

THE LEARNING AND TEACHING OF SLAVIC LANGUAGES AND CULTURES

EDITED BY

**OLGA KAGAN &
BENJAMIN RIFKIN**

WITH SUSAN BAUCKUS

Bloomington, Indiana, 2000



From Testing to Assessment, From Teaching to Learning

Thomas R. Beyer, Jr.

The state of testing in Russian and other Slavic languages in the United States has closely paralleled student interest in those languages. Historically new tests appeared with increased enrollments and available funds in the national defense budget. In the 1960s the Modern Language Association offered tests of reading and listening comprehension and of written and spoken Russian at three different levels. Until the 1970s Russian was one of the subject areas tested by the SAT Achievement series (now called SAT II's). In the 1980s the Educational Testing Service produced a series of Russian language exams based on authentic language and scaled to the Proficiency Guidelines of the American Council on the Teaching of Foreign Languages. By the 1990s, when language enrollments had declined in Russian and it and other Slavic languages disappeared from the national consciousness as a defense priority, none of these tests were available. The only standard nationally recognized instrument used to measure proficiency is the Oral Proficiency Interview (OPI) based the ACTFL scale and guidelines developed for Russian in the 1980s (1984, 1988) (Rikin 1998: 16).

Carol Klee describes the shift from audio-lingual discrete point testing in the 1960s toward communicative competence and functional use of language measured holistically by the OPI. Irene Thompson refers to the ETS exams (1986) and provides examples that in themselves were revolutionary in their focus upon proficiency, authentic language and authentic tasks. Klee envisions a future of "alternative assessment" that will be interactive, rely upon standards, and be learner centered. Moore and Bond show us a glimpse of that future in action using portfolios to raise several issues worthy of the profession's attention.

Klee asserts that even as "instructors began to incorporate some of the tenets of communicative language teaching into their classrooms, foreign language testing began to change as well" (this volume: 229). But she notes "[i]nnovations in testing, however, have lagged behind innovations in teaching" (this volume: 229). The assertion is no doubt true, but focuses exclusively on the impact of teaching on testing. Course content and the manner in which

language is taught clearly influence what and how that language is tested. Our students—the successful ones—develop early a sixth sense that has them study what they will be tested on. These students develop language learning strategies based on the tests. For many, if it is not tested, it is not worth learning, and to the extent that tests are evaluative, these students are making reasonable assumptions and sound decisions. Moreover, teachers responsible for constructing exams tend to test what they teach. At the classroom level, testing or assessment can explicitly address the areas that the teacher values. And as Moore and Bond demonstrate, students can participate actively in their own assessment in the portfolio process, where they determine in consultation with the instructor what goals they themselves hope to achieve.

When teachers are not directly responsible for the construction of the test, as in the case of any national standardized exam, many tend to teach to the test. Thus there already exists a close relationship between teaching and testing. Whether or not such instruments represent a broad consensus of what should be taught, learned, tested or assessed, national exams such as the SAT II's and the Advanced Placement exams of the College Board have significant impact on subject matter and skills that students must acquire to perform well. Here what is tested frequently dictates what is taught. Lacking either of these instruments for Russian or the other Slavic languages, many teachers welcomed the publication of the ACTFL Proficiency Guidelines (1982) and the existence of the Oral Proficiency Interview which directed attention to oral skill development in classrooms around the country.¹

In the field of Slavic, materials development supported by publishers has traditionally has lagged behind the more popular languages (French, German, Spanish); yet change in the classroom has nevertheless come rapidly in this past decade. Several new proficiency inspired texts address the issue of communication. Products of the 1990s, they have moved beyond reflecting a bygone Soviet reality.² This very aspect of change in the societies, cultures and languages of

¹ The word "proficiency" is frequently abused, having become the marketable catchword of many methods, exams etc. It is curious to note that the MLA used the title Cooperative Foreign Language Proficiency Exams in 1961—long before the term came into vogue in the 1980's for specific proficiency measurements derived originally from the U.S. government's scale for job performance readiness.

² I was reminded of this last spring when we used an older exam for practice in multiple choice questions. The students unanimously complained of a frequently used word that they had never learned or heard. The word "tovarišč" was a mainstay of socialist societies, but the newer generation rarely encounters "comrades." How many more of these, "Komsomol," "Pioneers," "Communist Party," "Central Committee," "Supreme Soviet," are linguistic relics that need historical notes if they are to be a part of our students' language?

Eastern Europe and the former Soviet republics may explain in part the survival of the OPI. Designed and modified on the spot for the individual test taker, the interview does not rely on any set of vocabulary, grammatical or cultural items for it to be administered. Other proficiency-based exams based on authentic language, for example, on newspaper articles and sound bytes from radio on sports or foreign affairs, age rapidly. The ETS Proficiency Exams constructed in the 1980s are now hopelessly outdated and would have been retired on that account alone. There were simply no financial resources for the constant updating that such exams require. In addition, the tests did not generate sufficient revenue to justify their administration, scoring and management. This financial reality continues to stand as an impediment to the development and maintenance of a nationally devised test by the College Board, ETS, or ACTFL. Russian and the other Slavic languages simply have enrollments too small to justify the cost of such undertakings.

Standardized proficiency testing in Russian was a bright page in our profession in the 1980s where Russian led many of the other languages in test development and design. Irene Thompson, a key figure in developing the ACTFL Proficiency Guidelines, the ACTFL Russian Proficiency Guidelines and the ETS Proficiency Exams, provides an invaluable historical appraisal along with a sober assessment of student performance, as she explores the relationship between time in class and proficiency, correlations between proficiency skills, and the problems of setting proficiency exit requirements. Her test takers are college/university students with one to five years of Russian language study and are tested using proficiency based instruments to measure speaking, listening, reading and writing abilities. While the results to many will not be surprising, they are disturbing. After three years of study only 25% of the students achieve Advanced (Level 2) proficiency in speaking. Only after five years of language study did 75% of the students score Advanced or Superior (Level 3). Thompson does not indicate which of these students had a significant immersion experience either in a Russian speaking country or here in the United States (although her higher level students were attending the Middlebury Russian School characterized by a pledge to use only Russian for the six week session). Interpreted another way, over a four-year language sequence there was little to no functional ability to use spoken Russian in the workplace beyond basic survival and conversational needs.

For reading, students took a modified version of questions similar to those first introduced in the ETS Proficiency exams. Results indicate that students begin to acquire Advanced proficiency when they move from the first and second year of language instruction, much of it heavily focused on oral centered tasks, to third year college classes and beyond where the emphasis shifts to actual reading. These findings reinforce the notion that students learn to do what

they practice doing. Nonetheless, even after five years of Russian more than half the students read below the Advanced level. Listening and writing skills were likewise disappointingly low, but here Thompson questions the adequacy of the test instruments. One might add to her comments that the content for the Russian listening exam presumed a depth and breadth of vocabulary rarely encountered in literature courses.

Thompson's article provides a thoughtful look into the problems of evaluation of proficiency, where even the ratings on the OPI tended to differ significantly between the tester, who was simultaneously conducting the interview, and those who listened to a tape of the interview for a second rating. In spite of the problems associated with the tests of other skills, the sample questions presented (Thompson, this volume: 283-84) indicate a real departure from discrete-point testing of the past toward a new focus on proficiency. The reading passages, for example, pose the questions in English ensuring that the student understands the task. The sample listening passage came from an actual recorded interview. The writing prompts require actual written statements to communicate information at the different levels of language ability. Thompson notes that the *ACTFL Guidelines* themselves are due for revision based on research and findings of the past ten years, and the new guidelines may serve as springboards for more effective assessment of language skills. Much of this focus on "proficiency" of the past ten years, however, is likely to be superseded by recent pedagogical developments and the need to assess student performance against emerging national standards.

Klee also reviews the OPI and cites some legitimate criticisms, including that the guidelines themselves were based on native speaker performance. In fact, the ACTFL levels do not adequately indicate progress even over substantial periods of classroom time—a fact corroborated by Thompson's study—and were never intended to measure student achievement in traditional language learning courses. Klee does not mention other constraints on the widespread implementation of such exams in the classroom: time, money, and the effort needed to become a certified tester. A face-to-face interview may last from ten to more than thirty minutes per individual. Listening to the tape and making an evaluation requires a similar amount of time. An official ACTFL certified interview presently costs \$115. Certified testers require extensive training, regular practice and undergo frequent, rigorous and expensive recertification. Klee does note the successful implementation of modified proficiency based testing and language requirements including the University of California Language Ability Assessment System. This model of real-world texts and tasks, Klee suggests, can also be adapted for designing exams of communicative competence at all levels.

Never intended as a pedagogical movement, adoption of "proficiency" had a significant positive influence on teaching and curriculum, what Klee calls "washback." Curricular reform notwithstanding, the realistic performance goals in Russian and other Slavic languages were set at Intermediate Low (Level 1) for the productive skills of speaking and writing. This was lower than the expectations in French, German and Spanish and as Thompson notes, in "setting proficiency requirements associated with different levels of study...the issue is the percentage of failure the profession is willing to tolerate" (this volume: 278). Standards that require only Intermediate level performance may be meaningless or useless outside of academia. Klee cites recent developments offering a new vision of testing that includes self-evaluation, peer assessment and portfolios. These "alternative assessments" are now widely used in other disciplines at the primary and secondary school levels. A key to alternative assessment is the focus on progress, not only what one can do—the product—but on the process of improving that performance. Klee outlines several tasks: role plays, interviews, skits for speaking; journals, e-mail correspondence, research reports for writing; skimming and scanning stories or articles for reading. All require real life authentic tasks and authentic text materials that themselves could be assessed in this new way via interactive authentic challenges.

Klee's theoretical overview finds concrete illustration in the description of a portfolio-based instructional model of which Moore and Bond provide an example and an indication of the possibilities and potential problems. The portfolio model addresses the needs of middle school (7th and 8th grade) students. The goals included "to recruit and retain" students for continued study of the language. In our own time, these goals will resonate at all levels of instruction. The portfolio program acknowledges the centrality of students in the learning process. Students permitted to set their own goals and account for their own learning styles should be motivated learners. Learner involvement began with establishing the guidelines for the portfolio and its content. Based on the goals of the "recruiting and retaining" the program was a complete success: all twenty students who enrolled in this particular Russian course continued their study into high school. The opportunity to succeed in a language as measured by portfolios clearly enfranchises and validates the work that students do. There was much in this experimental exercise that is praiseworthy: the focus on learning as opposed to teaching, the interdisciplinary and interactive nature of the work, the involvement of students, instructors and parents in a learning partnership.

Nonetheless, while wonderful in theory, the exercise described was problematic in practice. The authors themselves recognized the need to modify their own instructional and assessment model. The very goals have little direct rela-

tionship to linguistic skills development or language proficiency. The tasks agreed upon by students seemed to have little correspondence to the skills supposedly under development. The speaking assignments included reading aloud three passages, one of which was a poem by Puškin. The instructor notes the lack of communicative skills involved in the major projects, some of which were in part dictated by the need to prepare students for the U.S. Olympiada of Spoken Russian (another example of national testing directing curriculum). The writing component of the portfolio devoted significant time to questions of graphics and layout, which are extra-linguistic tasks. While these are useful skills they are not directly related to improving language skills. Portfolios in and of themselves do not guarantee excellent measurement instruments. The validity of portfolios is greatly dependent upon professional design—something that students may simply be incapable of providing. Another issue is that of ultimate responsibility for setting effective challenges. Even when students were permitted a voice in the establishment of assignments, a majority of them failed to meet their own or their teacher's expectations. In the first assignment, only nine out of twenty students worked the expected fifteen minutes or more. By the fifth exercise, only two out of twenty were working at the minimum amount of time expected by the teacher. A danger in portfolios and when students set standards is that they may aim too low—establishing minimal performance as the goal, rather than stimulating achievement resulting from standards slightly higher than the students' own expectations.

This portfolio system and the accompanying structure for learning and assessment are bold departures from the teacher-student model. It is student centered, and the teacher becomes a collaborator in the learning process. In spite of the shortcomings of the particular example documented by Moore and Bond, portfolios can and are being used effectively throughout the country when they are carefully constructed and designed. When professionals at the state or national level define standards of achievement that can be successfully modeled, portfolios give students something to aim at in improving their own performance. In combination with national standards that set high expectations, portfolios are one means of alternative assessment that is likely to be utilized ever more frequently in the next few years.

Let us now move past the scope of the three articles and recap the current state of testing/assessment and then look out beyond the horizon. There are any number of interesting assessment devices in use or development. The proficiency exam for writing used in the Thompson study was developed at George Washington University. Middlebury College developed an interactive computer proficiency exam of reading comprehension. The University of Wisconsin-Madison has a number of Russian tests for Grammar, Listening, and

Writing. The web site of the Center for Applied Linguistics (CAL)³ lists several institutions offering testing for a fee. New York University (NYU) offers testing and college credit in Czech, Polish, Russian, Serbo-Croatian and Ukrainian via their Foreign Language Proficiency Test of listening, reading, writing, grammar, culture, with aural, written and translation components.⁴ Brigham Young University (BYU) offers a Russian Achievement Test (FLATS), with parallel exams for Polish and Serbo-Croatian, designed to test all four language skills, using multiple-choice items, essays, and semi-direct oral testing procedures.⁵ At the national level, there are the Oral Proficiency Interviews and/or Simulated Oral Proficiency Interviews for Bulgarian, Czech, Polish, Russian, Serbo-Croatian and Ukrainian. Others clearly exist and this review does not even begin to gauge what is being developed in Russia for internal and commercial purposes. Moscow State University, to name one, has for the past few years offered to test the "proficiency" of students upon entrance into and exit from language programs at Russian universities.

At the pre-college level the best known and widely used instruments are not primarily tests: the U.S. Olympiada of Spoken Russian and the National Russian Essay Contest. These are nationally conducted competitions at the secondary level designed to motivate students and support interest in programs. At the same time, they offer a way to compare institutions across the country. The format and content of the tasks, as Bond and Moore indicate, and as corroborated in my own experience with high school students, directly influence decisions of high school and middle school teachers in their own selection of materials and emphasis on specific skills.

What about the future? We are currently poised to overcome traditional constraints of testing in the classroom. Ideal tests may not easily fit a one-hour class period. Nor is the classroom the only or best place to test. Recent advances in computer technology offer the promise of wide-scale computer assisted testing programs. Regardless of the new formats, all the test instruments mentioned have their limitations, as will no doubt any future exam. But I agree with Klee that "alternative assessment" will continue to gain wider support. At best

³ Center for Applied Linguistics <<http://www.cal.org>>. Other useful sites are ACTFL <<http://www.actfl.org>> and ACTR <<http://www.actr.org>> and <<http://www.russnet.org>>.

⁴ It can be used to fulfill foreign language requirements, gain advanced placement credit and credit waiver as well as to establish professional credentials such as teacher certification and other professional certifications. The exam can be administered at NYU or off-site. Two exam levels are available: one equivalent to four college semesters of language study, the other equivalent to advanced level credits.

⁵ It, too, provides credit by examination to beginning and intermediate college-level students and can be administered to both groups and individuals. Test components include a reading and grammar booklet, a listening tape and an oral test tape and booklet.

it should help students recognize their own strengths and weaknesses measured against clearly defined standards. Good assessment can help instructors to monitor content and their own methods. Results should be readily comprehensible and transferable across schools and across state lines and to a variety of audiences. It will involve self-evaluation and peer evaluation, and it will permit feedback that can be diagnostic and corrective. It will also be standards driven. The coming years will see a dialogue between what we assess and what our students learn, the one informing and influencing the other. The extent of that interplay may be determined by assessment of programs, and the need to demonstrate progress toward goals for which learners and instructors will be held accountable by local boards and institutions. While some things will change, some things won't. If we don't test it, our students will not expend the effort to learn it. If others don't test it, we ourselves are less likely to value those skills. We will encourage our students to learn what can and will be assessed.

But who will determine what is to be learned? Just as "proficiency" dominated the pedagogical *perestrojka* of the late 1980s and early 1990s, there appears to be a growing consensus at the elementary and secondary levels to embrace national standards. Such standards for foreign languages, including Russian, will drive the curriculum and determine the type of testing used in the decade to come. "Standards" have become a key way for schools, districts and states to measure and compare the progress of their students against others. This is not only an educational reform movement, but a political one with significant economic implications attracting the dollars for education at the federal, state and local levels. Given the federal support for identifying and implementing "standards" in schools across the nation, the inclusion of foreign languages in 1993 as the seventh and last subject area insured our participation in this national dialogue.

Standards for Foreign Language Learning: Preparing for the Twenty First Century under the directorship of June Phillips recognize that "Language and communication are at the heart of the human experience" (1996: 7). They identify the so-called "5-C's": Communication, Contents, Connections, Comparisons, Communities, as the essential components of the agenda for foreign language learning into the next century. The standards for grades K-12 (Kindergarten through the 12th grade) outline what students should know and be able to do, and they provide Sample Progress Indicators for grades four, eight and twelve, which define student progress in meeting the standards. For example, "Standard 1.1: Students engage in conversations, provide and obtain information, express feelings and emotions, and exchange opinions." At the 12th grade level the Progress Indicator expects that: "Students exchange, support and discuss their opinions and individual perspectives with peers and/or speakers of

the target language on a variety of topics dealing with contemporary and historical issues.”

The standards are less an indication of what students are currently capable of doing than the hope of what in the best of circumstances—including language education commencing in Kindergarten—students in the next century might be able to do. This same optimism is evident in the *Standards for Russian Language Learning* authored by Ruth Edelman, Peter Merrill and Jane Shuffelton (1998). Both the standards and the sample progress indicators can be measured and there will be increasing pressures to do so.

The national consensus on standards at the K–12 level still needs to be embraced by our colleagues in higher education. Unless the Higher Ed community knows and embraces the standards, there may be an even greater lack of articulation between high school and college than there is now. While the goal of many of our programs may be self-survival, recruiting and retention tools are always welcome. But other factors will have to be considered in the setting of goals for our students’ performance. These goals will have to account for the skills we require of our students, or they expect to learn from us. This may be to read Russian texts in the original, or to develop spoken and listening skills sufficient to permit study and communication with native speakers in the foreign country. When many metropolitan areas have Russian television and newspapers, and when the Internet provides access to live or taped radio and video broadcasts as well as print media, the ability to understand authentic Russian in a variety of contexts is a highly desirable skill. There may also be some emphasis in the standards on an appreciation of Russian or other Slavic languages as a system for expression of human thought. It is not yet clear how the new standards for Russian coincide with any of these advanced goals, and yet given preparation in these areas, future generations of students will look for familiar materials and skills emphasis.

What are we measuring for and what are we measuring? Traditionally most tests were means of control, to ensure that students are “doing the work.” On achievement exams, students demonstrate their knowledge of what we have taught—vocabulary, spelling, grammar skills. Proficiency exams measure what a student can do, ignoring when or how that skill has been acquired. According to Thompson, there is some, but no direct, correspondence between the number of years spent in a language classroom and the level of proficiency attained. Tests have also been used for identifying and ranking, accepting or rejecting students where backgrounds and language acquisition may have differed. The emphasis has been in eliciting information useful to the teachers or the testers.

Assessment can and should be more than that. It can be a diagnostic tool indicating students’ strengths and weaknesses as well as pointing toward improved performance in those areas. Good teachers have for years understood

how they themselves learn from tests taken by their students. Sophisticated tracking of answers for diagnostic purposes is for Russian and the other Slavic languages still more a wish than a reality. Assessment, broadly conceived, should be used for evaluation and placement, matching skills to tasks, whenever transition occurs, for example, when a student enters college, enters or exits a program abroad, or applies for a job. Such assessment must measure on a clearly defined and agreed upon scale or set of standards. Good assessment instruments, as we have noted, can drive curriculum in positive ways. Most successful students have learned how to take tests and do reasonably well—the culture of our educational system rewards those who do well on our exams. The goal of future developers of assessment instruments should be to measure only those things worth learning.

One final note. Evaluation, assessment, curriculum are all complex issues, which benefit from the participation of all concerned, certainly students and instructors, but also administrators, employers, editors, parents. However, the design and construction of a good test, like that of good instructional materials, ultimately requires experienced, competent professionals. So let the work begin of moving from testing to assessment and from teaching to learning.

Works Cited

- American Council on the Teaching of Foreign Languages. *ACTFL Proficiency Guidelines*. Hastings-on-Hudson, NY: ACTFL, 1982.
- American Council on the Teaching of Foreign Languages. *ACTFL Russian Proficiency Guidelines*. Yonkers, NY: ACTFL, 1984.
- American Council on the Teaching of Foreign Languages. *Revised ACTFL Russian Proficiency Guidelines*. 1986. *Foreign Language Annals* XXI, 2 (April 1988), 177–97.
- Edelman, Ruth, Peter Merrill, and Jane Shuffelton. *Standards for Russian Language Learning*. Yonkers: ACTFL, 1998.
- Educational Testing Service. *Advanced Russian Listening/Reading Proficiency Test*. Princeton: ETS, 1986.
- Phillips, June, ed. *Standards for Foreign Language Learning: Preparing for the Twenty First Century*. Yonkers: ACTFL: 1996.
- Rifkin, Benjamin. "Report on the AATSEEL/ACTR Roundtable on Testing in Russian." *AATSEEL Newsletter* XLI.1 (February 1998), 16–17.

Russian Department
Middlebury College
Middlebury, VT 05753
tom.beyer@middlebury.edu