

Oral Proficiency Testing in the Real World: Answers to Frequently Asked Questions

Elvira Swender

American Council on the Teaching of Foreign Languages

Abstract: *The ACTFL Oral Proficiency Interview (OPI) is used to assess the ability of individuals to use language for real-world purposes. Today, OPIs are used by academic institutions, government agencies, and private corporations for many purposes: academic placement, student assessment, program evaluation, professional certification, hiring, and promotional qualification. Through Language Testing International (LTI), the exclusive ACTFL testing office, ACTFL conducts, rates, and archives 8,000 to 10,000 oral proficiency interviews each year. This article addresses questions that are frequently asked by educators, test takers, employers, certification boards, and others who require information about an individual's level of oral proficiency.*

The frequently asked questions (FAQs) addressed in this article are (1) Does taking an OPI over the phone produce a different rating than a face-to-face interview? (2) Are there differences in testing performance from one testing occasion to another when there is no significant opportunity for learning or forgetting between the two tests? (3) How proficient are today's foreign language undergraduate majors? (4) What minimum levels of proficiency are required in the workplace?

The answers to questions 1 and 2 are based on the results of an ACTFL-sponsored testing project that compared face-to-face with telephonic interviews. The findings indicated that there is no significant difference in the ratings assigned using face-to-face versus telephone test administration. The data from the same study indicated that comparable results are obtained in test/retest situations. Questions 3 and 4 are answered using data from the ACTFL Test Archives. The majority of undergraduate language majors have achieved proficiency levels that cluster around the Intermediate-High/Advanced-Low border. Different jobs require different levels of proficiency. Charts are provided to summarize the findings.

About the ACTFL OPI

The ACTFL Oral Proficiency Interview (OPI) is a valid and reliable testing method that measures how well a person speaks a language (see Dandonoli & Henning, 1991; Thompson, 1995; Surface & Dierdorff, in this volume). It is a face-to-face or telephonic interview between a certified ACTFL tester and an examinee.

On the surface, the interview is interactive and continuously adapts to the interests, experiences, and abilities of the speaker. Below the surface, the OPI follows a specific set of procedures to measure general spoken language by determining patterns of strengths and weaknesses. It establishes a speaker's level of consistent functional ability as well as the upper limits of that ability. Although the conversational format of each ACTFL OPI follows an established protocol, there is no script or prescribed set of questions.

The ACTFL OPI is a criterion-referenced test. It compares a sample of speech with the 10 proficiency levels outlined in the ACTFL Proficiency Guidelines—Speaking, Revised (Breiner-Sanders et al., 2000). The Guidelines describe language proficiency from Novice-Low (little or no functional ability in spoken language) to Superior (ability to function in formal and informal contexts, and to state and support opinions and hypotheses in extended discourse, with a high degree of accuracy). The 10 proficiency levels of the ACTFL rating scale are Superior, Advanced-

Elvira Swender (Doctor of Arts, Syracuse University) is the Director of Professional Programs at the American Council on the Teaching of Foreign Languages, Yonkers, New York.

High, Advanced-Mid, Advanced-Low, Intermediate-High, Intermediate-Mid, Intermediate-Low, Novice-High, Novice-Mid, and Novice-Low.

The OPI assesses language proficiency in terms of the ability to respond to questions and tasks posed in a structured interview. The OPI does not compare one individual's performance with another's. Rather, it compares each individual's performance to the assessment criteria. The assessment criteria used to determine a rating include global tasks and functions, contexts and content areas, text type, and accuracy (see Swender, 1999).

Official ACTFL OPIs are conducted and rated by ACTFL certified proficiency testers. The testers are highly specialized language professionals who complete a rigorous training process and participate in ongoing quality control and tester-norming activities. Each interview is tape recorded and assigned an initial rating by the tester. Under the supervision of the ACTFL testing office, the interview is blindly rated by a second certified tester. If the ratings disagree, the interview is arbitrated by a third certified tester. When two ratings agree exactly, an official rating is assigned and an official ACTFL OPI certificate is issued.

FAQs 1 and 2

Does taking an OPI over the phone produce a different rating than a face-to-face interview?

Are there differences in testing performance from one testing occasion to another when there is no significant opportunity for learning or forgetting between the two tests?

Background

Until 1992, ACTFL OPIs were conducted primarily for academic purposes. The interviews between certified testers and individuals whose oral proficiency was being assessed took place almost entirely face-to-face. There are understandable reasons for conducting oral proficiency interviews face-to-face. The format fosters a sense that a conversation—rather than a test—is taking place, thereby creating a more relaxed performance atmosphere.

For some interviewees, the ability to give and understand nonverbal cues and gestures can ease perceived gaps in verbal communication. Within the academic context, when tester and student are located in the same school, face-to-face interviewing is logical and convenient.

In recent years however, the increased use¹ of the ACTFL OPI for academic, governmental, and private-sector purposes has made face-to-face testing less practical. Telephonic interviews allow for more flexible scheduling and more cost-effective pricing. Telephonic interviews can be arranged anywhere in the world. Many interviews can be administered within a short period of time by utilizing multiple testers, thus avoiding the tester fatigue factor that can result from day-long testing projects with individual testers. In a recent testing project (see Paulsen et al., 2002), some interviewees actually preferred telephone testing

because it eliminated the possibility of seeing any involuntary grimaces or other negative visual reactions from the interviewer.

Of course, telephone interviews may have disadvantages for the interviewee. For example, the speaker is unable to see the interviewer's gestures and facial expressions, and the physical limitations of telephone communication may restrict the full range of sounds. To partially compensate for these factors, testers extend the warm-up period of the interview and/or spend additional time establishing the level of consistent functional ability.

Telephonic interviews presently account for approximately 95% of official ACTFL OPIs administered by the ACTFL Testing Office. The shift from face-to-face to telephonic interviews raises the question of whether an examinee, tested in one modality according to the standard ACTFL interview protocol, would receive the same rating if he or she were tested via another modality. A previous study (see Jackson, 1998), found that the ratings in the two types of tests correlated positively but failed to meet the author's 70% standard of exact matches. Since such a study would require testing the same individuals twice, using two testing modalities, a second research question automatically emerges: Would examinees score higher on the second test due to practice or familiarity with the OPI process?

To address these questions, ACTFL sponsored a research study involving students in a summer language immersion program from the Spanish School of the Middlebury Language Schools. The students were in the fifth week of a 7-week summer immersion program. A total of 40 students volunteered for the project. The participants were undergraduate and graduate students, ages 19 to 35. They represented a range of language ability from Intermediate through Superior. No Novice-level speakers were tested.²

The incentive for participation in the study was the awarding of an official OPI rating and certificate. Participants were informed ahead of time that if the face-to-face and telephone interviews resulted in different final ratings, the higher rating would be reported as the official rating on the certificate. Thus, the participants were highly motivated to do their best on both performances.

Study Design

The students were randomly divided into two groups of 20 (Group A and Group B). The 40 students were scheduled for two interviews within a 48-hour period. The 48-hour period provided no significant opportunity for learning or forgetting between the two test occurrences. To control for the effect of practice, the researcher(s) scheduled Group A participants to take the face-to-face interview first and Group B to take the telephonic interview first. Despite the best intentions in study design, several students were unable to participate at the last minute; others began the process but were unable to complete it within the 48-hour

time frame. As a result, a total of 34 students (19 in Group A and 15 in Group B) completed the process.

Testing and Rating Procedures

Five experienced certified ACTFL testers conducted the interviews. These testers, active in the official ACTFL Testing Program, conduct hundreds of tests each year. To keep the number of test variables to a minimum, two adjustments were made to the standard ACTFL testing and rating protocol. First, the same tester conducted both the test and the retest interview for any given individual. Second, the testers did not assign any ratings. This decision was made to avoid the possibility that the testers might recall a previously assigned rating. A consequence of this decision was that some testers reported that it required a conscious effort on the tester's part to avoid topics already explored in the first interview.

In spite of these adjustments to the standard OPI administration protocol, all the samples were considered to be ratable; that is, all samples contained clear evidence of the speaker's highest level of sustained performance and the level at which the speaker could no longer sustain the requirements for the level.

Following testing, the researcher(s) forwarded the tapes of the interviews to six experienced raters, who listened to the interviews and assigned ratings. Each interview was independently double-rated, resulting in two ratings for each face-to-face interview and two for each telephonic. No rater listened to both samples of any one individual. The two ratings for each interview were reported to ACTFL.

Where disagreements occurred, they were within contiguous sublevels, that is, when the two raters disagreed, their ratings were not more than a sublevel apart. As is standard rating practice, cases in which there was a disagreement between the two ratings were arbitrated by a third rater. Thus, all final ratings were determined by two independent raters. These final (official) ratings are reported in Tables 1 and 2.

Findings

Tables 1 and 2 report the findings for this study. Table 1 compares ratings according to the testing modality. Table 2 reports the findings for the test/retest factor. The shaded areas in both tables show exact agreement. In both analyses, the correlation between ratings is significant beyond the .001 level.

As Table 1 shows, there was a high level of agreement between the ratings derived from the face-to-face and the telephone tests. There was no significant difference in ratings either by testing modality or test occurrence. There was no pattern of shift in one direction or the other. For the face-to-face versus telephonic testing, the final ratings agreed exactly in 32 of 34 cases. In the two cases of disagreement, the telephone rating was higher in one case and lower in the other.

As Table 2 shows, there was also a high level of agreement between the ratings assigned on the first and second interview occasions. For the test occurrence results, the ratings agreed exactly in 32 of 34 cases. In the two cases in which there were disagreements, the retest received the higher rating. While this difference might be attributable to a practice effect, a Wilcoxon Signed Ranks Tests showed that the difference is not statistically significant.

The ACTFL rating levels were coded as Intermediate-Low = 1.0, Intermediate-Mid = 1.4, Intermediate-High = 1.7, Advanced-Low = 2.0, etc; and then both Pearson Product Moment and Kendall's *tau-b* were run on the data in Tables 1 and 2.

The results comparing the ratings for the first and second test occurrences were: Pearson $r = 0.986$, significance = .000; Kendall's *tau-b* = 0.974, significance = .000.

The results comparing the face-to-face and telephonic ratings were: Pearson $r = 0.928$, significance = .000; Kendall's *tau-b* = 0.972, significance = .000.

These findings indicate that there was no significant difference in the ratings assigned whether the test modality

Table 1

SPANISH FACE-TO-FACE AND TELEPHONIC TEST RATINGS

Face-to-Face	Telephonic							Total
	Intermediate-Low	Intermediate-Mid	Intermediate-High	Advanced-Low	Advanced-Mid	Advanced-High	Superior	
Intermediate-Low	1							1
Intermediate-Mid		4						4
Intermediate-High			7					7
Advanced-Low			1	5	1			7
Advanced-Mid					8			8
Advanced-High						5		5
Superior							2	2
Total	1	4	8	5	9	5	2	34

Table 2

TEST OCCURRENCE RESULTS								
First Occurrence	Second Occurrence							Total
	Intermediate-Low	Intermediate-Mid	Intermediate-High	Advanced-Low	Advanced-Mid	Advanced-High	Superior	
Intermediate-Low	1							1
Intermediate-Mid		4						4
Intermediate-High			7	1				8
Advanced-Low				5	1			6
Advanced-Mid					8			8
Advanced-High						5		5
Superior							2	2
Total	1	4	7	6	9	5	2	34

ty was face-to-face or telephonic. In addition, the data indicate that comparable results were obtained in test–retest situations. The results suggest—at least within this language (Spanish), within this range of scores (Intermediate to Superior), and with highly trained interviewers and raters—that an individual is likely to receive the same rating whether tested face-to-face or telephonically, and is likely to receive the same rating when retested within a 48-hour period.

FAQ 3

How proficient are today's undergraduate foreign language majors?

Background

Each year a number of colleges and universities make arrangements with the ACTFL Testing Office to assess the oral proficiency of their graduating language majors. The tests are conducted either face-to-face or telephonically. Tests are double-rated according to standard ACTFL testing and rating protocol and are certified through the ACTFL Testing Office. Official OPI ratings are reported to the schools and shared with the individual students. Schools use this information for a variety of purposes: to document student outcomes, to inform program goals, to evaluate program effectiveness, and to justify program decisions (e.g., whether to invest in study abroad or immersion programs).

A number of testing initiatives involving undergraduate language majors over a period of years have produced data that help to answer the next frequently asked question. This study collected and analyzed the data from official ACTFL OPIs taken by foreign language majors at five different colleges.

Data

A total of 501 official OPI ratings in seven different languages (Chinese Mandarin, French, German, Italian, Japanese, Russian, and Spanish) were collected. The tests

were conducted face-to-face and telephonically. All interviews were double-rated and certified through the ACTFL Testing Office. The data were collected over a period of 5 years between 1998 and 2002. The students were foreign language majors in their junior or senior years when tested.

The students represented five liberal arts colleges. All of the schools offered study-abroad programs, and many of the students had participated in these programs. The schools reported that their student populations consisted primarily of individuals who had learned language in a traditional academic setting; few heritage learners were included in the sample.

Findings

Table 3 lists the 501 official OPI ratings, the percentage of students rated at each level of proficiency, and the cumulative percentages. Table 4 breaks these numbers down by language, indicating the number of students rated at each level of proficiency and the percentage of the total number in that language rated at that level.

What we observe from these ratings is that slightly less than half (47%) of the foreign language majors tested were rated above the Advanced level border and slightly more than half (53%) received ratings below Advanced. The greatest concentration of ratings (55.8%) was in the Intermediate-High/Advanced-Low range. No students tested were rated below the Intermediate level. The numbers of majors reported for languages other than Spanish and French were so small that trends in the data could not be readily identified.

FAQ 4

What minimum levels of proficiency are required in the workplace?

Background

A note at the bottom of a resume indicating that one is “flu-

Table 3

ALL RATINGS FOR ALL LANGUAGES			
ACTFL Rating	Number of Students	% of Total ⁶	Cumulative %
Superior	12	2%	2%
Advanced-High	24	5%	7%
Advanced-Mid	95	19%	26%
Advanced-Low	105	21%	47%
Intermediate-High	175	35%	82%
Intermediate-Mid	86	17%	99%
Intermediate-Low	4	1%	100%
Novice-High	0		100%
Novice-Mid	0		100%
Novice-Low	0		100%
Total	501	100%	

ent in French” or “proficient in Russian” may mean different things to different people. It may not mean, however, that the person has the language proficiency (i.e., the ability to function in a language) to perform the communicative tasks in the target language required for the job. Proficiency testing is increasingly in demand, as more companies, agencies, corporations, and other institutions are faced with the need to hire linguistically competent employees.

One of the challenges in meeting the demand for employees qualified for multilingual positions is to identify the level of proficiency that corresponds to the responsibilities of the job. For example, if a position requires that

an employee interact with the general public, resolve problematic situations, make clear explanations, report facts and other information, an Advanced level of proficiency may be sufficient to carry out the tasks successfully. If the position requires, however, that the employee be able to discuss topics from an abstract perspective, support opinions, and hypothesize, then Advanced-level proficiency may be inadequate.

In today's workplace, many companies, agencies, corporations, and other institutions are experiencing ever-increasing demands to hire personnel with language skills. Companies must ask the question, “How proficient does one need to be for the job?” The answer is, “It depends on the job.” As a service to schools, agencies, certification boards, and corporations who require linguistically competent employees, the ACTFL Testing Office provides analyses of the linguistic tasks and responsibilities of specified positions.

Analysis

Table 5, “Proficiency Levels in the Workplace,” reports information that has been compiled from collaborations between ACTFL and academic, government, and commercial agencies. It draws relationships between levels of proficiency and positions in the workplace based on the communication tasks, contexts, and accuracy requirements of the Proficiency Guidelines. The table includes the ACTFL and ILR³ ratings by level and sublevel, the functions or global tasks associated with each level, a listing of professions and positions that correspond to these levels, and an indication of what group of language learners is likely to function at that level of proficiency and how long it typically takes to reach that level.

It should be noted that the proficiency levels that correspond to specific job descriptions in Table 5 represent

Table 4

LANGUAGE-SPECIFIC RATINGS							
	Spanish n = 220	French n = 210	German n = 32	Italian n = 12	Russian n = 7	Mandarin n = 10	Japanese n = 10
Superior	6 (3%)	2 (1%)	4 (13%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Advanced-High	6 (3%)	15 (7%)	2 (6%)	0 (0%)	0 (0%)	0 (0%)	1 (10%)
Advanced-Mid	38 (17%)	46 (22%)	4 (13%)	3 (25%)	2 (29%)	2 (20%)	0 (0%)
Advanced-Low	43 (20%)	47 (22%)	4 (13%)	5 (42%)	2 (29%)	2 (20%)	2 (20%)
Intermediate-High	86 (39%)	68 (32%)	9 (28%)	4 (33%)	2 (29%)	3 (30%)	3 (30%)
Intermediate-Mid	40 (18%)	31 (14%)	9 (28%)	0 (0%)	1 (14%)	2 (20%)	3 (30%)
Intermediate-Low	1 (.4%)	1 (.4%)	0 (0%)	0 (0%)	0 (0%)	1 (10%)	1 (10%)
Novice-High	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Novice-Mid	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Novice-Low	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total	220 (100%)	210 (100%)	32 (100%)	12 (100%)	7 (100%)	10 (100%)	10 (100%)

Table 5

PROFICIENCY LEVELS IN THE WORKPLACE

Proficiency ACTFL	ILR	Language Functions	Corresponding Professions/Positions	Who is Likely to Function at the Level ?
Superior	5 4	<i>Ability to tailor language to specific audience, persuade, negotiate. Deal with nuance and subtlety.</i>	Diplomat, Contract Negotiator, International Specialist, Translator/Interpreter, Intelligence Specialist	Highly educated and professionally specialized native speakers, L2 ⁷ learners with extended professional and/or educational experience in the target culture
Advanced-High	3	<i>Discuss topics extensively, support opinions and hypothesize. Deal with a linguistically unfamiliar situation</i>	University FL, Professor, Business Executive, Lawyer, Judge, Financial Advisor	Educated native speakers; Educated L2 learners with extended professional and/or educational experience in the target language environment
Advanced-Mid	2+	<i>Narrate and describe in past, present and future and deal effectively with an unanticipated complication</i>	Human Resources Personnel, Court Stenographer/Interpreter Customer Service Representative, Social Worker, Claims Processor K-12 Language Teacher (NCATE Standards), Physician, Police Officer, Military Linguist, Billing Clerk	L2 learners with graduate degrees in language related area and extended educational experience in target environment Heritage speakers, informal learners, non-academic learners Undergraduate language majors with year-long study abroad experience
Intermediate-High	2		Aviation Personnel, Receptionist, Missionary	Undergraduate language amjors without year-long study abroad experience
Intermediate-Mid	1+	<i>Create with language, initiate, maintain and bring to a close simple conversations by asking and responding to simple questions</i>	Tour guide, Cashier, Sales clerk (highly predictable contexts)	L2 learners who have completed 6-8 year sequences of language study. In AP courses, etc.; L2 learners upon completion of basic college language (4-6 semester) sequence
Intermediate-Low	1			L2 learners after 4 years of high school language sequence; L2 learners after 2 semester college sequence
Novice-High	0+			
Novice-Mid		<i>Communicate minimally with formulate and rate utterances, lists and phrases</i>		L2 learners after 2 years of high school language study
Novice-Low	0			

minimal levels that have been established by subject matter experts from a variety of agencies, organizations, and companies for whom the ACTFL testing office provides testing or analysis.

For example, a group of subject matter experts from a financial company set Advanced-Mid as a minimal level required for a Customer Service Representative and Superior as a minimal level for a Financial Advisor. Some state teacher certification boards have established Intermediate-High as a minimal level of oral proficiency for foreign language teacher candidates. Other educational boards, such as NCATE,⁴ have set the minimal level at Advanced-Low. It should be noted that the minimum proficiency levels are not levels prescribed or recommended by ACTFL, but rather represent the most frequently selected minimal levels that have been established by the job subject matter experts for whom ACTFL/LTI conducts testing.

At some proficiency levels, the answer to the question posed in the column labeled "Who is likely to function at the level?" is stated in terms of the number of years of study, the number of courses taken, or the type of language program. This does not mean that students who complete a specific number of courses or a particular type of program will necessarily be proficient at the given level, but rather that these are realistic levels of expectations for students who have completed such programs. The levels indicated are based on a compilation of testing results, anecdotal information from the workplace, and teacher experience.

It should also be noted that these expectations were written specifically for Spanish, a Category I language⁵. For Category II, III and IV languages, one can expect that more time on task would be required to reach similar levels of proficiency.

As the work world better understands the linguistic requirements for certain multilingual positions and the time-on-task required to reach those levels of proficiency, many companies, agencies, schools, and other institutions have chosen to modify their hiring processes to first test for language proficiency before testing for work-specific qualifications. These companies report that it is more cost efficient to hire employees with language skills and train them for specific positions than it is to hire individuals for specific positions and then train them with regard to language skills.

General Conclusions

The answer to FAQ 1 indicates that the testing modality does not significantly affect test results. These findings are based on interviews conducted with higher-level speakers. It would be instructive to ask this same question with speakers in the Novice-Mid to Intermediate-Mid range.

The answer to FAQ 2 confirms that the ACTFL OPI gives consistent results in test/retest situations when the tests are administered in close proximity. These results also indicate that students desiring to improve their ratings on

retests should devote a significant amount of time to first improving their language skills.

The answers to FAQs 3 and 4 represent an invitation for more research and for institutions of higher learning to evaluate their programs in terms of student outcomes vis-à-vis the real-world language needs of the workplace.

Notes

1. The number of OPIs conducted through Language Testing International, the ACTFL Testing Office, increased from 3,000 in 1996 to more than 8,000 in 2001. Eight thousand to 10,000 OPIs are administered and rated annually through the ACTFL Testing Office.
2. The study was conducted in the fifth week of a 7-week summer immersion program. At the time of the study, few students, even those who began with no functional ability, remained in the Novice ranges.
3. ILR refers to the Interagency Language Roundtable, a consortium of U.S. Government agencies that deal with issues of language. The proficiency rating scale used by many U.S. government agencies is based on the ILR Proficiency Guidelines.
4. NCATE is the National Council for the Accreditation of Teacher Education Programs. Standards for foreign language teachers, approved by NCATE in 2002, established Advanced-Low as the minimum level of oral proficiency for foreign language teacher candidates.
5. Other Category I languages include Afrikaans, Danish, Dutch, French, Haitian Creole, Italian, Norwegian, Portuguese, Romanian, Swahili, and Swedish.
6. Percentages are rounded up or down to a whole percentage number.
7. L2 refers to second language.

References

- Breiner-Sanders, K., Lowe Jr., P., Miles, J., & Swender, E. (2000) ACTFL proficiency guidelines—Speaking, revised. *Foreign Language Annals*, 33, 13-18.
- Dandonoli, P., & Henning, G. (1991). An investigation of the construct validity of the ACTFL proficiency guidelines and oral interview procedure. *Foreign Language Annals*, 23, 11-22.
- Jackson, G. L. (1998). A score is a score. Isn't it? *The CALLer*, 5(4). Center for the Advancement of Language Learning.
- Paulsen, C., Malone, M., Kenyon, D., & Swender, E. (2002). Assessing students' conversational proficiency using telephone-mediated interaction: A cognitive laboratory evaluation. Paper presented at the annual meeting of the American Educational Research Association; New Orleans, LA.
- Swender, E., ed. (1999) *ACTFL Oral Proficiency Interview Tester Training Manual*. Yonkers, NY: American Council on the Teaching of Foreign Languages.
- Thompson, I. (1995). Interrater reliability of the ACTFL oral proficiency interview in five European languages: Data from English, French, German, Russian, and Spanish. *Foreign Language Annals*, 28, 407-22.