

Review of the ACTFL Oral Proficiency Interview

Timothy Brockley

EDU570: Assignment 1

Professor Kathleen Bailey

September 5, 2009

Word Count: 2109

"Oral proficiency consists of those aspects of communicative competence that are displayed and rated in oral proficiency interviews" (van Lier, 1989, p. 492).

Test Reviewed: ACTFL (Oral Proficiency Interview)

Test Reviewer: Timothy Brockley

Date: September 6, 2009

Intended Purpose:

"The ACTFL Oral Proficiency Interview was developed to evaluate speaking proficiency in a foreign language" (Yoffe, 1997, p.1).

Intended Audience:

According to the official website (ACTFL, 2009:1)

The ACTFL OPI is currently used worldwide by academic institutions, government agencies, and private corporations for purposes such as: academic placement, student assessment, program evaluation, professional certification, hiring and promotional qualification. The ACTFL OPI is recognized by the American Council on Education (ACE) for the awarding of college credit. More than 10,000 OPIs in 37 different languages are conducted through the ACTFL Testing Program.

Ordering Information:

By appointment. Application for testing can be found at:

http://www.languageTesting.com/acad_to_apply.htm

Pricing:

Individual Certified ACTFL OPI: \$134.00

History:

The approach to oral proficiency assessment in the United States is based on The Interagency Language Roundtable (ILR) interview. The ACTFL grew out of the ILR government-based criterion-referenced scale and emerged with a modified rating scale which "involved a refinement of the distinctions at the lower end of the scale in recognition of the need to test individuals with lower levels of proficiency" (O'Connell and Norwood, 2007). It coincides with what is called 'the proficiency movement': "...the OPI (especially the ILR interview) generates so many suggestions for teaching methodology and classroom practices that it has provided the impetus for a clutch of pedagogical recommendations with enough identifiable unity to be called 'the proficiency movement.'" (van Lier 1989, p. 490, 491). In historical terms, the influence the OPI has had on course and curricular design, most notably in the United States, is significant.

Description of the Oral Proficiency Interview:

The ACTFL OPI is a criterion-referenced assessment instrument. It can be administered

directly (in a face-to-face interview) or by telephone with a single trained and certified interviewer. The fact that the ACTFL Oral Proficiency Interview by Computer (OPIc) is also now available raises the level of its practicality (ACTFL, 2009:2):

The ACTFL OPIc is a test of oral proficiency that can be delivered electronically and on demand. This semi-direct test is individualized to the test taker and designed to elicit a 20-30 minute sample of ratable speech delivered via the internet. The candidate's digitally recorded speech is double rated by an ACTFL Certified Rater. An Official ACTFL OPIc Certificate, indicating one of seven possible proficiency levels, is issued.

This method of inter-rater reliability applies to F-to-F and telephonic interviews as well.

There are five stages to the interview: warm-up, level check, probes, role-play, and wind-down. According to Yoffe (1997, p.2):

The role of the 'warm-up' is to put the interviewee at ease, to familiarize him/her with the pronunciation and way of speaking of the interviewer, and to generate topics which can be explored later in the interview. The 'level checks' allow the interviewee to demonstrate his/her ability to manipulate tasks and contexts at a particular level.

If the interviewer is satisfied with the testee's sustained performance, an attempt will be made to discover the 'ceiling', i.e. to elicit response at the higher level. 'Probes', thus, makes the testee reveal a pattern of weaknesses. A 'role-play' serves as an additional check, to help

the interviewer confirm the testee's level. The 'wind-down' brings the interviewer down to a level comfortable for the testee so as to end the OPI on a positive note. The entire interview lasts about 15 minutes in the case of a novice, and can be as long as 35 minutes if a series of probes and level checks are necessary. The interview is taped and a decision is made if the interviewer and a second rater agree on the level. In the case of disagreement, the tape is sent to a third rater.

Proficiency Levels and Sublevels

There are four main levels and three sublevels within the first three main levels (Omar, 2009):

Levels: Novice (0 ~ 1) Intermediate (1 ~ 2) Advanced (2 ~ 3) Superior (3 ~4)

Sub-levels:

Low – just hanging on

Mid – length and strength; some features of the next level

High – functions most of the time at the next higher level

The scoring system works out such that there are actually 11 categories: the superior level has no sub-levels while the other three levels each utilize the 3 sublevels (low, mid, high) and the 11th category is a hypothetical '0' or 'Zero Proficiency'.

There are five aspects to each of the four main levels as well: function, content, context, accuracy and text type (see Appendix B). Yoffe (1997) distils these components in the following manner:

'function' refers to what the learner can do with the language. 'Content' and 'context' refer to the range of topics (personal, professional, and abstract) the learner can handle with confidence and in what setting (formal or informal). 'Accuracy' describes the extent of phonological and syntactical precision. Finally, 'text type' refers to the discourse complexity of the testee, i.e. whether the subject speaks in discrete words, unconnected sentences or extended, planned paragraphs (p.2).

The criteria for assessing test-takers are outlined in the revised *ACTFL Proficiency Guidelines* (Breiner-Sanders et. al, 2000). See Appendix A for a detailed table and view the pdf file here: <http://www.actfl.org/files/public/Guidelines.pdf>

Sample role plays from the test (Omar, 2009):

Intermediate/1 Role play:

You are in city X. You missed your plane to New York. Ask the person behind the counter three or four questions to find out what you need to know to get to New York quickly.

Advanced/2 Role play:

When you arrived at the airport in city Y, your luggage is not in the baggage claim area. You speak with a service representative, explain why you and your luggage did not arrive in the same flight, and make arrangements to have the bags delivered to your hotel.

Superior/3 Role plays:

1) You just received an award for “_____ of the Year”. Make a brief speech accepting this award.

2) You are leading a discussion at your club on a book/movie you read/saw recently. Describe a major theme of the book/movie, and discuss the significance of the theme to society.

Analysis of the test using Wesche’s four components:

Behavior	Format and ACTFL specifications
1) Stimulus material	Interviewer generated questions and role plays
2) Task	Level check, interview (as 'probe'), role play
3) Testee response	It is presumed the testees will respond at their optimal level
4) Criteria	<i>ACTFL Proficiency Guidelines</i> (Breiner-Sanders et. al, 2000). App. A or pdf file here: http://www.actfl.org/files/public/Guidelines.pdf

Evaluation of the Test using Swain’s four principles:

1. Start from somewhere

'Starting from somewhere' means having clearly defined constructs. The problem of construct validity centers around two controversies: one, neither proficiency nor communicative competence (referring to strategic, grammatical, sociolinguistic and discourse) have been defined by ACTFL: "Nowhere in the text, however, is the definition of 'oral proficiency' provided. If we do not know exactly what it is the OPI tests, then any claim of its usefulness as an accurate evaluative mechanism is highly suspect" (Yoffe, 1997, p.3). The problem of defining proficiency is further elucidated by Volmer via Yoffee (1997): "Volmer

(1981, p. 152) defines proficiency quite nicely: 'Proficiency is what proficiency tests measure.'" (p. 7). The argument is circular in nature; it states implicitly that neither universal definitions nor agreed upon parameters have yet been established in defining proficiency as a construct. The question is: Will there ever be?

The second controversy revolves around asymmetrical and pseudo-contingencies. The former contingency manifests as power exerted over the interviewee in terms of interview question formation, discourse trajectory, choice of content (in role plays) and the like, while the latter occurs whilst creating false contexts (as in role plays). It appears that the interview format can neither claim reactive nor mutual contingency, both of which can be seen as hallmarks of conversation and consequently, it could be said, of the related construct of oral proficiency. In Appendix C, van Lier (1989) provides a clear graphic model of contingencies via Jones and Gerard (1967).

2. Concentrate on content

In the following example (van Lier, 1989, p.499) we can see how content initiation in interviews may go terribly wrong:

I: Where is your mother? What does your mother do?
 S: She's dead.
 I: A:h - she's dead. Very good.

Here the interviewer brings up an inappropriate topic accidentally and then fails, on a social level, to address the interviewee (a six-year-old child). Van Lier (1989) rightly observes: "In a sense, in asymmetrical discourse, miscommunication and pragmatic failure are *by definition*

the controlling party's responsibility" (p. 499). It's very likely the child did not perform optimally after this exchange. Of course the interviewer could have followed up with: "I'm so sorry to hear that..." and, depending on the context, the girl may have fared the better.

If we consider Swain's comment "...the materials need to be motivating, substantive, integrated and interactive for the testee" (1984, p. 190) as a litmus test for content, then we run head on into reliability issues (as in the interview example above). The hypothetical argument goes like this: How can such a test be administered equally for everyone, when there may be content and/or socio-cultural differences between interviewer and interviewee? In light of these differences, how can the test claim reliability?

3. Bias for best

"This means do everything possible to elicit the learners' best performance" (Swain, 1984, p.195). The framework of the interview (warm-up, level check, probes, role play, wind down) satisfies this principle to some extent. The performance of the interviewee could be optimally elicited given the relative skill and experience of the interviewer. There is care taken at the beginning and end of the interview to make the testee feel comfortable, but the fact remains, and this is the key argument, that the oral interaction can only accurately measure interview proficiency (or 'performance in context' if you will) and NOT general oral proficiency or conversation skills.

4. Work for washback

Joffe (1997) discusses Lantolf and Frawley's view of the test's negative washback:

They quoted their 1985 article in which they called the OPI 'a criterion-reductive, analytically derived, norm-referenced test of how well an individual can deal with an imposition'... The authors (Lantolf and Frawley) call the OPI theoretically and empirically unsound and believe its institutionalization may be harmful (p. 6).

On the other hand, van Lier (1989) has suggested an academic form of positive washback: "Apart from providing an alternative to traditional tests, the OPI has focused our attention on the construct of proficiency and its relationship to another construct, communicative competence" (p. 490). In a similar fashion, Yoffe confesses a possible pedagogic form of positive washback: "...the washback effect on classroom teaching has been positive as the practitioners place more emphasis on speaking, encouraging student oral production in class" (Yoffe, 1997, p. 10).

In my view, the OPI has widened the scope of testing beyond the straitjacket of standardized norm-referenced multiple choice tests and has, at least, given institutions and test takers a format in which to judge listening and speaking skills as a performance in context.

Conclusion

"Tests, by their very nature, can provide only a limited range of interactive contexts. Designing an evaluative mechanism which would assess all competencies (discourse, grammatical, sociolinguistic, and strategic) in all settings (from very formal to very informal) and on all contents (factual, hypothetical, and abstract) is a virtual impossibility" (Yoffe, 1997, p. 10).

What Yoffe is describing here is the complexity of human interaction. It follows that attempting to test this phenomenon will always remain problematic. He goes on to suggest the possibility of evaluating OPI test scores against assessments in other contexts: "In defense of the OPI, it is perhaps necessary to look at the high positive correlation between the OPI ratings and subsequent measurements of success on job assignments." (Yoffe, 1997, p. 10).

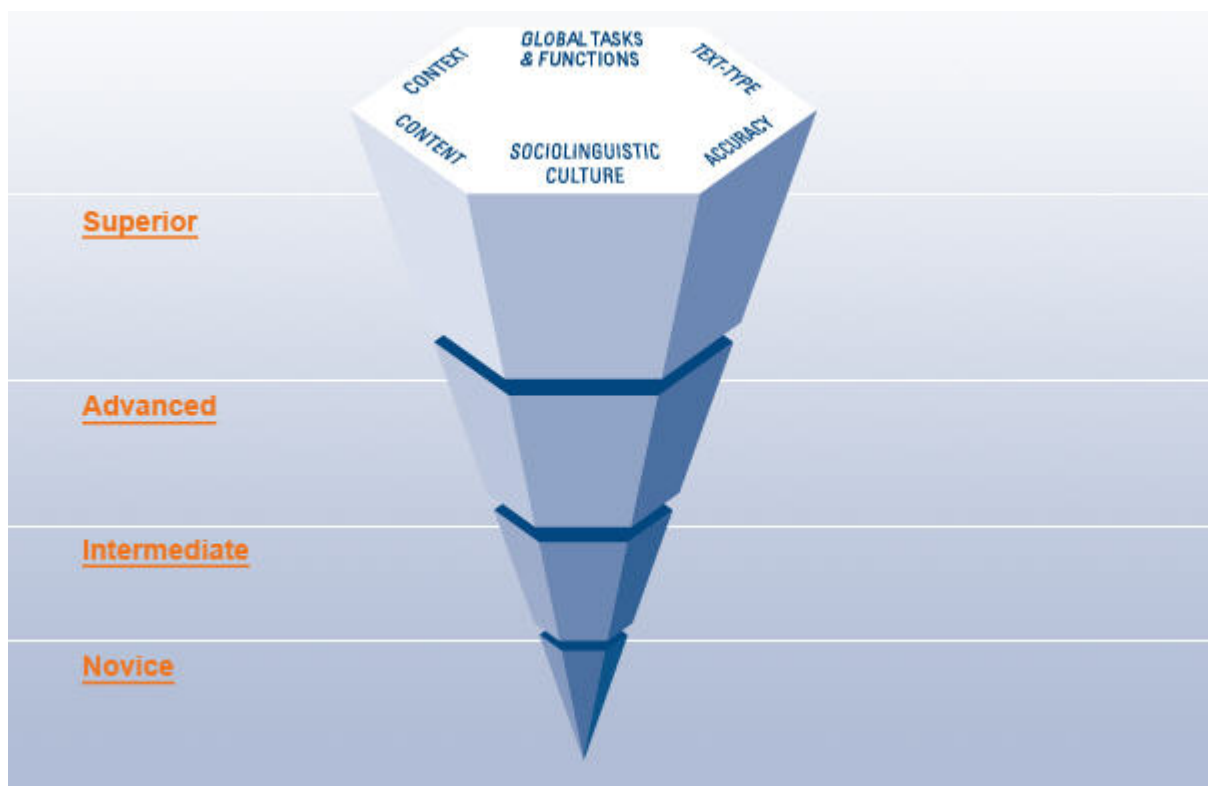
If there is a high correlation between the aforementioned variables, we could place more confidence in OPIs in general, but with guarded optimism: "How does one measure 'success on job assignments'?" and "What is the nature of the correlation?" would be a few of the questions that would require compromise and debate... not to mention the cost in terms of money, time and energy, which certainly raises the specter of practicality.

The fact that an interview is not a conversation and that a role play is based upon pseudo-contingency are the two stones in the shoes of the OPI... what's certain, is that researchers and assessment specialists will continue working (and walking) until they are able to shake those stones loose. And who knows? Someday they may succeed.

Appendix A:

Chart			
SUMMARY HIGHLIGHTS ACTFL PROFICIENCY GUIDELINES—SPEAKING (REVISED 1999)			
SUPERIOR	ADVANCED	INTERMEDIATE	NOVICE
<p>Superior-level speakers are characterized by the ability to:</p> <ul style="list-style-type: none"> • participate fully and effectively in conversations in formal and informal settings on topics related to practical needs and areas of professional and/or scholarly interests • provide a structured argument to explain and defend opinions and develop effective hypotheses within extended discourse • discuss topics concretely and abstractly • deal with a linguistically unfamiliar situation • maintain a high degree of linguistic accuracy • satisfy the linguistic demands of professional and/or scholarly life 	<p>Advanced-level speakers are characterized by the ability to:</p> <ul style="list-style-type: none"> • participate actively in conversations in most informal and some formal settings on topics of personal and public interest • narrate and describe in major time frames with good control of aspect • deal effectively with unanticipated complications through a variety of communicative devices • sustain communication by using, with suitable accuracy and confidence, connected discourse of paragraph length and substance • satisfy the demands of work and/or school situations 	<p>Intermediate-level speakers are characterized by the ability to:</p> <ul style="list-style-type: none"> • participate in simple, direct conversations on generally predictable topics related to daily activities and personal environment • create with the language and communicate personal meaning to sympathetic interlocutors by combining language elements in discrete sentences and strings of sentences • obtain and give information by asking and answering questions • sustain and bring to a close a number of basic, uncomplicated communicative exchanges, often in a reactive mode • satisfy simple personal needs and social demands to survive in the target language culture 	<p>Novice-level speakers are characterized by the ability to:</p> <ul style="list-style-type: none"> • respond to simple questions on the most common features of daily life • convey minimal meaning to interlocutors experienced with dealing with foreigners by using isolated words, lists of words, memorized phrases and some personalized recombinations of words and phrases • satisfy a very limited number of immediate needs
© ACTFL, Inc., 1999			

Appendix B:

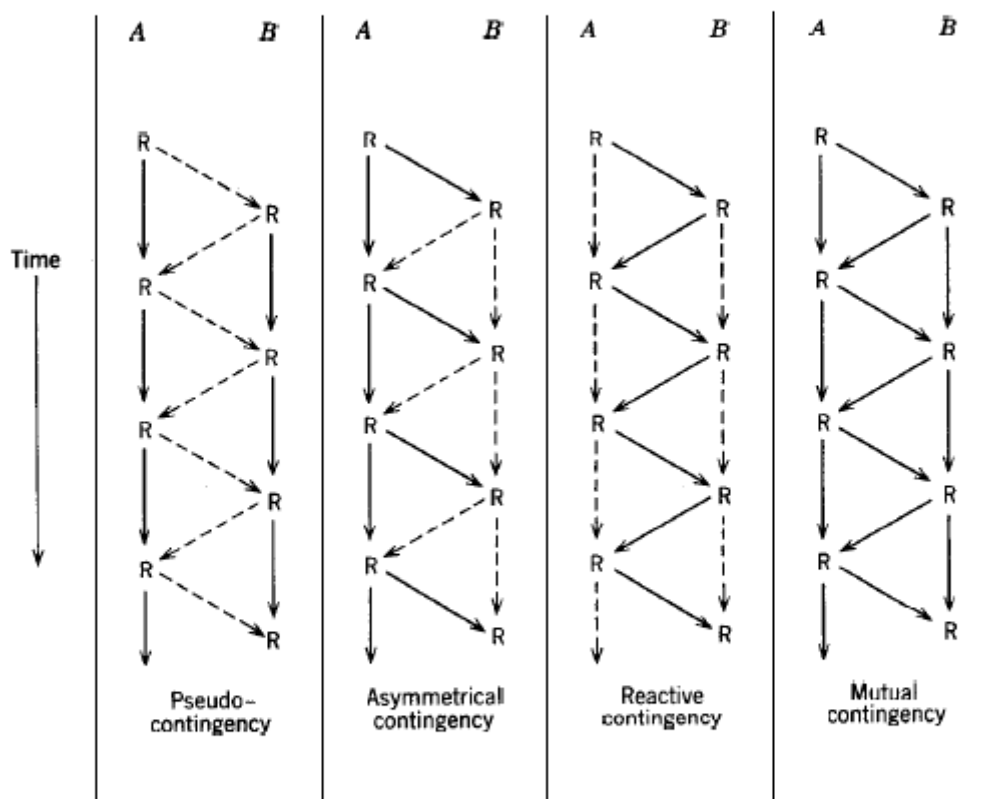


Appendix C:

FIGURE 1

Classes of Social Interaction in Terms of Contingency

From *Foundations of Social Psychology* (p. 507) by E. E. Jones and H. B. Gerard, 1967, New York: John Wiley & Sons. Copyright 1967 by John Wiley & Sons. Reprinted by permission.



Note: A and B represent interactants; R represents a reaction by one interactant to the other; \longrightarrow and \dashrightarrow represent primary and minimal (or no) contingency, respectively; *pseudocontingency* describes speech events such as acting in a play or rituals (e.g., greetings); *asymmetrical contingency* describes the interaction of (traditional) teaching and interviewing; *reactive contingency* describes events such as a rambling conversation; and *mutual contingency* describes negotiations, serious discussions, and the like.

Appendix D:

+ = **good to excellent work; no questions or doubts in these areas**
 √ = **fair to good work; some doubts and/or some confusion here**
 - = **poor to fair work; many doubts and/or much confusion**
 NA = **not applicable in this case**

1. I have chosen a commercially developed or widely used language test.

- + It is a test of the language that I teach (or will teach).
- + It is a test that will influence me or my students, whether now or in the future.

I have gathered information about the test (from the publisher, from published research articles, from teachers who have used it, from websites, from scholarly journals, etc.).

- + This information includes a summary of the history of the test and a sample test (if available) or at least sample items;
- + I have provided test development information on its reliability, validity, and item analyses;
- + I have located pricing information and order forms, administration manuals and/or manuals for test score users; and
- + I have read published articles about and/or reviews of the test.

3. I have gathered first-hand information about this test.

- NA I have administered the test (if possible).
- NA I have talked to teachers who have used this test.
- NA I have talked to students who have taken it.

- + I have tried to get (and report on) some ideas about the test's washback.
- + I have tried to get (and report on) some ideas about the test's practicality.

4. I have examined the scoring system.

- + I have determined how scores are derived and reported, what guidance for interpreting scores is provided to user agencies, what test security measures are used, etc.
- + I have decided whether this is a norm-referenced or a criterion-referenced test, and whether it provides usable diagnostic information.
- + I have explained the scoring system and any problems that it may present.
- + I have sought out and provided rater-reliability data (if it was available to me).

5. I have analyzed the test.

- + I have incorporated Wesche's four components to analyze my test.
- NA I have incorporated concepts from the Brown and Bailey textbooks as I analyzed my test.
- + I have used the four principles of communicative language testing (Swain, 1984) in my review.
- + I have also incorporated any other relevant constructs from our seminar in my analysis of this test.

6. I have prepared a professional-calibre written review.

- + I have evaluated the test based on all these sources of information.

- + I have cited appropriate literature in writing my test review.
- + The review is well-documented, well organized, and thorough.
- + The body of the review is a maximum of eight to twelve pages per person, typed and double-spaced. (Appendices and reference lists don't count against the page limit.)
- + My report includes accurate citations of appropriate references in APA format.
- + I realize this assignment is worth 20% of my final course grade.
- + I understand that the grade is final and that I may not resubmit this paper to improve the grade.
- + I have learned something in completing this project and I'm proud of the work I've done.

7. I have developed an informative summary to post on our class conference.

- NA The summary follows the prescribed format (see assignment sheet).
- NA I have carefully proofread AND spell-checked the summary.
- NA The summary is of professional quality for distribution to my peers via the course Moodle.

References

ACTFL site (2009:1) Oral Proficiency Interview (OPI) Testing for Proficiency. Retrieved on August 26, 2009 from <http://www.actfl.org/i4a/pages/index.cfm?pageid=3348>

ACTFL site (2009:2) Oral Proficiency Interview (OPI) ACTFL Oral Proficiency Interview by Computer (OPIc). Retrieved on August 26, 2009 from

http://www.languagetesting.com/speaking_proficiency_academic.cfm

ACTFL site (2009:3) Language Assessment Graphic (Appendix B) Retrieved on August 26, 2009 from http://www.languagetesting.com/assessments_academic.cfm.

Breiner-Sanders, K. E., Lowe, Jr., P., Miles, J., & Swender, E. (2000). ACTFL proficiency guidelines – Speaking revised 1999. *Foreign Language Annals*, 33(1), 13-18.

Jones, E. E., & Gerard, H. B. (1967). *Foundations of social psychology*. New York: John Wiley & Sons.

Lantolf, J. P. & Frawley, W. (1988). Proficiency: Understanding the construct. *Studies in Second Language Acquisition*, 10(2), 181-195.

O'Connell, M.E. and Norwood J.L., *editors* (2007) National Research Council International Education and Foreign Languages: Keys to Securing America's Future, pp.360, 361.

Retrieved on August 28, 2009 from

http://books.nap.edu/openbook.php?record_id=11841&page=360

Omar, Alwiya S. (2009) STARTALK Professional Development Course for Swahili

Instructors: Theory and Practice. PowerPoint presentation. Retrieved on August 26,

2009 from

<http://councilnet.org/startalk/Downloads/STRTALK+2009+proficiency+and+OPI+implications.ppt>.

Swain, M. (1984). Large-scale communicative language testing. In S. J. Savignon, & M. S. Berns (Eds.), *Initiatives in communicative language teaching*. Reading, MA: Addison-Wesley.

van Lier, L. (1989). Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly*, 23(3), 489-508.

Volmer (1981) Not Available

Wesche, M. B. (1983). Communicative testing in a second language. *The Modern Language Journal*, 67, 41-55.

Yoffee, L. (1997). An overview of ACTFL proficiency interviews: A test of speaking ability. *JALT Testing & Evaluation SIG September Newsletter*. Retrieved on August 25, 2009 from http://www.jalt.org/test/yof_1.htm.