Oral Proficiency Learning Outcomes and Curricular Design

Benjamin Rifkin
University of Wisconsin–Madison

Abstract: In this article, the author describes the place of the oral proficiency guidelines as the only nationally recognized criteria for the assessment of communicative competence in speaking across languages and discusses learning outcomes studies based on the proficiency guidelines. The guidelines are now 20 years old, but learning outcomes are not significantly better than they were before the guidelines were published—suggesting that two decades of guideline-influenced instruction have failed to have an impact on student learning. However, data suggest that students need more hours of language exposure and instruction than a college curriculum can provide in order to attain advanced-level oral proficiency. Nonetheless, the guidelines are a curricular framework that helps teachers prepare students to break through to advanced level performance, even if this occurs after completing the college-level curriculum.

Introduction

The ACTFL Proficiency Guidelines and the "proficiency movement" that followed are landmarks in the history of foreign language education in the United States. Although the ACTFL Proficiency Guidelines have been criticized on a variety of grounds (see Brindley, 1998; Liskin-Gasparro, 2000; Liskin-Gasparro, in this issue), they are nevertheless recognized by the United States government, many educational institutions (primary, secondary, postsecondary, and foreign language teacher certification programs), and many private sector firms as a reliable indicator of oral communication skills.

This broad acceptance is of critical importance in understanding the place of the proficiency guidelines, especially the guidelines for speaking (or oral proficiency guidelines), in American education today. For example, the government and educational and private sectors—recognizing oral proficiency ratings from an oral proficiency interview (OPI) on the ACTFL or Interagency Language Roundtable (ILR) scale as meaningful—hire, graduate, and promote individuals on the basis of these ratings. Moreover, ACTFL and Language Testing International (LTI) (ACTFL's testing agency) reported a 10% annual increase in oral proficiency testing for each of the past 14 years, with approximately one third of this growth coming from each of the three sectors: education, government, and business. This growth is evidence that those who use the results of the ACTFL OPI find these results meaningful as they hire and promote.

Thus, despite any theoretical reservations, the OPI and the proficiency guidelines for speaking on which the OPI is based have practical value in helping individuals and agencies distinguish among individuals with different language abilities. This implies that the construction of foreign language curricula in correlation with the proficiency guidelines also has practical value. Students expect to be trained in a way that will help them pass tests based on the proficiency guidelines so they can get a good job or a promotion. Instructors' success in the foreign language enterprise is based in part on students' success after they leave the foreign language classroom.

The proficiency guidelines remain at this time the only nationally recognized set of criteria for assessing oral communication skills across languages in the United States and the OPI (and tests derived from it, such as the SOPI [simulated oral proficiency interview], VOCI [video oral communication instrument], and so forth) is the only nationally recognized speaking test across

Benjamin Rifkin (PhD, University of Michigan) is Professor of Slavic Languages at the University of Wisconsin–Madison, Madison, Wisconsin.
languages. OPI results are therefore a chief indicator of possible learning outcomes for speaking (oral proficiency) in the foreign language curriculum.

The landmark study on learning outcomes in foreign languages was conducted by Carroll (1967), who reported on the language assessments of 2,784 college seniors majoring in a foreign language (French, German, Italian, Russian, Spanish). In this investigation, students took Form A of the MLA Foreign Language Proficiency Test for Teachers and Advanced Students, with subtests in listening, speaking, reading, and writing. In conjunction with this study, the Foreign Service Institute worked with a subset of the respondents in Carroll's pool to establish tentative correlations between the MLA skills tests and the speaking and reading proficiency ratings. Carroll (1967) was dissatisfied with the results. He observed:

> The most striking thing … is the generally low median level of attainment in audio-lingual skills that [the tests] reveal. The median graduate with a foreign language major can speak and comprehend the language only at about an FSI Speaking rating of “2+” (= ACTFL Advanced High), that is, somewhere between a “limited working proficiency” and a “minimum professional proficiency.” (p. 134)

The level of student achievement reported in Carroll’s study was probably not at what we would today call “advanced-high.” The study was based at least in part on a written test that was not proficiency-oriented in construct, and the oral assessment was likely not done with the precision of the well-articulated oral proficiency interview protocol (in either ACTFL or ILR versions) or with the precision of today’s well-trained corps of oral proficiency interview testers. Indeed, the study was conducted well before the development and adoption of the oral proficiency interview protocol and well before the advent of organized tester training. In the mid-1960s, North American students of Russian were starved for comprehensible input. Study abroad opportunities in the Soviet Union were extremely limited, there were few, if any, large and vibrant Russian émigré communities in the United States, and there were no Internet-based opportunities to read or listen to Russian media. In this context, it is very hard to believe that students of Russian in 1967 demonstrated advanced-level listening proficiency and nearly advanced-level oral proficiency in the Carroll study.

### Review of Research on Oral Proficiency Learning Outcomes

The publication of the ACTFL Provisional Proficiency Guidelines in 1982 was a landmark event for the establishment of a common metric for the assessment of learning outcomes in language study across languages. This event propelled the language field towards a more standardized approach to the assessment and reporting of learning outcomes in the foreign languages field. Indeed, the last 20 years have seen the publication of several studies examining learning outcomes in a variety of languages. Among these are studies by Magnan (1986) in French; Tschirner (1996) in German; and Brecht, Davidson, and Ginsberg (1993), Davidson (1998, 2002), Thompson (1996), and Rifkin (described in this article) in Russian. There are no published data in other languages such as ESL, Spanish, and Japanese. The picture from the published research is similar overall and is presented in Table 1.

Data for Russian upon entrance into the Middlebury Russian School are interesting as a random sampling of students from programs all over the United States, because these students were tested on the basis of what they had learned and acquired before beginning the program. In 2001, 24 students randomly selected from each of five levels of instruction were given OPIs (conducted by ACTFL-certified OPI testers and second-rated by ACTFL-certified testers). In 2002, 34 students were similarly randomly selected for oral proficiency testing. Both sets of students (2001, 2002) came from a broad range of colleges and universities across the United States.

Prior to the administration of the oral proficiency interviews, the students in the 2002 session were asked, in surveys, to identify how many hours of formal classroom instruction in Russian they had had up to that time (i.e., before the beginning of the summer immersion program at Middlebury). Of the 30 students in the 2002 sample with prior Russian instruction, three had entrance oral proficiency ratings at the advanced level or higher and one of these was a native speaker of another Slavic language. The 27 other students, whose oral proficiency ratings ranged from Novice-Mid to Intermediate-High, claimed from 60 to 600 hours of prior classroom instruction. Students entering at the Intermediate-Mid level alone represented a range of 180 to 600 hours of prior classroom instruction (with an average of 350 hours of classroom instruction). This sampling, a small cross-section of American college-level instruction in Russian, suggests that college-level curricula in Group 3 Languages can rarely achieve learning outcomes at the Advanced level.

The data from Middlebury entrance testing conforms to the results of advisory (unofficial testing) I have conducted with students at a large public university in the Midwest. This public university’s program offers 150 hours of instruction for each of the first-, second-, and third-year courses, and an additional 90 hours of instruction in the fourth-year program, for a total of 540 hours of instruction (without study abroad) in four years of instruction. Students in this program have typically demonstrated Novice-High (average) or Intermediate-Low (optimal) oral proficiency at the end of first-year Russian, Intermediate-Low (average) or Intermediate-Mid (optimal) oral profi-
ciency at the end of second-year Russian, and Intermediate-Mid oral proficiency in the third and fourth year of Russian, unless they study abroad for an entire academic year. Few students who go abroad to Russia for only a semester achieve Advanced-level function; only some who study abroad a year achieve this level. This pattern of data was confirmed by Brecht, Davidson, and Ginsberg (1993) in their analysis of students’ proficiency before and after study abroad in Russia.

This picture of foreign language learning outcomes is consistent with data from the US Department of State Foreign Service Institute (FSI). In its analysis of the learning patterns of its students—career diplomats and foreign service officers with varying degrees of "language aptitude" (however defined at the time) learning a variety of foreign languages of varying degrees of difficulty — the FSI identified patterns of achievement correlated with hours of instruction. From this analysis, the FSI created a table of expected learning outcomes correlated with expected hours of classroom instruction that was widely circulated in Omaggio-Hadley's (2001) methods textbook, Teaching Language in Context. It is reprinted in Table 2.

The classification of certain languages into groups based on difficulty presumes that the learners, in all cases, are native speakers of English without knowledge of another language more closely related to the language of study than English. According to the FSI data presented in Table 2, the number of hours of classroom instruction predicted for learning outcomes at the Advanced and Superior levels, for students of average language learning aptitude, far exceeds the number of classroom hours in the typical college curriculum, especially for languages in Groups 2 to 4.

### Table 1

**ORAL PROFICIENCY LEARNING OUTCOME STUDIES UPON COMPLETION OF 1 THROUGH 4 YEARS OF INSTRUCTION**

<table>
<thead>
<tr>
<th>Results for Students Having Completed 1 to 4 Years of Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Year</td>
</tr>
<tr>
<td>French (Magnan)(^a)</td>
</tr>
<tr>
<td>German (Tschirner)(^b)</td>
</tr>
<tr>
<td>Russian Study Abroad (Brecht et al.)(^c)</td>
</tr>
</tbody>
</table>

**Notes**

\(^a\)Magnan (1986) and Thompson (1996): These studies were published before the 1999 revision of the Proficiency Guidelines for Speaking with the introduction of the Advanced-Mid level.

\(^b\)Tschirner (1996) tested only at first- and second-year levels.

\(^c\)Brecht et al. (1993) tested students prior to departure on study abroad; results are not broken down by years of instruction, but most students had completed 2 to 4 years of classroom instruction.

\(^d\)Middlebury Russian School Entrance and Exit Test Results for 2001 and 2002 are published for the first time in the present study (1996). Entrance test results are for students entering into second- through fourth-year Russian who had completed first- through third-year Russian prior to testing.

\(^e\)Thompson: (1996) First- and second-year students tested at summer program at University of Iowa, third- and fourth-year students tested at Middlebury Russian School.
Even in the case of a Group 1 Language, students of less than superior language-learning aptitude who start their study of the language in college would be hard pressed to attain Advanced-level oral proficiency within the constraints of the four-year college curriculum. If a college foreign-language program requires four hours per week in first- and second-year classes and three hours per week in third- and fourth-year classes, the sum total of hours of classroom instruction would only equal approximately 420 hours, which falls far short of the 480 hours required for an individual of average language aptitude to reach advanced-low oral proficiency in a Group 1 language, let alone what would be required to take an individual of average language aptitude to that same level of proficiency in a Group 2, 3, or 4 language.

The comparison of learning outcomes of students in American colleges and universities with the learning outcomes of government employees in specialized and intensive language learning programs, however, is problematic for several reasons (see Liskin-Gasparro, 1984). Notably, the FSI focused on government employees learning language without distraction from homework.

The FSI data, confirmed by data from the learning outcomes studies cited in Table 1, clearly refute the persistent myth that somehow languages can be "picked up" without significant effort. The advertisements in the in-flight magazines are seductive: "Speak Spanish like a diplomat in only 48 hours!" American culture, broadly speaking, accepts the myth that somehow languages can be "picked up" without "content"—I believe that Robin is missing a critical point. Given all the data on learning outcomes, it is clear that to attain high levels of oral proficiency (advanced or higher), students need more hours than those available through the liberal arts college curriculum. If students begin the study of a language before entering college, they may be more likely to achieve higher levels of proficiency in college (since they will have more time “on task”). The truth is, however, that the vast majority of North American students who do study a foreign language in high school study either Group 1 or Group 2 languages (e.g., Spanish, French, and German)—languages for which fewer hours of classroom instruction are required for the attainment of high levels of oral proficiency. Group 3 and Group 4 languages are rarely taught on the precollege level, but it is precisely these languages that require the longest learning sequences for students to achieve professional competence.

In proficiency-oriented programs of classroom instruction, research (e.g., Magnan, 1986; Tschirner, 1996) suggested that students of average language learning aptitude studying Group 1 and Group 2 languages typically achieve intermediate oral proficiency by the end of the first year of college-level instruction; learners of Group 3 languages typically achieve this breakthrough some time in the second year of instruction (as suggested by the data for Russian reported in this paper). Students of Group 4 languages may achieve Intermediate oral proficiency only close to the time of graduation (upon completing 480 hours of instruction). Having achieved Intermediate oral proficiency, students of most languages, especially Group 2, 3 and 4 languages, typically "stall" unless they have an extended immersion experience in the United States or abroad. Even with study abroad, as argued by Brecht, Davidson, and Ginsberg (1993) and Davidson (1998, 2002), nearly two thirds of students fail to break through into Advanced-level oral proficiency.

Robin (2000) argued that the failure during the past 30 years to improve learning outcomes as measured by the proficiency scale represents a failure of the proficiency movement. He claimed that the proficiency “revolution”—that is, the improved understanding of foreign language education and the principles underlying proficiency-oriented instruction—should have resulted in significantly improved learning outcomes for students of foreign languages in America's colleges and universities:

In 1967, the typical Russian major had no proficiency-based instruction, no opportunities for length stays in country, no possibility of homestays, little in the way of strategy instruction for reading authentic language, and little recourse to authentic flow-of-speech Russian. Yet even with the advent of goal-oriented instruction and materials designed to help us meet those goals, an outside observer would be hard pressed to explain our apparent lack of meaningful progress. As we in pedagogy look at outcomes, we can't help but ask if the last three decades [since the publication of Carroll's study] have been a waste of time. (p. 29)
<table>
<thead>
<tr>
<th>Category 1 Languages(^a)</th>
<th>Minimum Aptitude</th>
<th>Average Aptitude</th>
<th>Superior Aptitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>480 hours</td>
<td>1+ (Intermediate-High)</td>
<td>2 (Advanced-Low to Advanced-Mid)</td>
<td>2+ (Advanced-High)</td>
</tr>
<tr>
<td>720 hours</td>
<td>2 (Advanced-Low to Advanced-High)</td>
<td>2+ (Advanced-High)</td>
<td>3 (Superior)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category 2 Languages(^b)</th>
<th>Minimum Aptitude</th>
<th>Average Aptitude</th>
<th>Superior Aptitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>480 hours</td>
<td>1 (Intermediate-Low to Intermediate-Mid)</td>
<td>1/1+ (Intermediate-Low to Intermediate-High)</td>
<td>1+/2 (Intermediate-High/Advanced-Low to Advanced-Mid)</td>
</tr>
<tr>
<td>720 hours</td>
<td>1+ (Intermediate-High)</td>
<td>2 (Advanced-Low to Advanced-High)</td>
<td>2+/3 (Advanced-High/ Superior)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category 3 Languages(^c)</th>
<th>Minimum Aptitude</th>
<th>Average Aptitude</th>
<th>Superior Aptitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>480 hours</td>
<td>0+ (Novice-High)</td>
<td>1 (Intermediate-Low to Intermediate-Mid)</td>
<td>1/1+ (Intermediate-Low/Intermediate-Mid/Intermediate-High)</td>
</tr>
<tr>
<td>720 hours</td>
<td>1+ (Intermediate-High)</td>
<td>2 (Advanced-Low to Advanced-Mid)</td>
<td>2/2+ (Advanced-Low/Advanced-Mid/Advanced-High)</td>
</tr>
<tr>
<td>1320 hours</td>
<td>2 (Advanced-Low to Advanced-Mid)</td>
<td>2+ (Advanced-High)</td>
<td>3 (Superior)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category 4 Languages(^d)</th>
<th>Minimum Aptitude</th>
<th>Average Aptitude</th>
<th>Superior Aptitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>480 hours</td>
<td>0+ (Novice-High)</td>
<td>1 (Intermediate-Low to Intermediate-Mid)</td>
<td>1 (Intermediate-Low to Intermediate-Mid)</td>
</tr>
<tr>
<td>720 hours</td>
<td>1 (Intermediate-Low to Intermediate-Mid)</td>
<td>1+ (Intermediate-High)</td>
<td>1+ (Intermediate-High)</td>
</tr>
<tr>
<td>1320 hours</td>
<td>1+ (Intermediate-High)</td>
<td>2 (Advanced-Low to Advanced-Mid)</td>
<td>2+ (Advanced-High)</td>
</tr>
</tbody>
</table>

**Key**

\(^a\)Languages such as Romance languages, Scandinavian languages (exc. Finnish), and Swahili

\(^b\)Languages such as German, Greek, and Indonesian

\(^c\)Languages such as Burmese, Finnish, Hebrew, Russian, and Turkish

\(^d\)Languages such as Arabic, Chinese, Japanese, and Korean

This table was originally compiled from Foreign Service Institute data and appeared in International Language Roundtable testing materials.
Teaching Towards Advanced-Level Oral Proficiency

It is challenging for teachers to lead college students to advanced-level function within the constraints of the four-year college curriculum, in part because classroom time is so limited. In one 50-minute class, it is possible to ask each of 16 students to answer a question with a single word, phrase, or sentence (Novice and Intermediate level discourse, respectively) numerous times. In the same 50 minutes, it is possible to ask each of 16 students to provide a paragraph-length narration or description only once or twice if these narrations or descriptions are not conducted simultaneously (in pair or group work). The time constraints compel instructors to use partner work as a means to increase the number of times students are able to speak in class, but this very strategy reduces the amount of close teacher observation of that interaction, making it harder for teachers to monitor the discourse of each student struggling to produce a paragraph or avoiding the task with a series of unconnected simple sentences.

The FSI data could give many foreign language educators grounds for despair: If we cannot enable our students to reach Advanced or Superior level oral proficiency within the time constraints of our curriculum, why, some might ask, should we bother to teach them anything at all? Some have argued that there should be no classroom-based foreign language instruction at all: Indeed, at least one institution has eliminated foreign language instruction on the assumption that it is better to send students abroad for immersion (Schneider, 2001). It can be argued that teachers should strive to teach students how to function at the Advanced level in isolated, “topic-based” islands (as will be described later in this article), because this prepares them to break through into advanced function, regardless of topic area, in the immersion context. Thus the proficiency guidelines provide the foreign language field with a framework for constructing curricula that lead students towards the attainment of higher levels of proficiency, even if students actually attain these higher levels after completing the four-year college curriculum.

Harking back to Vygotsky’s principal of the zone of proximal development, instructors can use the framework of the proficiency guidelines to establish a sense of the goals and objectives of teaching foreign languages, to bring students from one milestone to the next. As soon as instructors see that students have achieved Intermediate-level oral proficiency, instructors and students should focus their sights on Advanced-level functions of description and narration (as defined by the Proficiency Guidelines for Speaking).

While few students will attain Advanced-level oral proficiency in the classroom (or even while still enrolled in college), teachers can help improve the odds that they will, in fact, cross that threshold into Advanced-level performance by preparing them with “beachheads” of advanced function or “islands of ease,” as suggested by Shekhtman et al. (2002, p. 121–23). The more “islands” of Advanced-level performance students can build, the easier it will be for them to navigate the challenging waters at this level.

For example, topic-based lessons on food can require students to describe the preparation of a particular dish; students can prepare and rehearse a memorized performance of a paragraph-long description. This paragraph could be expanded to a story about how the student usually prepares the dish and how he or she prepared the dish on one particular day when a key ingredient was missing. Similar activities can be designed for other topic areas, each enticing the student to prepare a paragraph-long narration or narration and then expanding the paragraph with additional details and elaborations.

The memorized recitation of such texts cannot substitute for Advanced-level performance, and there is no research proving that the memorization of such monologues can lead to spontaneous production of Advanced-level discourse in the context of an oral proficiency interview. However, it is possible that the preparation of such rehearsed oral texts can prepare students to make the leap to advanced function while living in the target culture.

A key factor in teaching for advanced- and superior-level function is the nature of the discourse and the interaction patterns instructors foster in second language classrooms. At the Novice level, our students need to learn to ask and answer questions, to get into and out of predictable communicative situations. Accordingly, the discourse patterns observed in second language classrooms at this level are dominated by questions and answers, individual exchanges of one or two sentences in length, often mediated by teacher turns between every student turn.

However, after students have achieved Intermediate-level function, instructional objectives should shift to Advanced-level tasks and the discourse and interaction patterns must change. Teachers’ questions must be different at this level: As OPI trainers teach workshop participants, good questions must be open-ended. Teachers cannot merely ask “what” or “when” in the Intermediate classroom. They must ask “how” and “why” and, perhaps more importantly, they must insist on—and wait for—student responses of paragraph-length or longer. This means shifting the interaction patterns in the classroom. A teacher at this level cannot be a “sage on the stage,” but must serve as the “guide on the side,” facilitating interaction among students in which the teacher is not the center of attention. The teacher can set the tone, call on students to respond to one another, and engage the class in discussion, but the discussion must be deep as well as broad. The challenge for teachers at this level is to provide students with appropriate stimuli for serious discussion, discussion that will compel communication at a level requiring paragraph-length
discourse (I have described this process elsewhere; see Rifkin, 2000, 2002).

While the learning outcomes data show that it is extraordinarily difficult to bring students to high levels of oral proficiency while they are still in college, the proficiency guidelines are a framework for curricular planning that can help prepare students for the ultimate attainment of this goal. The guidelines have served the profession well by providing us with a common metric for studying learning outcomes in a variety of languages and learning contexts. There remains, of course, much work to be done: We need learning outcome studies in other languages (e.g., Arabic, Chinese, Japanese, and Spanish) in traditional classroom and study abroad contexts, as well as in nontraditional contexts (e.g., internships, weekend immersions). Moreover, the work that has been done thus far has focused on learning outcomes in speaking. To more completely understand foreign language learning outcomes, more research on outcomes in listening, reading, and writing is needed. With the development of Web-based testing in the areas of listening and reading and the technology to facilitate the transmission of sound files (for listening) and reading texts, it is hoped that studies in these areas are already underway and will soon appear on the pages of this and other journals.

Acknowledgments

Oral proficiency interviews conducted with students of the Middlebury Russian School in the summer of 2001 were paid for in part by a generous research grant from ACTFL. Statistical analysis of proficiency testing at Middlebury in 2001 and 2002 was supported by the Middlebury College Dean of Language Schools and Schools Abroad and the University of Wisconsin–Madison Center for Russia, Eastern Europe, and Central Asia. Statistical analysis of proficiency testing of students at the University of Wisconsin–Madison was supported by the Center for the proficiency paradox. In O. Kagan & B. Rifkin (eds.), The learning and teaching of Slavic languages and cultures (pp. 9–28). Bloomington, IN: Slavica.


Notes

1. While Proficiency Guidelines for listening, reading, and writing do exist, no test has been developed for these modalities that has been adopted across languages. While this situation is to be regretted, it foregrounds the importance of data generated from oral proficiency interview testing as one of the most important sources of information about learning outcomes across languages and institutions.

References


