

## **An overview of the ACTFL proficiency interview: A test of speaking ability**

Leo Yoffe

The author chose to review the ACTFL OPI because, as a measure of speaking proficiency in a foreign language, this procedure has become extremely influential. Drafters of the ACTFL Proficiency Guidelines as well as the proponents of the OPI are well aware of its washback effect on the curriculum and syllabus design. Liskin-Gasparro (1984) writes that the guidelines took "proficiency-based approach to the functional-notional syllabus" as the point of departure. The proficiency movement spearheaded by ACTFL has generated a great deal of controversy among both practitioners and scholars (for further discussion see Lantolf and Frawley, 1988; Savignon, 1985; Bachman and Savignon, 1986; Byrnes, 1989; Omaggio, 1986). Today the ACTFL Proficiency Guidelines have a strong effect on the content and the teaching methodology of many foreign language courses. The guidelines are used to evaluate foreign language proficiency of secondary teachers in a number of states, and have been accepted as a standard measure to evaluate candidates' suitability for various government and administrative posts requiring FL speaking proficiency.

*"the ACTFL Proficiency Guidelines have a strong effect on the content and the teaching methodology of many foreign language courses."*

This paper will provide a brief outline of the history of the ACTFL OPI, describe the interview process drawing on personal experience and accounts of other test-takers, and discuss the controversy surrounding the use of the OPI in an academic setting. The discussion will incorporate a critical review of the relevant literature with particular attention on reliability, validity and instructional washback of the oral interview. One section will be devoted to the certification process for the benefit of those interested in becoming OPI raters.

### **OPI: Claims and Procedures**

The ACTFL Oral Proficiency Interview was developed to evaluate speaking proficiency in a foreign language. It is a criterion-referenced, direct, face-to-face interview with only one interviewer present. The interview consists of five stages: the warm-up, level checks, probes, role-play, and wind-down. 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. It costs \$120 to take the ACTFL OPI. If the testee resides in an area where no certified OPI rater is available, the interview will be conducted by telephone. The same procedure applies.

### **The ACTFL OPI's History**

ACTFL OPI proficiency scales developed out of the FSI (Foreign Service Institute) levels of oral proficiency. These levels ranged from 1 (very basic proficiency) to 5 (native-like proficiency). In the early 1980's, the Common Yardstick project aimed to make the FSI testing procedure more accessible in academia and other foreign-language related fields. To this end, it was considered necessary to establish finer gradations at the lower end of the FSI scale in order to provide test takers with realistically reachable goals. Thus the 0 - 1 range on the FSI scale was subdivided into four levels in the ACTFL OPI Guidelines: 0 - Novice Low; Novice Mid; Novice High. Conversely, the upper end of the FSI scale (levels 3 - 5) were subsumed under the 'superior' level in the ACTFL Guidelines. Labels replaced numerical values to make the test more appealing to the uninitiated user. The elicitation and scoring procedures were further elaborated and defined by Higgs and Clifford (1982).

Each proficiency level consists of five components: function, content, context, accuracy and text type. In the ACTFL OPI Guidelines, '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.

### **Certification Process**

To become a certified OPI rater a candidate must already be rated as 'superior' or 'native' in the language. The candidate must first attend a four-day workshop organized by ACTFL. The

first day is devoted to the general explanation of the Oral Proficiency Interview, its intended purpose and generic assessment criteria. All subsequent sessions are language specific, i.e. candidates are grouped according to the language of their expertise. During these training sessions mock interviews are conducted with volunteer testees. Each interview is approximately 15 minutes long, and followed by a discussion of the strong and weak points of the tester's behavior. The workshop participants will be guided by the group leaders (typically experienced raters from ACTFL) in rating each sample according to the guidelines.

The second phase of the certification process has to be completed within six months. During this time prospective raters are required to conduct 25 interviews in two cycles, and send the recordings to the ACTFL office. Of these, 5+3 randomly chosen samples will be evaluated. On two of the first five samples the ratings (of the interviewer and the trainer) must agree precisely; in the other three sub-level disagreement can be acceptable, e.g. mid-intermediate vs. high-intermediate but not low-intermediate vs. high-intermediate. Tapes are returned with detailed comments.

Certification is valid for two years, ostensibly to ensure continued proficiency of the rater. After that, a rater needs to re-qualify. Certification process including tester training manual, cue cards and role-play cards costs between US \$575 and US \$1,500 if taken in the United States, but quite a bit higher when taken in Japan.

### **OPI: Critique and Literature Review**

Despite the wide use of the OPI, questions have been raised about the construct validity of this procedure. Does the interview actually measure what it is supposed to? Tester training manual defines the OPI as:

. . . a standardized procedure for the global assessment of functional speaking ability or oral proficiency.  
(Ch. 1-1)

Nowhere in the text, however, is the definition of the '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. Van Lier (1989) goes as far as to suggest the following, somewhat facetious, yet defensible definition:

oral proficiency consists of those aspects of communicative competence that are displayed and rated in oral proficiency interviews. (p. 493)

While an exaggeration, this definition points to the problem of identifying the abilities which are subsumed under the nebulous heading of 'oral proficiency'.

As most of the talk occurs in the interactive setting, it is reasonable to assume that a conversation is the most representative demonstration of one's

*"the OPI can hardly claim to be a 'natural conversation in the target language'".*

functional speaking ability. A conversation is characterized by a variety of factors such as mutual topic nomination and equal duties in negotiation of meaning (van Lier, 1988). Seen from this perspective the OPI can hardly claim to be a "natural conversation in the target language."

### **Procedure of an ACTFL OPI**

First of all, in the course of an interview, it is an interviewer who brings in, then chooses to pursue or abandon a topic. The interviewee's input is limited to answering questions. Thus, topic nomination rests solely within the power of the interviewer. I can personally attest to the asymmetric roles of the interviewer and the interviewee in the OPI format. During an oral proficiency interview in Russian which lasted over 35 minutes the tester nominated all topics and carefully monitored any shift in the content of the conversation, I was given sufficient time to respond to a question but no opportunity to elaborate or introduce a new topic.

During the interview the tester and the testee are clearly not in equal position. The asymmetry is not specific to the OPI but is inherent in the notion of an 'interview' as an exchange wherein one person solicits information in order to arrive at a decision whereas the interlocutor produces what he/she perceives as most valued. The interviewee is, in most cases, acutely aware of the ramifications of the OPI rating and is, consequently, under a great deal of stress. On those occasions when an interview is taken for self-diagnostic purposes, the stress factor is considerably lessened.

Another problem lies in the form-function dichotomy. The interview purports to assess functional speaking ability yet the tester-training manual strongly encourages the raters to pay careful attention to the form of the language produced rather than to the message conveyed. A Russian OPI rater reported that she tries to elicit samples of 'past tense' 'subjunctive' and progressively more complex structures to ascertain the appropriate level of proficiency (Lovick, personal communication, 1994).

Role-play situations also present a problem. While they are considered an additional level check, the very nature of this activity demands some acting ability. Cultural as well as personal resistance to assuming a role may interfere with the intended purpose of this speech elicitation device. It is, therefore, important to ensure that the type of behavior called for in a role-play is not incompatible with the sociocultural and personal parameters of the interviewee.

The government (FSI) scale and the ACTFL guidelines presume that a 'zero' level and a 'perfect' level of proficiency exist. As the OPI is a criterion-referenced measure, what norm is the point of reference? The lowest point on the ILR scale, an immediate predecessor to the ACTFL OPI is 0 characterized by the absence of speaking proficiency. It is, however, very hard to determine if the testee's level corresponds to the 'absolute zero'. Even if no oral production is observable, the testee may be able to function using the sound system of the language, or by relying on cognates. Conversely, language is a constantly evolving system, so it is impossible to achieve total proficiency in it.

The norm against which the testee's performance is rated is the functional contextual and accuracy level of proficiency of an educated native speaker (ENS). ACTFL guidelines do not provide an operationalized definition of an ENS. Using ENS as a reference point opens doors to a host of related questions: Whom do we consider to be an educated native speaker? Educated native speakers of a language do not form a homogeneous group. It is possible that a lawyer speaks differently from an educated housewife. Ingram and Wylie note that ". . . it is perfectly possible to have 'native housewives who are '3s', semi-educated 'natives' who are '3+s'". How do we interpret colloquial speech among native speakers? It may deviate from the standards prescribed by the institutional grammarians. Not unimportant is also the issue of pronunciation (accuracy). The plethora of language varieties makes the evaluation of language performance a very sensitive and problematic issue.

In the light of these considerations, Bachman and Savignon (1986) make the following recommendations for revising the guidelines.

- \* the guidelines should include a working definition of 'communicative language proficiency'.
- \* the rating scale should be based on this definition with a separate score for each component, i.e. discourse, sociolinguistic competence, etc.
- \* content and context areas should be omitted in the criteria. Scales should be defined in terms of components of communicative language proficiency, e.g. cohesion, grammar, sensitivity to register.
- \* OPI scales should be defined in terms of control and range by the testee.
- \* tester-training manual should include a section on the content and context areas which could be exploited for soliciting a speech sample.

### **OPI: Definitional Approach vs. Principled Approach**

In 1988 James Lantolf and William Frawley argued against a definitional approach to oral proficiency and in favor of a principled approach, based on sound theoretical considerations. 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.' In the present article they labeled the OPI inauthentic in that experimenters cannot assume that subjects have defined a situation in the way the experimenters intended (Wertsch, Minick, and Arns, 1984, p. 160).

They noted that several states are moving toward the establishment of oral proficiency standards for teacher certification and for bilingual certification. The Universities of Minnesota and Pennsylvania are moving toward the establishment of 'intermediate-mid' as a minimum standard for students exiting its two-year language programs.

The authors call the OPI theoretically and empirically unsound and believe its institutionalization may be harmful. In the future, people who cannot meet specified requirements may be prevented from the study of languages, even if such study is only for personal enjoyment. Other people who fail to measure up to requirements may interpret such results to mean that they have no aptitude for languages.

They quote Schulz (1986) as saying that the OPI Guidelines should be prevented from penetrating any further into the foreign language curriculum until a sound theory of proficiency has been established and adequately evaluated. The fact that it is possible to devise tests on which individuals score arbitrary points does not mean that the quality being measured by the test is really metric. The illusion is provided by the scale. (Lewontin, Rose, and Kamin, 1984, p. 91) ACTFL provides no justification that oral proficiency can be so scaled. They must justify the number of levels in the scale. Why 5 or 12 or 3? Why not 6 or 95?

Volmer (1981, p. 152) defines proficiency quite nicely: 'Proficiency is what proficiency tests measure.' A theory of proficiency must be developed, independent of psychometrics. Once a theory of proficiency is proven by empirical research, then psychometrics can be re-introduced. But even this will be difficult because proficiency is not a construct that can be formalized in terms of a taxonomy of items, no matter how long or detailed.

### **The OPI does not measure communicative ability**

In 1988 Raffaldini wrote an article discussing the ACTFL OPI in relation to current models of communicative skills, arguing that the OPI fails to evaluate important aspects of the communicative abilities of learner. To its credit, the OPI does allow a thorough assessment of grammatical competence, but it provides only a partial assessment of discourse competence. The learner in an OPI is dependent upon the tester. Exchanges are initiated by the tester, and the learner never gets a chance to initiate and react to the tester's responses. Thus the OPI only assesses how the learner answers tester-initiated questions.

The testing situation created by the OPI is extremely artificial. Information and opinion exchanges are purposeless, inauthentic, and not directed toward a particular outcome. The OPI range of discourse and socio-cultural contexts is extremely limited, so it does not adequately evaluate communicative ability. On the contrary, the OPI only assesses the subject's ability to exchange factual information and opinions in formal situations using polite language with a stranger.

Raffaldini goes on to report on two situational tests which were developed and administered along with the OPI to American learners who had returned to the U.S. after a year of living in France. The first one was a written multiple-choice test, administered to both native and non-native speakers of French. The second was an oral test. These situational tests assessed more areas of language proficiency in a wider range of language-use situations than did the OPI. They assessed socio-linguistic as well as grammatical competence, and most aspects of discourse competence.

## **Validity**

In 1988 Bachman examined how the ACTFL OPI (as embodied in the ACTFL Proficiency Guidelines) provided valid indicators of the ability to communicate through speaking. He concluded that validation was impossible because traits and test methods are confounded in the design of the interview and in the interpretation of the ratings. The confounding of function, content, and structure is a fundamental flaw which makes validation impossible.

*"The confounding of function, content, and structure is a fundamental flaw which makes validation [of the ACTFL OPI] impossible".*

The ACTFL uses a single global rating of oral proficiency with no basis in current linguistic theory or empirical research in language testing. In order to validate a test like the OPI, we must begin with definitions of language ability and testing procedures that give direction for

test design and hypotheses about test performance. A high rating of grammatical competence based on an OPI is a highly reliable indicator of a very narrow ability: a subject's skill in using grammatical structures accurately in contexts and under the conditions that are included in the testing procedure. Thus the ACTFL OPI, as currently designed, is not a valid measure of communicative language ability.

The OPI claims to be based on procedures developed over years of practice and governmental experience, but is this experience evidence of validity? Governmental experience with the OPI only provides (1) evidence that OPI findings can predict something about an individual in a government position, and (2) circumstantial evidence that ratings measure some ability of interest to the government's language schools.

The question is, is this type of predictability relevant to the academic language teaching community and second language acquisition researchers?

To its credit, the proficiency movement has infused new interest into the field of foreign language teaching. However, because of the litigious society in which we live, ACTFL must be prepared to show that a given rating is a valid indicator of proficiency in speaking and valid for the purpose of making decisions about minimum competency or employability.

### **Construct Validity**

In 1990 Dandonoli (ACTFL) and Henning (Educational Testing Service) reported on the results of research conducted by ACTFL on the construct validity of the ACTFL Proficiency Guidelines and OPI. This was one of the few studies which examined the validity of the ACTFL OPI using statistical procedures. A multitrait-multimethod validation study was the basis of the research design and analysis, including speaking, writing, listening, and reading tests in English as a Second Language and French. She acknowledged that this validation study was done in answer to criticisms by several authors, including Bachman and Savignon, and Lantolf and Frawley. For purposes of the data analyses, proficiency ratings (ordinal scale) were converted to numerical scores (interval scale) as follows:

<b>Proficiency Rating</b>	<b>Numerical Equivalent</b>
Novice Low	0.1
Novice Mid	0.3
Novice High	0.8
Intermediate Low	1.1
Intermediate Mid	1.3
Intermediate High	1.8
Advanced	2.3
Advanced High	2.8
Superior	3.3



No other rationale for converting the ordinal scale to an interval scale, nor the fractional distances between the intervals, was given. There were fewer proficiency levels in the table for listening and reading than for speaking and writing because passages were purposely selected at only five levels, the 'major borders' in listening and reading. In French, the correlation matrix revealed that speaking, writing, and reading showed discriminant validity in all required comparisons, but in listening, only three of the required 12 comparisons were met. The ESL skills of speaking and reading showed convergent/discriminant construct validity in all thirteen required comparisons, but writing only exhibited such validity in eleven of thirteen comparisons and listening in ten of thirteen comparisons. In French listening, only four of the required thirteen comparisons were met. The calibration figures indicated that there was generally adequate progression in the right direction on the latent ability and difficulty continua associated with the descriptors provided in the Guidelines, but in English there were three exceptions: (1) ratings of Novice High speaking tended to be higher on the scale than ratings of Intermediate Low speaking; (2) Novice High writing higher than Intermediate Low, and (3) items related to Intermediate Listening passages tended to have lower mean difficulty than those related to Novice listening. In French there were also two exceptions: (1) Novice High writing tended to be higher than Intermediate Low writing and (2) Intermediate reading passages tended to have higher mean difficulty than those related to Advanced reading.

## **Conclusion**

Very few statistically based research studies are available about the validity, reliability, and test-method effect of the ACTFL Oral Proficiency Interview. Many experts in the field urge a revision of the guidelines, a redefinition of the proficiency scales (Bachman & Savignon). Of course, until more empirical data is collected and a sound theoretical framework of what constitutes oral proficiency is constructed, it is premature to make sweeping generalizations about the effectiveness of this testing procedure. We must be particularly cautious about basing our decisions, either administrative or academic, solely on the OPI results.

This said, the author would like to look at some positive aspects of the ACTFL OPI. It is a direct, integrative test of speaking. In spite of scholarly criticism, no suggestion has been made to do away with the OPI, possibly because no better alternative presently exists. Most authorities in the field of language testing recognize the limitations of the test format, and advocate various changes, yet the feasibility of developing a testing procedure which would respond to all their specifications has not yet been realized. Raffaldini (1988) for example, is

critical of the OPI because it does not adequately assess all components of communicative competence. Her point is well taken, however, is it possible to construct a test which would? 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. 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. It is also instructive to examine the correlation between the OPI results and the score on more traditional tests such as College Board Achievement Test (Huebner and Jensen, 1982) Lastly, the washback effect on classroom teaching has been positive as the practitioners place more emphasis on speaking, encouraging student oral production in class.

The ACTFL OPI is a test of speaking ability within severely restricted parameters. It is, therefore, necessary to treat it as such rather than as an ultimate test of oral proficiency. Until more research into its validity, reliability, and theoretical underpinnings can warrant a wider interpretation its scores should be judged very judiciously.

## References

- American Council on the Teaching of Foreign Languages. (1986). *ACTFL Proficiency Guidelines*. Hastings-on-Hudson, NY: Author.
- American Council on the Teaching of Foreign Languages. (1988). *Oral Proficiency Interview Tester Training Manual*. Hastings-on-Hudson, NY: Author.
- Bachman, L.F.(1988). Problems in examining the validity of the ACTFL Oral Proficiency Interview. *Studies in Second Language Acquisition, 10* (2), 149-161.
- Bachman, L.F. & Savignon, S.J. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL Oral Interview. *Modern Language Journal, 70* (4), 380-389.
- Byrnes, H. (1989). Evidence for discourse competence in the Oral Proficiency Interview. *Applied Language Learning, 1* (1), 1-13.
- Clark, J. & Clifford, R. (1988). The FSI/ILR/ACTFL proficiency scales and testing techniques. *Studies in Second Language Acquisition, 10*, (2) 129-147.
- Dandonoli, P. & Henning, G. (1990). An investigation of the construct validity of the ACTFL Proficiency Guidelines and oral interview procedure. *Foreign Language Annals, 23* (1), 11-21.
- Lantolf, J. P. & Frawley, W. (1988). Proficiency: Understanding the construct. *Studies in Second Language Acquisition, 10* (2), 181-195.
- Lowe, P., Jr. (1986). Proficiency: Panacea, Framework, Process? A reply to Kramersch, Schulz, and particularly to Bachman and Savignon. *Modern Language Journal, 70* (4), 391-397.

Lowe, P., Jr. (1987). The unassimilated history. In P. Lowe & C. Stansfield (Eds.) *Second Language Proficiency Assessment*. Old Tappan, New Jersey: Prentice Hall Regents.

Meredith, R.A. (1990). The Oral Proficiency Interview in real life: Sharpening the scale. *Modern Language Journal*, 74 (3) 288-295.

Rafaldini, T. (1988). The use of situation tests as measures of communicative ability. *Studies in Second Language Acquisition*. 10 (2), 197-211.

Savignon, S.J. (1985). Evaluation of communicative competence: The ACTFL Provisional Proficiency Guidelines. *Modern Language Journal*, 69 (2), 129-134.

Sieloff Magnan, S. (1987). Rater reliability of the ACTFL Oral Proficiency Interview. *The Canadian Modern Language Review*. 3 (2), 525-537.

van Lier, L. (1989). Reeling, writhing, drawing, stretching, and fainting in coils: Oral Proficiency Interviews as conversation. *TESOL Quarterly*, 23 (3), 489-507.

## **Appendix: Oral Proficiency Interview Facets**

### **FACETS OF THE TEST RUBRIC**

Time Allocation - 15 to 35 minutes, depending on the level

### **FACETS OF THE INPUT FORMAT**

Channel of presentation:	aural
Mode of presentation:	receptive
Form of presentation Language:	aural
Vehicle of presentation:	live
Language of presentation:	L2
Identification of problem:	N/A
Degree of speededness:	unspeeded

### **NATURE OF LANGUAGE**

Length:	varied, usually relatively short.
Propositional content vocabulary:	familiar
Degree of contextualization:	context embedded
Distribution of new information:	N/A
Type of information:	concrete <-> abstract
Topic:	general, nominated by the testee (varied)
Genre:	conversation

### **ORGANIZATIONAL CHARACTERISTICS**

Grammar:	formal to semi-formal
Cohesion:	yes
Rhetorical organization:	yes
Pragmatic organization illocutionary force:	functional
Sociolinguistic characteristics:	usually standard var. of the TL, semi-formal register

**FACETS OF THE EXPECTED RESPONSE FORMAT**

Channel:	aural
Mode:	productive
Type of response:	unrestricted
Form of response:	oral
Language of response:	L2

**NATURE OF LANGUAGE**

Length:	varied, mostly relatively short sentences
Propositional content	
Vocabulary:	familiar/general
Degree of contextualization:	mostly embedded
Distribution of new information:	N/A
Type of information:	concrete & abstract, depending on interviewee
Topic:	varied
Genre:	conversation
Organizational characteristics	
Grammar:	yes (from formal to informal registers)
Cohesion:	yes
Rhetorical organization:	yes
Pragmatic characteristics	
Illocutionary force:	functional
Sociolinguistic characteristics:	standard variety of the TL

**RESTRICTIONS ON RESPONSE**

Channel:	restricted
Format:	unrestricted
Organizational characteristics:	restricted
Propositional and illocutionary:	fairly unrestricted
Time or length of response:	fairly unrestricted
Relationship between input and response:	reciprocal

**HTML:** [http://www.jalt.org/test/yof\\_1.htm](http://www.jalt.org/test/yof_1.htm) **PDF:** <http://www.jalt.org/test/PDF/Yoffe1.pdf>