To OPI or Not to OPI: Proficiency-Oriented Instruction and Assessment in U.S. University-Level Russian Programs

Oleksandra Wallo
Molly Godwin-Jones

Follow this and additional works at: https://scholarsarchive.byu.edu/rlj

Part of the Slavic Languages and Societies Commons

Recommended Citation
Wallo, Oleksandra and Godwin-Jones, Molly (2021) "To OPI or Not to OPI: Proficiency-Oriented Instruction and Assessment in U.S. University-Level Russian Programs," Russian Language Journal: Vol. 71: Iss. 1, Article 8.
DOI: https://doi.org/10.70163/0036-0252.1318
Available at: https://scholarsarchive.byu.edu/rlj/vol71/iss1/8

This Article is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in Russian Language Journal by an authorized editor of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.
To OPI or Not to OPI: Proficiency-Oriented Instruction and Assessment in U.S. University-Level Russian Programs

OLEKSANDRA WALLO, MOLLY GODWIN-JONES

1. Introduction

Back in 1991, Thompson claimed that the impact of the proficiency movement on how Russian was taught in the United States had resulted in something more akin to Soviet glasnost rather than perestroika. She meant that while the introduction of ACTFL Proficiency Guidelines and an emphasis on functional ability in a foreign language spurred much discussion in the Russian teaching profession in the 1980s, these developments did not lead to “the actual restructuring of curricula and assessment along functional lines” (375). Thompson mentioned several obstacles to the adoption of the proficiency-based approach for Russian teaching at that time, including a lack of communicative, proficiency-based Russian textbooks; the organization of U.S. language programs around “diffuse goals,” such as “developing an understanding of other countries and their cultures” rather than “functional language use” (386); and limited hours devoted to language study in the U.S. educational system.

Since then, many changes have taken place in the Russian teaching field. A new generation of elementary Russian textbooks appeared in the mid-1990s, each book declaring a proficiency orientation and an emphasis on developing communicative competence (Comer 2012). Despite some of their shortcomings, pointed out by Comer (2012), all of these textbooks included communicative activities. A proficiency-oriented web-based open-access textbook for beginning Russian, *Mezhdu nami*, became publicly available in 2015. This was the first textbook to apply the processing instruction approach to Russian language teaching. Some new Russian textbooks for levels above the elementary, such as *Russian: From Intermediate to Advanced* (2014), made students’ progress along the proficiency scale into their central focus. In the professional literature, analyses of post-secondary Russian language learners’ low scores on proficiency tests, such

---

1 Processing instruction is an approach to teaching L2 grammar that takes into account how learners tend to process language input (VanPatten 1996) and aims to train them to correctly connect form and meaning by structuring input in specific ways.
as the Oral Proficiency Interview (OPI) (Thompson 1996; Rifkin 2005), and calls to lower the expectations of proficiency outcomes attainable within traditional university-level Russian language curricula (Rifkin 2003; Isurin 2013) have been increasingly supplemented with reports about creative restructuring of existing curricula. Introducing intensive courses and incorporating hybrid instruction, for instance, has permitted some post-secondary Russian programs to maximize students’ language learning time and thereby improve their proficiency outcomes (Garza 2013).

*Perestroika*-like curricular innovation in some university-level Russian programs began in earnest in the early 2000s, with the launch of the federally funded Language Flagship Program for a number of critical languages, including Russian. Its ambitious goal of enabling students to reach the proficiency level of ACTFL Superior (ILR 3) in the target language by the time they graduate has required the reorganization of program timelines, instruction delivery methods, assessment practices, and more (Murphy et al. 2017). Impressively, the Flagship experiment is proving to be very successful: over three quarters of students in all Flagship programs and languages who graduated in 2014 attained “the goal of ILR 3, with over 95% of students (...) reaching Level 2+” (Nugent and Slater 2017, 23). Having demonstrated that professional competence in a foreign language by graduation is an attainable goal even for Category III languages like Russian, the Flagship model is bound to have an increasing impact on the Russian teaching profession in the years to come. Due to recent and ongoing budgetary crises and the resulting pressure on university-level language programs to justify their existence, the Language Flagship’s efficiency and its capacity to deliver measurable, high-proficiency results are becoming attractive to post-secondary Russian programs. For example, some elements of the Flagship model, such as intensive instruction that increases

---

2 The Language Flagship is an initiative sponsored by the Defense Language and National Security Education Office and administered by the Institute of International Education.

3 Unlike the ACTFL Proficiency Guidelines, the Interagency Language Roundtable (ILR) proficiency scale, widely used by governmental agencies, is numerical. The equivalency between the ACTFL oral proficiency levels and the ILR levels is discussed later in this article.

4 The U.S. Foreign Service Institute divides languages into four categories, based on the average length of time that it takes an L1 English speaker to achieve “professional working proficiency” (ILR 3) in the language. Russian and other Slavic languages are assigned to Category III as “languages with significant linguistic and/or cultural differences from English,” requiring about 1,100 hours of instruction before ILR Level 3 can be reached (“Foreign Language Training”).
the amount of classroom hours and the level of student engagement in language use, have already been adopted by the Russian program at the University of Texas at Austin (Garza 2017).

The rise of federally funded language programs, such as Project Global Officer (Project GO) and especially the Language Flagship, has also helped bridge the divide between what used to be two very different types of language learners. Thompson (1991) called these two groups “the academic learner” and “the government learner”:

The typical learner in the government setting is an adult in an intensive program who has a utilitarian motive for studying a second language: the need to meet job requirements in the target-language country. The academic learner, on the other hand, is an individual, usually in the late teens or early twenties, who studies a foreign language for a few years in a non-intensive program as part of a more general education in the humanities. (Thompson 1991, 377)

While some of the distinctions between these two kinds of learners remain, a sharp divide between them no longer exists. For example, some academic learners are becoming more like government learners, approaching the study of a foreign language with aspirations to utilize it in future careers in the government sector. Furthermore, in federally funded language programs, the two types of students sometimes learn side by side, using the same standards-based curricula and working towards the same proficiency benchmarks, as measured by proficiency tests, such as the OPI.

Despite the changes outlined previously, it remains unclear to what extent the post-secondary Russian language teaching field as a whole has embraced the proficiency-based approach to instruction and assessment. There is also a lack of information about the reasons that guide collegiate Russian programs to implement this approach or not—or, as we put it in the title of this article, “to OPI or not to OPI.” The exploratory two-part study reported in this article aims to begin to fill this gap (1) by examining the trends in the use of some proficiency-oriented practices in university-level Russian programs around the country and (2) by analyzing our recent experience of supervising or teaching in federally funded intensive Russian language programs with externally set proficiency benchmarks, measured by standardized proficiency tests. This experience has given

---

5 Project GO is an initiative sponsored by the Defense Language and National Security Education Office and administered by the Institute of International Education.
us first-hand knowledge of the so-called WASHBACK EFFECTS from the proficiency guidelines and tests, has required us to undertake considerable curricular restructuring, and has prompted us to answer for ourselves the questions of whether it is and why it might be worth it to OPI. Thus, in the first part of our study, we conducted a small-scale survey of U.S. university-level Russian programs to ask about their implementation of several common elements of the proficiency-oriented approach; we begin with a brief review of these elements and then discuss our survey and its results. In the second part of our study, we report on the action research (McNiff 1993) of our own teaching or supervision of teaching for speaking proficiency, as measured by a course-final OPI or OPIc, in two intensive Russian language programs. We discuss the programs’ oral proficiency outcomes, describe the pedagogical principles and practices that we have found most effective in helping students achieve these outcomes, and reflect on the OPI washback effects.

2. The ACTFL proficiency guidelines and the OPI: pro and contra

The ACTFL Proficiency Guidelines, which define language proficiency in terms of “functional language ability,” provide one increasingly popular way to establish externally recognized, proficiency-based language program standards (ACTFL 2012, 3). The Guidelines describe five major levels of proficiency (Novice, Intermediate, Advanced, Superior, and Distinguished) in each of the four language skills (speaking, listening, reading, and writing). Each proficiency level is characterized by specific language abilities, which increase in difficulty in each ascending level. For example, while speakers at the Novice level are able to use “memorized” and “isolated words and phrases” to “communicate short messages on highly predictable, everyday topics” (ACTFL 2012, 9), Intermediate-level speakers can “produce sentence-level language” and “create with the language when talking about familiar topics related to their daily life” (ACTFL 2012, 7). Novice, Intermediate, and Advanced levels are further divided into sublevels of Low, Mid, and High, which capture how well an individual controls the abilities of the given major level. Table 1 shows how the levels of the ILR scale, which is numerical, correspond to the ACTFL oral proficiency levels.
Table 1. Equivalency between the ACTFL levels and the ILR scale.*

<table>
<thead>
<tr>
<th>ACTFL</th>
<th>ILR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distinguished</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Superior</td>
<td>3</td>
</tr>
<tr>
<td>Advanced High</td>
<td>2+</td>
</tr>
<tr>
<td>Advanced Mid</td>
<td>2</td>
</tr>
<tr>
<td>Advanced Low</td>
<td></td>
</tr>
<tr>
<td>Intermediate High</td>
<td>1+</td>
</tr>
<tr>
<td>Intermediate Mid</td>
<td>1+</td>
</tr>
<tr>
<td>Intermediate Low</td>
<td>1</td>
</tr>
<tr>
<td>Novice High</td>
<td>0+</td>
</tr>
<tr>
<td>Novice Mid</td>
<td>0</td>
</tr>
<tr>
<td>Novice Low</td>
<td></td>
</tr>
</tbody>
</table>

*Adapted from ACTFL (2015)

The OPI is a standardized test of oral proficiency administered through Language Testing International (or by government-certified testers) that is tied to the ACTFL guidelines in speaking (or the ILR scale). It assesses a person’s ability to communicate in a foreign language through a 25–35-minute-long interview between the examinee and a certified tester, which is conducted either in person or by telephone. The interview may cover a wide variety of topics but has a consistent structure: it alternates between level checks and probes. Level-checking questions establish the highest level at which the examinee can sustain performance without linguistic breakdown while probes test the examinee’s abilities at the next higher level to ascertain that this level of proficiency has in fact not been reached (Malone and Montee 2010). In the Oral Proficiency Interview by Computer (OPIc), the tester is replaced with a digital avatar and the examinee takes one of five test forms of different proficiency ranges, selected on the basis of the examinee’s self-assessment. Prompt topics are tailored to the examinee’s interests, as reported in a background survey (Isbell et al. 2019). Both the OPI and the OPIc are recorded and scored independently by two raters certified for the specific format of the test. When the scores of the two raters do not match, a third rater is enlisted.

The ACTFL Proficiency Guidelines and the OPI procedure have been repeatedly revised in response to early criticism and are quite reliable in their current form (Malone 2003; Malone and Montee 2010). The more philosophical critiques that have questioned their validity, however, have
been harder to refute. Malone and Montee (2010) provide a thorough overview of these critiques, ranging from the charge that the guidelines are not grounded in empirical research on language acquisition to the argument that the format of the OPI—a formal interview—does not capture many real-life speaking situations and thus is hardly a global test of speaking proficiency (Johnson 2001; Liskin-Gasparro 2003).

Of special relevance to our study is the ongoing debate about the washback effects from the OPI. On the one hand, “the ACTFL OPI may have a potential positive impact by encouraging students to speak and classroom instructors to emphasize speaking” (Malone and Montee 2010, 979). Scholars who have found college learners’ OPI scores falling short of the required benchmarks, such as in teacher certification, advocate for allowing the OPI to transform post-secondary language departments into “clearly articulated, proficiency-based programs” (Kissau 2014, 540). On the other hand, some scholars caution against using the guidelines and the OPI to establish the language program’s proficiency standards, suggesting that doing so may devalue the wide scope of the program and its curricular foci beyond proficiency (Norris and Pfeiffer 2003). At the same time, these critics acknowledge the value of the OPI as a form of external feedback on the quality of the curriculum, instruction, and learning in a program, and as a way to provide students with a widely recognized standard rating of their speaking skills (Norris and Pfeiffer 2003).

Finally, there are considerable practical constraints to implementing the ACTFL OPI. Official testing and tester training require an investment of time and finances (Malone and Montee 2010), and regularly administering even unofficial OPIs to large classes may be too time-consuming for instructors.

3. Intensive instruction

In addition to her commentary on the lack of proficiency-based textbooks and of a program focus on functional language ability, Thompson (1991) discussed the issue of limited time devoted to language study in the U.S. educational system, which “mitigate[s] against the acquisition of language skills that could be professionally applied upon graduation” (386). For post-secondary U.S. language programs, this problem is perhaps even more acute now than it was in 1991, with fewer undergraduate students choosing to major in a foreign language and a rising trend of double or triple majors, which places severe constraints on how many years students can devote to studying a language. One solution is intensive instruction, which has been used for decades with government learners and is becoming increasingly
common in academic language programs.

In addition to the fact that intensive courses permit learners to cover more material in a shorter time period, there is growing empirical evidence about the advantage of intensive instruction over traditional “drip-feed” instruction (Stern 1985) for proficiency development (Serrano 2011). Such advantages have been reported both in studies that compared five-week intensive courses with about 25 hours of instruction per week to seven-month courses with 4 hours of instruction per week (Serrano and Muñoz 2007) and in studies that described less intensive year-long courses with 6 hours of instruction per week (Norris and Pfeiffer 2003; Garza 2013), which may be more typical of post-secondary language programs. Garza (2013) found that the OPI ratings of students who completed two semesters of intensive Russian for beginners, with six contact hours per week, averaged higher than the OPI results of students after the traditional two-year sequence of Russian at the University of Texas at Austin, although the innovative curriculum of this course likely accounted for some of this difference. Norris and Pfeiffer (2003) reported a similar advantage for oral proficiency in year-long German intensive courses at Georgetown University: considerably more “intensive-track” learners (two semesters, six hours per week) than “nonintensive-track” students (four semesters, three hours per week) received Intermediate High ratings or above after the same total amount of class hours, even though the curriculum in both tracks was the same (577). This evidence seems to suggest that even a relatively modest concentration of instruction (i.e., six hours per week) may be more effective in fostering students’ oral proficiency than the more common three- or four-hour courses taught over a longer period.

Although there is still little empirical research on the connection between intensive instruction and proficiency, teaching intensively for language proficiency has been standard practice in the government sector as well as in some academic summer programs, both domestic and abroad. Consequently, we decided to add intensive instruction to our investigation of proficiency-oriented practices at U.S. university-level Russian programs. We explored the trends in the use of these practices in the Russian post-secondary teaching field by conducting a survey. This part of our study sought an answer to the following research question:

To what extent have post-secondary Russian language programs implemented the proficiency-based approach to instruction and assessment, as evidenced by the teaching of intensive courses, the use of the ACTFL Proficiency Guidelines as benchmarks for student progress, and the incorporation of the OPI into assessment practices?
4. Survey results
The survey was conducted in spring and summer of 2019 using Google Forms. The results do not represent a random sample of post-secondary Russian language programs, but rather a convenience sample. The survey was advertised via the SEELANGs listserv and a Facebook group for teachers of Russian. In addition, some individual colleagues were contacted by email. This technique of collecting data may have resulted in a higher response rate from institutions that already use the proficiency-oriented practices under study or are in favor of them. Some limitations were inherent in the format of the survey, such as close-ended questions, which limited the potential for feedback and did not allow us to ask specific follow-up questions. Nonetheless, the survey does reveal some trends regarding the use of the proficiency-oriented practices under examination.

The survey was completed by 26 participants, as illustrated in Table 2. The majority of respondents represented public research universities (n = 15), with eight participants from private research universities and three from small liberal arts colleges. The departments represented vary from Slavic and East European (n = 8) to Modern and World Languages (n = 14) to a combination of Slavic and other languages (n = 4), such as Slavic and German or Romance languages.

Table 2. Participant representation.

<table>
<thead>
<tr>
<th>University Profile</th>
<th>Department Profile</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Small liberal arts</td>
<td>3  Slavic/East European</td>
<td>8</td>
</tr>
<tr>
<td>Public research</td>
<td>15* Slavic and one/two other languages</td>
<td>4</td>
</tr>
<tr>
<td>Private research</td>
<td>8 All languages (i.e., “Modern Languages Dept”)</td>
<td>14</td>
</tr>
<tr>
<td>TOTAL</td>
<td>26 TOTAL</td>
<td>26</td>
</tr>
</tbody>
</table>

*Including 3 Russian Flagships

Of the 26 participating institutions, almost half (n = 12) teach Russian intensively, while 14 do not (see Table 3). Of our participants, only one small liberal arts college teaches Russian intensively, but four private research universities and seven public universities teach intensively (including all three Russian Flagship Programs that participated in this survey).
Table 3. Number of institutions that teach Russian intensively (not including study abroad).

<table>
<thead>
<tr>
<th>University Profile</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Small liberal arts</td>
<td>1</td>
</tr>
<tr>
<td>Public research</td>
<td>7*</td>
</tr>
<tr>
<td>Private research</td>
<td>4</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>12</td>
</tr>
</tbody>
</table>

*Including 3 Russian Flagships and one Summer Language Institute

During the academic year, students in intensive courses are often traditional undergraduate students, including both Russian majors and minors and non-majors. Military students, such as Project GO participants, also make up student populations in intensive Russian courses, in both the summer and the academic year. One institution also mentioned graduate students as typical participants in intensive Russian courses, in both the summer and the academic year. Additionally, several institutions mentioned non-traditional and non-degree seeking students, especially in summer intensive Russian courses. With three schools reporting, the most common combination for student population during the academic year is traditional undergraduate students, Russian majors/minors, non-majors/minors, and military students.

Table 4. Levels taught intensively (not including study abroad).*

<table>
<thead>
<tr>
<th></th>
<th>Academic Year</th>
<th>Summer</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA/no intensive courses offered during the term</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Beginner</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Intermediate Low</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Intermediate High</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Advanced</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

*Levels are labeled “Beginner, Intermediate Low,” etc. rather than “First year, second year,” etc., to allow for a more accurate descriptive approach to the levels being taught.

During the academic year, the most common level of Russian that is offered intensively is Beginner, with ten schools reporting this option, as
evidenced in Table 4. Beginner is also the most common level offered in the summer (n = 6), closely followed by Intermediate Low (n = 5). Intermediate High and Advanced are not offered as commonly as the lower levels, in either the academic year or the summer.

Participants reported a range of options regarding the total hours per week and total number of weeks that intensive Russian courses meet. Responses ranged from the lower end of five hours per week for 16 weeks to eight hours a week for 15 weeks. This evidence points to the fact that there is no single standard for semester-long intensive courses in U.S. university-level Russian programs: five hours per week might be considered a regular course at one institution and an intensive one at another. By contrast, summer intensive Russian courses appeared to be more standardized in length, with four institutions reporting courses meeting for 20 hours per week for eight weeks.

Ten institutions reported using the ACTFL guidelines as benchmarks for student progress in intensive Russian courses, while two institutions reported not using them.

Table 5. Specifics of OPI administration.

<table>
<thead>
<tr>
<th>Program Profile</th>
<th>Specifics:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>Live OPI  1</td>
</tr>
<tr>
<td></td>
<td>OPIc      1</td>
</tr>
<tr>
<td>Private</td>
<td>Either live or OPIc 1</td>
</tr>
<tr>
<td></td>
<td>OPI(c) for some students 2</td>
</tr>
<tr>
<td></td>
<td>For all students 3</td>
</tr>
<tr>
<td></td>
<td>For Flagship students 2</td>
</tr>
<tr>
<td></td>
<td>Throughout the semester 1</td>
</tr>
</tbody>
</table>

*Including 1 Flagship  **Including 2 Flagships

Half of the institutions that teach Russian intensively administer OPIs at the end of the course, as illustrated in Table 5. Five institutions administer official OPIs (either computerized or live for some or all students), and six administer informal OPIs, which are conducted inhouse for training or informal assessment purposes by someone familiar with the
ACTFL guidelines and the OPI technique but not necessarily certified. For levels above elementary, four institutions reported administering either an official or an informal OPI at the beginning of the semester; two institutions conduct OPIs with all students initially, and two administer OPIs to only some students.

For non-intensive courses, six institutions reported using informal OPIs during the semester. One institution uses informal OPIs in all levels of Russian, while four institutions use them for advanced levels. One institution reported using unofficial OPI scores for graduate student classes and assessment.

In overall program assessment, more than half of all respondents (n = 14) reported using some type of OPI as a form of program-final functional language assessment, as represented in Table 6. The split is roughly even between official and unofficial OPIs. More public research universities (n = 9) use OPIs as a form of program-final assessment than private research universities (n = 4) or small liberal arts colleges (n = 1).

Table 6. Administration of the OPI or other tests as a form of program-final language assessment.

<table>
<thead>
<tr>
<th>Program Profile</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>9</td>
</tr>
<tr>
<td>Private</td>
<td>4</td>
</tr>
<tr>
<td>Small liberal arts</td>
<td>1</td>
</tr>
</tbody>
</table>

Most institutions that do not administer OPIs as a form of program-final assessment reported using traditional pen-and-paper tests (n = 9), such as final exams. Four institutions reported a different type of speaking test, similar to the OPI but not necessarily based on the ACTFL guidelines or...
the OPI structure. Finally, two programs reported using the Avant STAMP test, an online proficiency test in all four modalities based on the ACTFL guidelines.6

Institutions that reported not administering OPIs as a form of program-final functional language assessment listed cost as the main factor affecting this decision. Five institutions reported a lack of other resources for OPI administration, and six reported a lack of instructor training in the ACTFL OPI. Notably, two participants stated that OPIs are not administered because the OPI is not an effective measure of student progress. Other responses included the need for testing all modalities as a reason for not conducting OPIs as a form of program-final assessment. Approximately half (n = 6) of the institutions that do not administer program-final OPIs are interested in initiating this practice.

These survey results suggest that proficiency-oriented practices are being implemented by a considerable number of post-secondary U.S. Russian language programs. With almost half of our participants teaching intensive Russian courses, of different levels and to various student populations, and with the majority of these institutions using the ACTFL guidelines to measure student progress in these courses, one can say that the proficiency orientation has been gaining ground in Russian programs. The OPI is being utilized as a measure of speaking proficiency by more than a third of the programs in our sample, with nine institutions using it in either intensive or non-intensive courses, or both. Finally, 14 out of 26 institutions surveyed reported using some type of OPI as a form of their program-final functional language assessment, with several more using other proficiency-based speaking tests. Six institutions stated their interest in starting to use the OPI for students at the end of the students’ language programs. We conclude from this evidence that despite its drawbacks (such as lingering doubts regarding its validity and the cost of official testing), the OPI is growing in popularity across U.S. university-level Russian programs.

5. Teaching for speaking proficiency: two programs’ results and practices
The second part of our study represents action research—an approach to investigating practical aspects of teaching by the practitioners themselves (McNiff 1993). In this section, we scrutinize our own experience of making

---

6 The Avant STAMP (Standards-Based Measurement of Proficiency) is an online, computer adaptive language test in reading, writing, listening, and speaking that was developed at the University of Oregon. For more information, please see https://avantassessment.com/stamp.
the transition from working almost exclusively with academic learners within more traditional post-secondary language curricula to supervising or teaching in federally funded language programs, aimed primarily at “the government learner.” Oleksandra Wallo, a faculty member in the Slavic department of a large public university, has designed and supervised a university-run off-site intensive Russian program for military personnel (Program 1). Molly Godwin-Jones, an advanced graduate student in the same department, has served as the lead instructor for Project GO summer intensive first-year Russian courses (Program 2). Learners in both programs enrolled as students in short, intensive university courses (8-, 11-, or 14-week long). By the end of each course, learners were required to demonstrate gains in speaking proficiency by taking the official OPI or OPIc. Ensuring that our students in these programs were able to meet the set proficiency benchmarks necessitated curricular restructuring and provided occasion to experience the OPI washback effects first-hand.

Our investigation was guided by the following research questions:

1. What percentage of learners in each intensive Russian program were able to meet the set speaking proficiency benchmark (as measured by the course-final official OPI or OPIc)?

2. What pedagogical principles, used in the two proficiency-oriented intensive Russian programs, helped students meet the speaking proficiency benchmarks and why were they effective?

3. What washback effects did the mandatory course-final official OPI or OPIc have on the programs’ curricular design and instruction?

5.1. Program results in speaking proficiency

Intensive elementary Russian courses with no study abroad component in Program 1 lasted either 11 or 14 weeks for a total of approximately 330 contact hours per course, with 6 hours of instruction on most days. There were usually two parallel sections in the course, with 3–7 learners per section. All students were adult learners and true beginners in Russian. By the end of the course, learners were required to demonstrate that they have reached ILR level 1 in speaking by taking a double-rated official telephonic OPI. The program did not focus exclusively on oral communication, but rather provided well-rounded training in several skills. Besides making gains in oral proficiency, learners were also expected to reach ILR level 1
in reading and listening, which they demonstrated by taking the Defense Language Proficiency Test (DLPT).\footnote{A detailed discussion of reading and listening proficiency is beyond the scope of this article.}

As Table 7 shows, the success rate of the program in meeting the benchmark for speaking proficiency was high. Of the 55 students who were tested with the official OPI, over 96% reached ILR level 1 (ACTFL Intermediate Low) and 43% exceeded this benchmark, with most reaching ILR level 1+ (ACTFL Intermediate Mid or High).

Table 7. Course-final OPI results in Program 1 (Data from October 2016 to December 2018).

<table>
<thead>
<tr>
<th>Contact Hours</th>
<th># of Students</th>
<th># Tested</th>
<th>ILR Goal</th>
<th>0+</th>
<th>1</th>
<th>1+</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>330</td>
<td>57</td>
<td>55</td>
<td>1</td>
<td>2</td>
<td>29</td>
<td>23</td>
<td>1</td>
</tr>
</tbody>
</table>

Intensive elementary Russian courses with no study abroad component in Program 2 lasted eight weeks for a total of 160 contact hours, with about 3 hours of instruction and 1 hour of one-on-one tutoring each day. There was only one section each summer that included three or four Project GO students and one or two non-Project GO students. Project Global Officer is specifically geared towards ROTC students and “aim[s] at improving the language skills, regional expertise, and intercultural communication skills of future military officers” (“About Project GO”). The courses trained all four skills, with a somewhat greater emphasis on speaking, especially in the final two weeks, as students prepared for a course-final official OPIc. This was the only standardized proficiency test administered at the end of Program 2. Another difference between the programs was the amount of time spent on homework, with only 3–4 hours per week for Program 1 students and about 15 hours per week for Program 2 learners.

Program 2 also demonstrated impressive course-final OPI results. Even though Project GO students are expected to reach the speaking proficiency goal of ILR 1 after four semesters of language study or its equivalent, the majority of students (82%) in Program 2 tested at this level after only one 8-week intensive summer course (equivalent to two semesters). As Table 8 shows, two out of 11 students were even able to achieve ILR level 1+ after just one summer of intensive study.
Table 8. Course-final OPIc results in Program 2 (Data from summers of 2017, 2018, and 2019).

<table>
<thead>
<tr>
<th>Contact Hours</th>
<th># of Students</th>
<th># Tested</th>
<th>ILR Goal*</th>
<th>0+</th>
<th>1</th>
<th>1+</th>
</tr>
</thead>
<tbody>
<tr>
<td>160</td>
<td>11</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>

*Project GO proficiency goal after four semesters of Russian or its equivalent

Below we discuss the pedagogical principles that drove speaking instruction in these programs and some of the teaching practices that evolved from year to year during our work there. We adopted these instructional practices because the OPI was set as the final assessment for our courses by external parties. Yet both our program outcomes and existing research on these practices confirm that they are effective in facilitating the development of speaking proficiency—regardless of the type of final assessment used. Some of these practices were recently designated by foreign language education researchers as belonging to High-Leverage Teaching Practices, or HLTPs (Glisan and Donato 2017), and the positive impact of these practices on students’ proficiency development is beginning to be closely examined (Vyn, Wesely, and Neubauer 2019).

5.2. The flipped model
Because of the challenges which the Russian alphabet, pronunciation, vocabulary, and grammar present to Anglophone students, it is easy for most of class time in traditional beginner courses to be taken up by more rudimentary work on these building blocks of language. Even though the new generation of elementary Russian textbooks do contain a considerable number of communicative activities, Comer’s (2012) quantitative analysis of communicative and “language-focused” exercises in select chapters from Golosa, Troika, Nachalo, and Live from Moscow revealed that the latter activities significantly outnumber the former (147–8). If an instructor were to closely follow one of these textbooks, it would automatically result in much class time devoted to “the teaching of language form” (Comer 2012, 156).

One solution, implemented in intensive courses by language programs at the University of Texas at Austin, has been the flipped model. As Garza (2017) explains, in this model, “the bulk of overt instruction of grammar, phonetics and lexicon is relegated to outside of class time,” whereas time in class “is spent almost entirely in actively engaging the
students, individually and in groups” (231). Although instructors are sometimes concerned about their students’ ability or willingness to adequately process new grammar and vocabulary on their own, research on flipped language instruction has repeatedly found the flipped model to be at least as effective as and often superior to traditional instruction for the learning of both grammar (Al-Harbi and Alshumaimeri 2016; Moranski and Kim 2016; Webb and Doman 2016) and vocabulary (Mori, Omori, and Sato 2016; Alnuhayt 2018; Kırmızı and Kömeç 2019). For grammar learning, Tonkin, Page, and Forsey (2019) link the advantages of the flipped approach to the fact that the latter reduces students’ cognitive load by allowing them to work through new grammar at their own pace and in advance of being required to use it in cognitively more demanding communicative activities with peers. There is also emerging evidence of the flipped language classroom’s positive effect on students’ speaking skills (Russell 2018; Sabahattin 2019).

The flipped classroom approach is built into Mezhdu nami, one of the newest elementary Russian language textbooks on the market. This textbook was utilized in Program 2. Before coming to class, students read and listened to short dialogues on the Mezhdu nami website and then worked through the accompanying reading comprehension and grammar sections, which include self-check exercises. Students were then required to complete an Online Review Check (ORC), consisting of multiple choice, matching, and cloze activities, posted by the instructor on the course management system. The ORCs were graded, and the ORC scores were included in the overall homework grade. Because the questions on the ORCs were not overly challenging and mirrored the grammar explanations or self-check exercises from Mezhdu nami, they were a quick and convenient tool to reinforce the new material and to hold students accountable for learning the grammar on their own, as advocated by Mori, Omori, and Sato (2016). Instructors could look at the results of the ORCs to see if any questions were frequently missed and needed to be discussed in class the next day. In the experience of Godwin-Jones, students were usually able to complete the ORCs and rarely required additional instruction. Having thus ascertained before the class that students had a grasp of the assigned building blocks of language, the instructor could engage them in pair or group communicative activities from the first minutes of class.

Besides following the flipped model, Mezhdu nami prioritizes class activities that simultaneously focus on the communication of meaning and on language form. These contextualized input and output activities for elementary Russian, based on the principles of input processing (VanPatten
1996), constitute a major pedagogical innovation of this textbook. These types of activities are built in such a way that in the process of generating meaningful utterances, however short, students are prompted to notice how particular grammatical forms encode specific meanings. As a result, students acquire key grammar structures as they engage in communication, which makes grammar subordinate to the development of speaking skills.

5.3. Deliberate vocabulary learning at home

One aspect of the flipped approach that deserves special attention is students’ work on vocabulary. Nation (2008) suggests that the first 2,000 most frequently used words are best taught directly. According to Hacking and Tschirner (2017), Russian L2 learners’ receptive mastery of the most frequent 1,000 to 2,000 Russian words tends to correspond to intermediate-range reading proficiency on the ACTFL scale. Therefore, we advocate explicit, targeted vocabulary teaching and learning at the lower levels. As Meara (2005) explains, the first step in knowing a word is being able to recognize that it is in fact a word. Explicit vocabulary practice helps students begin to recognize word shapes. Frequency of exposure to a word also plays a significant role in vocabulary learning. According to Nation (2008), “[w]e need to see the learning of any particular word as being a cumulative process where knowledge is built up over a series of varied meetings with the word. At best, teaching can provide only one or two of these meetings” (97). Several other meetings can happen during students’ work with the textbook—if the latter recycles vocabulary sufficiently from one unit to the next. For instance, Comer (2019) found a 38% carry-over rate for vocabulary in Mezhdu nami. While this percentage is significant, it still leaves it up to students to learn the remaining vocabulary. Consequently, students need out-of-class settings for additional word meetings, which can be provided to them through many freely available online vocabulary-learning tools, such as Quizlet.

Students in Program 1 spent the few weekly homework hours learning vocabulary through Quizlet, where instructors created sets of active vocabulary for each topic. Program 2 used a collective Quizlet folder to make the vocabulary-learning process less intimidating and to help students practice good study skills. Rather than having each student create his or her own set of online flashcards for each unit, the instructor facilitated a group flashcard-making system. Students had to sign up for a certain number of vocabulary sets over the course of the semester, depending on the number of students in the class. The instructor provided all students with a daily vocabulary list containing active words for that
class day. The student assigned to Quizlet was responsible for converting the list into a set of online flashcards and sharing the set to the class folder, which automatically granted access to other class members.

5.4. Setting ambitious speaking goals and building in accountability
Having flipped the classroom by assigning most of the vocabulary learning (and grammar learning in Program 2) to homework, instructors freed up class time for using the material learned at home to communicate in pairs or groups. While these are always challenging activities for beginners, setting clear, specific, yet ambitious speaking goals for each textbook unit and then developing carefully scaffolded tasks for students through backward design ensures that even novice learners can spend much of the class communicating in Russian. This approach was adopted in Program 1, which used the topic-based lessons from a popular elementary-level Russian textbook (Troika) for some language input and overall course structure but focused on communicative tasks designed to meet the program’s benchmark in speaking proficiency. During each week of this intensive course, students worked on one or two major topics, such as daily routines or leisure activities. The speaking goals set for each topic purposely exceeded both the chapter-final checklists of abilities given in Troika and the speaking functions typical of ILR Level 1. For example, after working through the unit on daily routines during the week, students were expected to do more than simply “ask and answer questions about the time when some activities take place” (Nummikoski 2012, 202); they were expected to produce very short narrations about their weekend routines. Although not easy to reach for beginners, such a goal made sense to the learners because it described a useful ability in a topic area of interest to them. (Everyone wanted to talk about the weekend after spending many hours per day in the classroom!) As a result of this ambitious goal setting, more than a third of the learners in Program 1 routinely exceeded the speaking benchmark set for the course, testing at ILR level 1+ or above on their course-final OPI.

Students were held accountable for reaching the communication goals for each unit through weekly speaking sessions, usually held on Fridays. During each thirty-minute session, someone other than the students’ instructor would speak to each learner by telephone, engaging them in a conversation that tested the student’s ability to converse on the

---

8 Program 1 funding made it possible to employ graduate teaching assistants to conduct the weekly speaking sessions, but it is possible to arrange such “external” assessment even without additional funding—for example, by swapping instructors between sections of the same course one day a week.
topic(s) covered during the week. The fact that students had to speak to someone outside their class community on a weekly basis both increased their motivation to reach the weekly speaking goals and provided them with valuable experience ahead of their official telephonic OPI.

5.5. Interactional speaking at the center of classroom activities
In Program 1, the week's activities were designed to build up to the weekly goals and were carefully scaffolded from day to day. The study of the core vocabulary for a new topic would begin at home, over the weekend. On Monday and Tuesday, students would work in pairs to read short dialogues, taken from the textbook or prepared by the instructors, that contained this new vocabulary. In groups, they would also read about the week's topic in short texts, which contained more occurrences of the new vocabulary, and discuss comprehension questions. While not yet fully communicative, these activities flooded the students with relevant input and gave them opportunities to practice pronouncing the new words and recalling them while responding to comprehension questions. On Wednesday and Thursday, students would spend much class time discussing the topic in pairs following general English prompts written by the instructors. Learners would also complete role-play scenarios on the topic. Toward the end of the week, students would take a quiz on the active vocabulary of the week; complete end-of-unit oral exercises from *Troika*, which are focused on production; and take turns conversing on the topic one-on-one with the instructor, in preparation for the Friday speaking sessions by telephone.

Besides the speaking activities described above, two further types of oral communication tasks utilized in Program 1 were found particularly effective in preparing students for the speaking interactions characteristic of the OPI at lower proficiency levels. The first type was dubbed “the hot seat”: students would take turns sitting in front of the class and answering a series of questions from their peers on the topic(s) studied. The class was instructed to ask questions in a logical sequence, as they might be asked in a real interview, and to build on the responses given. The students in the hot seat were permitted to assume different personas and get creative in their responses to keep the activity interesting and maintain a real information gap. As a result, this activity would meet the main criteria for effective oral interpersonal tasks described by Glisan and Donato (2017): it was engaging, it required everyone to listen to what others were saying, it was highly interactive, and it often involved negotiation of meaning (55-56). The second type of activity involved bringing a guest to class—either a native or a highly proficient speaker of Russian—and placing the guest...
in the hot seat. This activity was even more motivating for students, as it would closely resemble authentic communication.

While critics might say that several of the activities described above target only a very limited range of functional abilities (i.e., oral communication in the interpersonal mode) typically tested by the OPI and are therefore nothing other than teaching to the test, we would stress that these abilities are at the core of the language learning process. The first two of the six high-leverage teaching practices (HLTPs) in foreign language education, singled out by Glisan and Donato (2017), are “facilitating target language comprehensibility” and “building a classroom discourse community.” Both of these practices are enacted in the activities that helped our students develop their speaking proficiency in the interpersonal mode, thus preparing them for the course-final OPI. Furthermore, as Glisan and Donato point out, giving a prominent place to oral interaction in the classroom makes sense for other reasons, such as the fact that it “will mediate the learning of other aspects of language” (56). We therefore consider such an emphasis in language instruction enabling rather than limiting.

5.6. Targeted OPI preparation

Toward the end of Program 1 and Program 2, we did introduce a small number of more targeted OPI practice activities, following the principle of teaching for how our students will be tested. For students in Program 2, Godwin-Jones used four class sessions to facilitate preparation for the official OPIc. Before engaging in any speaking activities, students reviewed the OPI structure and level descriptions, and the instructor explained what to expect when interacting with the avatar.

One of the most important skills to target in focused OPI preparation is students’ ability to speak with greater ease. In a traditional classroom setting, students may feel that they should only speak if their utterance is “correct,” or they may be uncomfortable or fearful of speaking. OPI practice activities need to be scaffolded in such a way as to promote students’ focus on fluency, even if there might be grammatical errors. Even in truly communicative classrooms, it is unlikely that novice students are accustomed to speaking independently for longer than a minute, as they may have to do during the official OPI, especially when they are probed at a higher level. Giving students the safe space to overcome any fear of speaking and to practice it in a low-stakes situation is an important aspect of these targeted OPI practice sessions.

To aid students in creating more sustained speech, Godwin-Jones used grammar-function pairings with specific lexical topics to help novice
learners create meaningful utterances. Each activity began with either a lexical or grammar recall warm-up. Lexical warm-ups could include word popcorn (instructor states a topic, then each student must say a vocabulary item associated with this topic that has not been previously mentioned; whoever cannot think of a word is out until only one student is left); creating pros and cons lists; and using visual prompts or diagrams. Grammatical recall warm-ups required students to think of specific grammar constructions (i.e., y + genitive case) and “translate” them into speaking functions (i.e., to talk about who owns what). After some warm-up time, students would pair up to speak for as long as possible on a specific topic, incorporating the material reviewed during the lexical or grammatical warm-ups. One partner would speak while the other would keep the time, and then they would switch roles. Students had to write down their time after each session. Partners were not allowed to interrupt or correct each other, nor were they allowed to ask for help from the instructor.

While the timed speaking activities described above aim to develop fluency, instructors in Program 1 used another form of targeted OPI practice to develop accuracy of speech. Towards the end of each thematic unit, students wrote short essays responding to typical OPI prompts, such as describing one's apartment, house, or city. They would then meet with the instructor one-on-one to discuss patterns of errors they made in the essays and to audio-record a sample mini-conversation on this topic with the instructor. For this recording, the student and the instructor would switch roles, with the student playing the role of the interviewer and the instructor responding to the student interviewer’s questions based on the information in this student’s essay. The purpose of this activity was to give the student a useful model of phonetically, lexically, grammatically, and, in this case, even factually accurate responses to potential OPI questions. Recorded using the free Audacity software, these mini-conversations were then listened to by students for review and further preparation before the official OPI.

6. Conclusion
Even the staunchest critics of the OPI have acknowledged its value as one of the motivating factors to purposely and persistently “work toward development of conversational skills” (Meredith 1990, 295). In the intensive Russian programs described in this study, we found that having the official OPI as a course-final assessment measure had positive washback effects on teaching and learning alike. It forced a focus on target-language communication in the classroom, compelling program supervisors and instructors to consistently rely on the flipped model in
order to make more time for many and varied interactional speaking activities. It also prompted all parties involved to analyze speaking in the way hierarchized by the ACTFL Proficiency Guidelines: for instructors, the guidelines helped better structure in-class speaking activities and provide appropriate scaffolding; for students, the guidelines helped demystify this often intimidating language skill. Having a course-final OPI or OPIc built accountability into the courses and provided students with a standard to reach or even surpass. Consequently, it fostered what Language Flagship Program administrators have called “a culture of high expectations” (Murphy et al. 2017, 33), motivating students to take risks and challenge themselves—which yielded impressive speaking proficiency outcomes.

Our experience of working in the intensive format is in line with previous findings about the benefits of intensive courses for oral proficiency development. However, more empirical research is needed to explain why intensive instruction seems to yield better oral proficiency outcomes and to determine what degree of intensity can produce the best results. One limitation of most research on this issue conducted thus far, including of our study, is the fact that it focuses on lower-level courses. The rapid progress of the students in our intensive courses would likely look very different at the advanced level, and a course-final OPI might not be able to capture it. Thus, the preference among the participants in our survey to use the OPI for program-level rather than course-level assessment is justified.

Yet, as our experience demonstrates, the OPI can be a very useful tool, especially in beginner-level intensive courses. Building the inhouse capacity to administer official, or even informal, advisory OPIs by paying for instructors to get certified or to at least take the OPI training workshop is a smart investment for language programs that seek to become more proficiency-oriented and better guide their students in acquiring oral proficiency in Russian.

References


Murphy, Diane, Karen Evans-Romaine, Valerie Anishchenkova, and Zhuo


To OPI or Not to OPI: Proficiency-Oriented Instruction and Assessment

Wallo, Godwin-Jones


