, or a *Question followed by utterance(s)* where the tester poses a question and then elaborates on the posed question:

Tester's Turn Unit Type

Tester	
Question	737
<i>Utterance(s) followed by Question(s)</i>	384
<i>Question(s) followed by Utterance(s)</i>	35
Total	1156

The following represents a sample of each of two variations on a straightforward *Question*:

(1) *Utterance(s) followed by Question(s)* taken from tape 307 (turn 76):

76. INTER: Mm-hm. One thing we often hear in the United States is the fact that many of our best students ones.. who excel in high school and college are Asian Americans. We often are reminded that .. Asian Americans have more of a work ethic than perhaps other students have. **Do you believe this to be true and what could we learn from Asian families to [inspire] our children to study more?**

(2) *Question(s) followed by Utterance(s)* taken from tape 406 (turn 103):

103.INTER: Do you believe that uh uh Would it be a very valid move to make as a society to legalize drugs? Uh by drugs mean of course [marijuana] cocaine heroin.
The idea = [Hm:!]
INTER: = being for example as in the Netherlands people who are addicted to drugs would be treated as a medical problem. They would be perhaps sent to treatment centers or given prescriptions for the drugs they're addicted to and there would be no illicit drug trade.

In contrast, the candidate's most frequently used Turn unit is *Utterance(s)*. Questions are relatively infrequently used by the candidate:

Candidate's Turn Unit Type

Question	179
Utterance followed by Question(s)	29
Question followed by utterance(s)	01
Total	209

The chi-square results ($\chi^2=982.887$, $df=8$ $p < .05$) indicate that there is a predictable pattern as to the tester's and the candidate's behavior in terms of *Turn Unit*. This contradicts one major prototypical characteristic of conversation that the OPI claims to represent. In conversation, turn unit type (along with turn location, turn distribution) is unpredictable; it is not fixed. Based on the results presented above, the typical pattern of *Turn Unit* in the OPI emerges. Whenever there is a question posed, one may predict that the tester poses the question, and whenever there is a response in terms of utterance(s), it will be uttered by the candidate.

Because there are only two participants in the OPIs used in the study, there is no competition for obtaining the floor, as one might have expected in a group interaction. However, there is a pattern of how the floor turn is being distributed in the OPI, which is closely associated with *Turn Unit* discussed previously. The most typical way for the candidate to obtain the floor is to be directly nominated by the tester. When the candidate finishes the response, the turn goes back to the tester who has the right either to accept it or give it back to the candidate to continue. This right does not, however, apply to the candidate. Thus when the candidate returns the turn to the tester he does it indirectly, simply by completing his/her response. The selected portion from the tape 201 illustrates this pattern:

- **15. INTER: How long does it take exactly to get from Salt Lake to Provo?**
16. CAND: I took a bus this morning so it took me about an hour and twenty minutes to get here.
- **17. INTER: Oh you rode the bus?**
18. CAND: Yeah I did.
- **19. INTER: Did they have a good bus service from between the two cities?**
20. CAND:: Yeah they have UTA Utah Transit Service and it's real good.
- **21. INTER: (clears throat) What kind of buses are they uh do they have. Are they big ones?**

In this excerpt, the tester selects the next speaker (i.e., the candidate) directly by asking questions (Turns 15, 17, 19, and 21). The candidate selects the tester indirectly. That is, after completing the response the turn goes back to the tester. The chi-square results ($\chi^2 = 1030.321$ $df=2$ $p < .05$) indicate that there is a predictable pattern of turn distribution in the OPI, which stands in sharp contrast to what one may expect to observe in a real life conversation. This predictable pattern of turn distribution points in the direction of an interview where the interviewer is expected to be in a total control over the turn distribution.

5.2. Topic nomination

Chi-square results for topic nomination ($\chi^2 = 9.552$ $df=3$ $p < .05$) indicate that there is a predictable pattern as to who is most likely to change a topic in the OPI. If a new topic is introduced the person most likely to introduce it is the tester. The tester is in control of changing topic, and thus determining what will be and what will be not talked about in the OPI. Chi-square results ($\chi^2 = 362.312$ $df=7$ $p < .05$) also show that the topic change introduced by the tester is most typically *contrived*, as opposed to the candidate's topic change. The tester tends to follow his/her own agenda. The change of the topic on the tester's part has little to do with what has been discussed previously. The selected portion (i.e., Turns 117,

137, 139, and 145) from the tape 402 illustrates this pattern:

- **117. INTER:** Uh-huh. OK! Good. Uh **if you: were to buy a car uh what kind of a car would you get? Would you buy an expensive one a beautiful one or a well-made one?**
- 137. INTER:** Uh-huh. Good. OK. Very good. Uh next question. Uh we hear a lot about immigration these days in the news uh many problem many uh uh comments by different people. Uh immigrants themselves feel that in this country they're not
- treated well. And they feel like they are victims. **Uh could you tell me why you think uh: uh immigrants feel like they are victims in this country?** They were victims where they came from and [c] now when they come here they feel like victims. Could you tell me why you think they feel that way?
- 139. INTER:** Uh-huh. OK. Uh did you read this uh weekend about that those militant groups uh in our country who uh claim to dislike uh: our government.. And they feel that uh.. uh: they .. uh now these are not immigrants but they're unhappy with the situation
- [c] with the government here. Uh I was thinking uh **do you think that people should stay in this country if they're unhappy with the government? I'm talking about these militant people** [Laughs]. Do you think they should stay if they =
- CAND:** [Laughs]
- **145. INTER:** = [c] It's interesting to hear their comments. Uh OK: uh uh **twenty-five years ago there was a court case called Brown versus teh Board of Education and the phrase that they used a lot was "separate but equal." Uh do you know what that means?**

In this excerpt, after the topic of buying a new car (turn 117), the tester introduces the topic of immigration (turn 137), followed by the topic of militant groups (turn 139) which, in turn, is followed by the topic of education (turn 145). This excerpt seems to indicate that a new topic is *not* negotiated, but controlled by the tester. The tester, *not* the candidate, has a right to say "O.K. I will give you another question." The tester, not the candidate, has a right to introduce a topic that has little relevance to what has been discussed previously. This treatment of *Topic* contradicts the ETS claim that the OPI is conversational in nature and points toward a specific type of interview - a survey research interview.

5.3. Discourse unit type

The findings of the discourse analysis also provide answers as to the most typical discourse unit of the OPIs. The chi-square results ($\chi^2 = 991$ df 42 $p < .05$) show that the OPI most typical discourse unit is *Tester asks a question and candidate answers* (and its variation *Tester presents a task/role play, which replaces a question, and candidate answers*) that was originally developed to capture a prototypical discourse unit of an interview that the OPI may represent. This type of discourse unit is primarily associated with the *Level Check* and *Probe* phases.

6. Discussion

6.1. Level check and probes

The picture that emerged from the quantitative findings of the discourse analysis is that the *Level Check* and *Probe* represent a formal type of interview: *A survey research interview*.

These phases of the OPI seem to represent a *formal behavioristic* verbal exchange (where one party is responsible for providing stimuli and the other responses), *not* a discourse exchange where participants have the same rights and duties (Mishler 1986). In this formal verbal exchange, there is little negotiation allowed on the part of the candidate. The following example illustrates this point.

Tape 403

244. INTER: O.K. let's imagine this say he asks you you talked me you talked about immigration I don't understand that word. Explain that to me daddy.
- 245. CAND: **Well I don't think my son is uh at the stage where he knows about immigration yet.**
246. INTER: Is there any way you can think of for me that that you might try to tell him what immigration is?
247. CAND: **At three years** [old]
248. INTER: [If] he hears you say that word how would you tell him or you could say [you could exp]
249. CAND: [Yeah I know] what you mean like if he keeps uttering that word all the time_z
250. INTER: Yeah_z
251. CAND: and he wants to know the meaning of [c] immigration?
252. INTER: Yeah!
253. CAND: Although I'm not assuming that he will uter that word all the time because

Despite the candidate's objection as to the "validity" of the task presented to him, (i.e., the candidate clearly finds it improbable that his three-year-old son will ask for the explanation of the word immigration), the candidate is forced to respond. His comments regarding the credibility of the task are being totally ignored by the tester who acts as if he has a right to "do" whatever he finds important to his own agenda.

In many instances, the tester's topic selection sounds rather challenging and insensitive. The candidate frequently expresses his/her discomfort (detected in her/his voice, longer pauses etc.) while providing the answer to the tester's stimuli - questions. This is illustrated in turn 177 of another interview:

Tape 205

177. CAND: O.K. all right **uh uh uh**: I think that uh my the first thing that **uh: uh** really comes to my mind when **uh uh I I am I'm uh uh ...** I agree on abortion in case of the following, uh case of rape...**uh**: in case of a very ... after a thorough examination by several specialists that the fetus.. **Uh: uh**: the young fetus that I had or somebody had **uh uh ...** is very deformed and the likelihood that he is the baby will live for y'know a short period of time is very unlikely then I would consider an abortion for that **uh uh**: and the other case is when the mother's **uh uh uh ...** the mother's uh uh: uh life is in danger and o.k. I think **uh ...**

Some testers often explicitly tell the candidate that they do not have to "tell the truth" while expressing their opinions about a given topic as illustrated in turn 170 tape 205: "You don't have to tell me your real views on this O.K. just make them up." This stands in a sharp contradiction to what participants in a real conversation are expected to do: The *Level Check* and *Probes* phases are the most important for the process of the OPI

rating. Some testers even make very explicit remarks that separate this phase from the *Warm-up* and the *Wind-down* phases which seem to have little impact on the outcome of the OPI rating. The following portion of the OPI, tape 303 illustrates this point.

170. INTER: O.K. Good O.K. uh let's go on to another question. O.K. uh this is on abortion O.K. **you don't have to tell me your real views on this O.K. just make them up.** In Your opinion what circumstances should exist before a physician or a doctor suggests to his patient that she should consider an abortion.

The *Level Check* and *Probes* phrases are the most important for the process of the OPI rating. Some testers even make very explicit remarks that separate this phase from the *Warm-up* and the *Wind-down* phases which seem to have little impact on the outcome of the OPI rating. The following portion of the OPI, tape 303 illustrates this point.

Tape 303

58. CAND: O.K. Uh: well I never returned to my original place where I live which is the capitol Mexico City. I never return and (laughs) but I would like to return and visit especially those historical places and remember by culture. It's a beautiful culture there are very beautiful architecture old historical places like the Parliament and all uh those uh places. It's nice to be there and make a trip.

→ **59. INTER: You and I sir are going to be talking in English for just a while a little while this afternoon ..**

After a lengthy *Warm-up* phase, the tester's remarks "You and I sir are going to be talking in English for just a little while this afternoon" sound out of place because they contradict what they have been doing so far. They have been talking in English for some time now, but apparently, it does not have any effect on the process of rating. The "real" talk in English will begin as of the next phase (i.e., the *Level Check*).

The *Level Check* and *Probes* also exhibit another typical feature of a survey research interview - the asymmetrical distribution of power. This asymmetry is evident in the tester's exclusive control over when and for how long the candidate holds the floor, what topics are to be discussed, and what is, or is not relevant to the interview. The following example from the transcribed OPI illustrates this asymmetry of power. The tester has the power to interrupt the candidate in any place he/she considers relevant as **Turn 136** illustrates:

Tape 409

135. : CAND:... educating children about weapons and about violence and giving them statistics and showing them uh movies and the same way that these children are being taught to be violent how carrying weapons to school how gangs are cool.. Well we should show them also the other side of the coin. I believe this should be also part of the school's program. I think it should be part of the education and I think the only way children are going to stop imitating is if they understand we can't force anything into anyone and nobody learns with somebody else's [head]

→ **136. INTER:** [O.K.] let's just shorten it to right here and switch gears just a little ...

7. Conclusion

The findings of this research study contradict the ETS claim. The OPI does not test speaking ability in the real-life context of a conversation. The OPI tests speaking ability in the context of an interview.⁴ Considering the fact that the candidate's level of proficiency is determined within the *Level Check* and *Probes*, the OPI tests speaking ability in the form of a unique type of interview - a survey research interview, which is based on the behavioristic theory of stimuli and responses. This raises the question of the validity of the OPI testing instrument.

From a theoretical point of view, the findings of this study raise another question: *Is the conversational nature of interaction the most desirable or valid measure of speaking ability?* Perhaps most desirable is a combination of several speech events rated separately on different types of rating scales. Perhaps such a combination of different speech events could provide a more detailed and more accurate picture of the candidate's speaking ability. However, the decision as to the format of oral proficiency tests should *not* be based on "intuition" but facts. These facts, in turn, should be based on the findings obtained from such fields as: Interaction analysis, conversational analysis, ethnography of communication, etc.

Within a spirit of cooperation with other scientific fields, all tasks, topics, and role-play situations recommended by the OPI level descriptions should be carefully examined to determine whether indeed they can be most appropriately and efficiently executed within the format of a conversation, an interview, or something else. The nature of interaction, thus, should be thoroughly investigated, described, and understood.

The investigation of the nature of interaction in second language performance assessment has been advocated by many researchers such as Lazaraton (1992), Young and Milanovic (1992), Young (1995a, 1995b), and McNamara (1996). McNamara advocates a broader view of performance in second language performance assessment - the view that "permits renewed focus on the social dimension of interaction" (McNamara 1996: 16). McNamara criticizes current approaches in the second language performance assessment that view interaction as a cognitive and psychological issue not as a *social* issue. He claims that "a danger of too exclusive a focus on defining the nature of candidate's ability in cognitive terms is that the performance is seen as in some way a simple projection of the candidate's ability.... It is as if the candidate is exclusively responsible for the performance, and can be held accountable accordingly" (McNamara 1996: 16).

McNamara (1996) also sees the importance of Vygotsky's sociocultural theory for second language assessment (despite the fact that Vygotsky's theory is a theory of learning, not of performance). Vygotskian theory focuses on the difference between the current state of the individual's development and the individual's potential development as determined by his/her ability to perform problem solving activities in *interaction* with more competent individuals. These differences constitute the basis for Vygotsky's zone of proximal

⁴ The limited space does not allow for a detailed analysis of the *Warm-up* and *Wind-down* phases. However, the findings of the discourse analysis reveal that the *Warm-up* and *Wind-down* phases have a discourse unit in common, which distinguishes them from the *Level-Check* and *Probes* phases. Although the *Warm-up* and *Wind-down* contain more conversation-like features, the controlled nature of turn taking and lack of topic negotiation prevent these two phases from being viewed as an everyday, friendly conversation.

development (ZPD). Clearly, the application of Vygotsky's theory with its focus on the individual's *potential* development would require reevaluation of the existing theories of performance that, in turn, would have some impact on the naming and definition of performance levels in rating scales (McNamara 1996).

Van Lier (1996) also promotes Vygotsky's theory, which should constitute, in his opinion, a theoretical basis for second language acquisition theory and practice. Van Lier's (1996) ideas are relevant to our discussion because of his claims as to one form of interaction. Van Lier (1996) views the symmetrical distribution of power in conversation along with its unplanned, unpredictable nature as an ideal form of interaction for learning. According to van Lier (1996), conversation promotes the necessary ingredient of learning within the ZPD - the movement from the reliance on others (i.e., other-regulation) to the reliance on self (i.e., self-regulation). If van Lier's ideas were to be applied to second language performance assessment, then perhaps conversation would be the most desirable form of interaction.

However, it is important to note that the purpose of this study is not to advocate the superiority or appropriateness of one form of speech event over another. The purpose of this study is to find out what the OPI measures. Also, the purpose of this study is to warn against spreading imprecise information about the nature of the OPI speech event. From a testing perspective, such a misrepresentation can have a profound impact on the users' ability to generalize scores from a testing context to the outside world. For example, since the OPI claims to be conversational in nature, the users of the test may be under the impression that the candidate who obtained a level 2 or higher is able to fully participate in a conversation. That is, the candidate is able to "compete" for the floor, negotiate a new topic, etc. The findings of the study show that the candidate does not have many opportunities to prove the mastery of these skills. The phases in which the candidate may exhibit more initiative - *Warm-up* and the *Wind-down* - are minimized in the process of assigning a global rating. The final rating is based on the candidate's performance within the *Level Check* and *Probes* - the phases wherein the candidate's power to negotiate is almost nonexistent.

From a practical point of view, the findings of the study may be used to improve testers' elicitation techniques, especially within the *Level Check* and *Probe* phases of the OPI. Improvement is urgently needed regarding the so-called lead-in questions (i.e., questions that lead to a given task). Improvement in this area might alleviate the impression that the testers are working from a prescribed set of questions based on their current reading of political and economic news. Testers should also be encouraged to avoid making any evaluative remarks such as: "Very good, let's move to the next question, Very good. We made it. Good, very good. OK I have asked you all the questions from my list." Such remarks clearly point toward a test and *not* a casual conversation.

This study recommends further research in the area of the construct validity of the OPI and language proficiency interviews (LPI). The investigation of construct validity should be based, as Messick (1989) suggests, on empirical evidence and theoretical rationales. In the light of many findings in the field of interaction analysis, conversation analysis, and ethnography of communication, the time has come for the OPI to reexamine the nature of its interaction and to concentrate its efforts on improving the OPI's construct validity. Research into the nature of the OPI's construct validity may not only improve the quality of the OPI testing instrument, but may also prove to be indispensable for designing

better instruments for assessing language speaking ability.

Appendix 1

The following represents the operational definitions of *Floor turn, Topic and Discourse Unit*. The operational definition of a floor turn in this study was based on Goffman’s definition of a turn. Goffman (1981) defined a turn as “an opportunity to hold the floor, not what is said while holding it” (p. 23).

Topic category was created to capture the pattern of topic development in the OPI. In conversation, topic is negotiated, locally managed, and unpredictable. The participants have equal rights to introduce a new topic. There is no one person solely responsible for introducing or changing topics.

Two subcategories of *Discourse Unit* were developed. (1) The subcategory of *Tester asks question(s), candidate answers* was developed to determine if the OPI discourse type unit resembles prototypical interview discourse unit characteristics (i.e., the interviewer asks questions and the candidate provides answers) as exemplified by turn 19-20 taken from the OPI tape 406:

19. INTER: What was the most difficult thing for you the aspect of life in the United States for you to adjust to aside from the language?

20. CAND: Right uh probably the schools because um: I had grown up with a [c] in a Catholic school Catholic German school and the discipline there is it’s extreme it’s it’s very strict and I mean you can’t even write in notebooks you you get a grade of the grade that you are giving uh reflects on um the discipline that you have with the notebooks

In the floor turn 19, the interviewer poses a question, to which the candidate provides a response.

(2) The subcategory of *Tester and candidate alternate in asking questions/making comments* was created to capture a prototypical, unfixed, and unpredictable discourse pattern in conversation as turns 31-33 in tape 209 illustrate:

31. INTER: Oh in Virginia, so am I, I live in Fairfax County.

32. CAND: Hah uh. Me too. I live in uh .. west of Springfield.

33. INTER: Oh that is more or less the area I live in. I’m in Burke.

Appendix 2

Coding

One turn (i.e., turn 109 from the tape 309) was selected to describe in detail the coding process conducted in this study :

109. INTER: Mm-hm all right. And what about health care what would you say about health care if you were a person who wanted more government involvement?

The following illustrates how turn 109, presented above, was coded onto a coding sheet. (For the purpose of this example the coding sheet has been divided into two rows.)

FLOOR TURN	SPEAKER	TURN UNIT	TURN ALLOCATION	REPAIR TYPE	REPAIR LOCATION
109	1	5	2		

NEW TOPIC NOMINATION	TOPIC NATURAL/ CONTRIVED	TOPIC CHANGE METHOD	QUESTION TYPE	DISCOURSE UNIT
1	1	2		1

Turn 109 belongs to the tester. The number 1 under the *Speaker* category indicates this. The number 5 under the category of the *Turn Unit* indicate that the tester asked a question. By virtue of asking a question, the tester selected the candidate as the next speaker. The number 2 under the *Turn Allocation* category indicate this. There is no repair in this turn so the categories *Repair Type* and *Repair Location* do not show any number. The tester initiates a new topic of health care that is reflected by the number 1 under the *Topic Nomination* category. The topic of health care is connected with the topic discussed previously in turns 107 and 108, and therefore the new topic introduced by the tester was coded as 1 (i.e., *Natural*). The method of introducing a new topic was in the form of a question that is reflected by the number 2 under the category of the *Topic Change Method*. The number 1 under the category of *Discourse Unit* indicates that the tester asked a question which was followed by the candidate's response in turn 110.

References

- Bachman, L. (1988) Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition* 10: 149-174.
- Bachman, L. (1990) *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L., & S. Savignon (1986) The evaluation of communicative language proficiency: A critique of the ACTFL oral interview. *Modern Language Journal* 70.4: 380-390.
- Brown, G., & G. Yule (1983) *Discourse analysis*. Cambridge: Cambridge University Press.
- Clark, J.L.D. (1972) *Foreign language testing: Theory and practice*. Philadelphia, PA: Center for Curriculum Development.
- Clark, J.L.D., & R. Clifford (1988) The FSI/ILR/ACTFL proficiency scales and testing techniques: Development, current status, and needed research. *Studies in Second Language Acquisition* 10: 129-147.
- Educational Testing Service (1989) *Oral proficiency testing manual*. Princeton, NJ: Educational Testing Service.
- Hughes, A. (1989) *Testing for language teachers*. Cambridge: Cambridge University Press.
- Johnson, M. (1997) What kind of speech event is the Oral Proficiency Interview: Problems of construct validity. Unpublished Georgetown University Dissertation.
- Johnson, M. (1997) Interaction in the Oral Proficiency Interview. Paper presented at the *American Association for Applied Linguistics*, Annual Conference, Orlando 1997.
- Johnson, M., & A. Tyler (1998) Re-analyzing the OPI: How much does it look like natural conversation? In R. Young & A.W. He (eds), *Taking and Testing: Discourse Approaches to the Assessment of Oral Proficiency*. Amsterdam & Philadelphia: John Benjamins Publishing Company.

- Tyler, A., & M. Johnson (1998) Natural context?: Collaboration, conversational involvement and the OPI. Paper presented at the *American Association for Applied Linguistics*, Seattle 1998.
- Jones, E.E., & H.B. Gerard (1967) *Foundation of social psychology*. New York: Wiley & Sons.
- Lantolf, J., & W. Frawley (1985) Oral-proficiency testing: A critical analysis. *Modern Language Journal* 69.40: 337-345.
- Lantolf, J., & W. Frawley (1988) Proficiency: Understanding the construct. *Studies in Second Language Acquisition*, 10: 181-195.
- Lazaraton, A. (1992) The structural organization of a language interview: A conversation analytic perspective. *System* 20: 373-386.
- Levinson, S. (1983) *Pragmatics*. Cambridge: Cambridge University Press.
- Liskin-Gasparro, J. (1984) The ACTFL proficiency guidelines: Gateway to testing and curriculum. *Foreign Language Annals* 17: 475-89.
- Lowe, P. (1983) The ILR interview: Origins, applications, pitfalls, and implications. *Die Unterrichts Praxis* 2: 230-240.
- Lowe, P. (1986) Proficiency: Panacea, framework, process? A reply to Kramsh, Schultz, and in particular, to Bachman and Savignon. *Modern Language Journal* 70.4: 391-396.
- Lowe, P. (1988) *ILR handbook on oral interview testing*. Washington, DC: DLI/LS Oral Interview.
- McNamara, T. (1996) *Measuring second language performance*. London: Addison Wesley Longman.
- McNamara, T. (1996) *Interaction in second language performance assessment*. Plenary address. American Association for Applied Linguistics Annual Conference. Chicago IL, 26 March 1996.
- Messick, S. (1989) Validity. In R.L. Linn (ed.), *Educational measurement*. New York: Macmillan, pp. 13-104.
- Mishler, E. (1986) *Research interviewing: Context and narrative*. Cambridge, MA: Harvard University Press.
- Ross, S. (1992) Accommodative questions in oral proficiency interview. *Language Testing* 9: 173-186.
- Ross, S., & R. Berwick (1992) The discourse of accommodation in oral proficiency examinations. *Studies in Second Language Acquisition* 14: 159-176.
- Sacks, H., E. Schegloff, & G. Jefferson (1974) A simplest systematics for the organization of turn taking in conversation. *Language* 50.4: 696-735.
- Savignon, S. (1985) Evaluation of communicative competence: The ACTFL provisional proficiency guidelines. *Modern Language Journal* 69: 129-133.
- Schegloff, E., & H. Sacks (1973) Opening up closings. *Semiotics* 8: 287-327.
- Schegloff, E., G. Jefferson, & H. Sacks (1977) The preference for self-correction in the organization of repair in conversation. *Language* 53: 361-82.
- Schiffrin, D. (1987) *Discourse markers*. Cambridge: Cambridge University Press.

- Schiffrin, D. (1990) Conversational analysis. *Annual Review of Applied Linguistics* 11: 3-16.
- Schiffrin, D. (1994) *Approaches to discourse*. Oxford: Blackwell Publishers.
- Shohamy, E. (1982) Affective considerations in language testing. *Modern Language Journal* 66: 13-17.
- Shohamy, E. (1983) The stability of oral proficiency assessment on the oral interview testing procedures. *Language Learning* 33: 527-40.
- Shohamy, E. (1988) A proposed framework for testing the oral language of second/foreign language learners. *Studies in Second Language Acquisition* 10: 165-179.
- Valdman, A. (1988) Introduction to the assessment of foreign language oral proficiency. *Studies in Second Language Acquisition* 10.2: 121-128.
- Van Lier, L. (1989) Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly* 23: 489-508.
- Van Lier, L. (1988) *The classroom and the language learner*. London: Longman.
- Van Lier, L. (1996) *Interaction in the language curriculum: Awareness, autonomy, and authenticity*. London: Longman.
- Young, R., & Milanovic (1992) *Discourse variation in Oral Proficiency Interviews*. Studies in Second Language Acquisition.
- Young, R. (1995a). Conversational styles in language proficiency interviews. *Language Learning* 45.1: 3-42.
- Young, R. (1995b) Discontinuous interlanguage development and its implications for oral proficiency rating scales. *Applied Language Learning* 6: 13-26.