

## Examining the Validity of the 2010 Prototype AP Russian Exam through a College Comparability Study<sup>1</sup>

*Camelot Marshall*

### Introduction

Since its inception twelve years ago, the Prototype AP<sup>®</sup> Russian Language and Culture Examination has developed into an assessment instrument that has increasingly become the culminating focus and a hallmark of high school Russian language study in select schools across the United States.<sup>2</sup> Even more so, through the years of field-testing, piloting, and making the tests operational, the design, content, development, administration, and analyses of the exam have evolved into the model for American Councils' online assessments of language proficiency not only for Russian, but also for Flagship programs. These tests are already being developed in Arabic, Chinese, Persian, Russian, Swahili, and Yoruba.<sup>3</sup>

The Prototype AP<sup>®</sup> Russian Program provides high school students with advanced Russian-language instruction, comparable to that of a college-level language class. The course, instead of being curriculum-based, reflects intermediate level proficiency skills expected and described in the course description of the Prototype AP<sup>®</sup> Russian Course and Exam. The target proficiency level of the exam for the intermediate level was determined from analyses of language performance of students with four and six semesters of college Russian language study and no Russian language study prior to

---

<sup>1</sup> The author would like to extend a special thank-you to the Russian Language Specialists and Test Development Members, the American Councils Language Assessment Support System (ACLASS), and to Saodat I. Bazarova for a concise analysis of the data.

<sup>2</sup> Throughout its development and administrations, the Prototype AP<sup>®</sup> Russian Exam has held different names reflecting the various stages of existence. Since December 2006, the exam has been known as the Prototype AP<sup>®</sup> Russian Language and Culture Examination. Initial funding for the project was provided by the U.S. Government through International Research Study Title VI and the National Security Education Program (NSEP).

<sup>3</sup> For a detailed discussion of the Russian proficiency tests developed and administered by American Councils, see Saodat I. Bazarova, Maria D. Lekic, and Camelot Marshall, "The Online Proficiency-based Reading, Listening, and Integrated Writing External Assessment Program for Russian: a Report to the Field." *Russian Language Journal*, Vol. 59 (2009): 59-78.

starting in college.<sup>4</sup> Over the years of its administration, the number of test takers continues to grow, as well as the number of colleges and universities that grant placement and/or credit based on results on the Prototype AP<sup>®</sup> Russian Exam. In order to assess how well the Prototype AP<sup>®</sup> Russian Program and Exam prepare students to continue Russian language study at the college level and to assure that the Prototype AP<sup>®</sup> final grade placement suggestions meet the current performance outcomes of university-level Russian language programs, a short comparability study was conducted with college students in summer 2010. That study is the focus of this report.

### **Overview of College Board's AP<sup>®</sup> Foreign Language Comparability Studies**

Every five to seven years, the College Board conducts college-grade comparability studies to assure that the final AP grades earned by test takers reflect the standards and practices of colleges and universities and the expectations of the parallel course they offer. The methodology of the College Board's comparability study for the foreign language exams starts with the administration of a subtest of the exam, which includes the multiple-choice sections and the free response section, to students with at least two years of college language study.<sup>5</sup> For the newer language exams, such as Chinese or Japanese, students are expected to have completed approximately 300 hours of college-level foreign language study.<sup>6</sup> In addition to the administration of a subtest, the college professor is asked to provide a grade for the subtest taken by each student and a course grade for each student. Course grades are requested, because they serve as the criterion used for determining the cut scores on the exams to earn college credit. To motivate the students to take the subtest seriously, they were told that their performance on the exam will be part of a final course grade. Although the multiple-choice subtests were graded like full AP exams, they were not full-length exams; thus, the mean scores and composite scores were calculated from equipercentile scaling of the subtest scores to the scores of the AP examinees. The free-response section was graded

---

<sup>4</sup> Camelot Marshall, "The Online Prototype AP<sup>®</sup> Russian Examination: An Evaluation of Its Reliability and Validity." *Mnemosynon: Studies on Language and Culture in the Russophone World*, ed. Richard D. Brecht, Ljudmila A. Verbitskaja, Maria D. Lekić, and William P. Rivers (Moscow, Russia: Azbukovnik, 2009), 132-145.

<sup>5</sup> Deborah Lokai Bishof *et al.*, "Validating AP Modern Foreign Language Examinations Through College Comparability Studies," *Foreign Language Annals*, 37, 4, (2004): 616-622.

<sup>6</sup> David Baum, "Establishing AP Chinese Language and Culture Exam Validity Using College Professors' Grading Standards," 2007, [http://www.apcentral.org/apc/.../collegeCompStudy\\_chineseLang\\_07.pdf](http://www.apcentral.org/apc/.../collegeCompStudy_chineseLang_07.pdf).

by AP readers in the subject area, using the same scoring rubric as the AP examinees. The results on each section were combined to create a composite score, which was used in combination with course grades to determine cut scores for the different AP final grade levels 1-5.<sup>7</sup>

For the college comparability study, data were categorized into three groups: 1) all college students; 2) all AP students, and 3) a standard group of AP students. The standard group of AP students were those who responded “no” to both background questions on their foreign language exam: 1) “Have you lived or studied for one month or more in a country where the language of the exam you are now taking is spoken?; 2) Do you regularly speak or hear the language at home?”<sup>8</sup> Based on the results of the comparability studies, the grading standards are adjusted in order to maintain the performance levels expected of AP students. That is, when comparing the results of the college students with those of the high school AP students, the College Board expects that high school AP students’ composite mean scores on the exams for each AP final grade should be higher than the average course grade of college students in the corresponding college level.

While these comparability studies continue to validate AP examinations today, they remain subjective to the extent that they use final course grades as determined by grading rubrics, which vary from professor to professor. The compatibility study presented here looks at student performance on the multiple-choice sections of the Prototype AP<sup>®</sup> Russian Exam, which are proficiency-based, using the American Council on the Teaching of Foreign Languages (ACTFL) Proficiency Guidelines, generally accepted and applied as a measurement of functional language.

### **Test Design, Content, and Measurement**

The testing objective of the Prototype AP<sup>®</sup> Russian Language and Culture Exam is to provide a summative, proficiency-based measurement of high school students’ language proficiency in the four skills: reading, listening, writing, and speaking. Students taking the Prototype AP<sup>®</sup> Russian Exam have completed at least two to three years of high school Russian language study (or the equivalent) with the final year of study, typically the fourth year, an advanced-level course, comparable to that of a college Russian language course. Heritage speakers may present a different course of language study, depending on the extent to which

---

<sup>7</sup> Bischof *et al.*, “College Comparability Studies,” 618-619.

<sup>8</sup> See “How AP College Comparability Studies Affect AP Grades for Foreign Language Exams,” [www.apcentral.collegeboard.com/apc/public/courses/11622.html](http://www.apcentral.collegeboard.com/apc/public/courses/11622.html).

the language is used and/or taught at home or outside a regular school setting.

Test specifications reflect “The College Board World Languages Framework” and the proficiency guidelines developed by the ACTFL. Test items are based on the tasks, skill domains, and performance indicators published in *Standards for Foreign Language Learning in the 21<sup>st</sup> Century*.<sup>9</sup> The Prototype AP<sup>®</sup> Russian Examination consists of four subtests, all of which are aimed at the direct assessment of functional proficiency within interpersonal, interpretive, and presentational modes of communication. The key to the test design and content is that it is proficiency-based, so there is not a prescribed program of study or target language units to master. Exam questions range in difficulty from the Novice level to Advanced, with the majority of questions focusing on the intermediate level of proficiency. The intermediate level of proficiency was set as the target level for the AP exam, following analyses of student performance data in the American Councils or Teachers of Russian (ACTR) database. College students who began their Russian language study in college and who had two or three years of study performed at the Intermediate-Mid level of proficiency.<sup>10</sup> The content of the exam, like the content of the course, is not prescribed but rather assesses the student’s ability to perform in all communicative aspects of the language (reading, listening, writing, and speaking). The exam includes authentic stimulus materials that are of interest to high school students and that could be encountered in real-life situations.

The Prototype AP<sup>®</sup> Russian Exam is approximately three hours, and includes the online administration of three sections—reading comprehension, listening comprehension, and integrated written communication—and a separate oral proficiency interview (OPI), which is administered by telephone. The reading and listening comprehension sections are multiple-choice sections. All of the stimulus material is in Russian, while instructions and multiple-choice questions are in English. Test questions are posed in English, so as to assess students’ comprehension of the reading or listening stimulus materials, rather than assess the added task of understanding the comprehension question. Reading texts may be brief articles, notes, advertisements, or various other types of authentic Russian sources. Likewise, audio clips may be short conversations or interviews, announcements, and advertisements; all of which are self-contained messages. Each text or audio may be followed by one to three multiple-choice

---

<sup>9</sup> See the “Russian-Specific Guidelines,” *Standards for Foreign Language Learning: Preparing for the 21<sup>st</sup> Century*. National Standards in Foreign Language Education Project, Allen Press, Inc., 2006: 433-474. *The College Board’s World Languages Framework*, 2006, [www.collegeboard.com](http://www.collegeboard.com).

<sup>10</sup> Marshall, 2009.

items. Items are developed by experienced item writers and undergo a lengthy review process before they are finalized and ready for an exam form. Items are regularly reviewed for proficiency level, topic appropriateness, equity, and fairness, following *ETS Standards for Quality and Fairness, 2002*.

The reading and listening comprehension sections of the Prototype AP<sup>®</sup> Russian Exam are assembled, adhering to test development processes to ensure quality control and statistical reliability. The reading comprehension section consisted of 34 items, ranging in proficiency levels from Novice-Mid through Advanced, and the listening comprehension section consisted of 36 items, also ranging in proficiency levels from Novice-Mid through Advanced. Prior to appearing on a test form, all items are calibrated by piloting among college students studying Russian. Items selected for inclusion on an exam form must meet the selection criteria of item difficulty between 0.30 and 0.90 and point biserial (discrimination index) of no less than 0.25. Each Prototype AP<sup>®</sup> Russian Exam form is equated across administrations from year to year, using a set of common items, which represent the content and item characteristics of the earlier form. Common items with a previous test form make up at least 40% of the items in the reading and listening comprehension sections.

A development committee composed of Russian language specialists at the high school and college levels reviews the exam forms and approves them. Following the exam administration, the equating provides raw score-to-raw score conversions for the old and new forms. This procedure produces raw score equivalents on the new form to the anchor points set on the original form based on the standard setting study. American Councils assembled a panel of experts to determine the cut scores that mark the boundaries between proficiency levels for the reading comprehension and listening comprehension sections for the Prototype AP<sup>®</sup> Russian Exam.<sup>11</sup> Using the cut scores from the standard setting, the raw scores of the multiple-choice reading and listening comprehension exams can be placed on the ACTFL scale.

The integrated written communication, composed of a dialogue completion and integrated tasks, requires that test takers produce language within a meaningful, real-life context. For the dialogue completion, test takers supply missing lines of a dialogue, the topic of which is highly plausible in a real-life situation. For the integrated tasks, students respond in writing to reading and listening prompts to create a realistic scenario of incorporating all three language skills. The oral proficiency interview is administered by a certified ACTFL tester and is personalized to each test taker, building on the test

---

<sup>11</sup> Bazarova, Lekic, and Marshall, "External Assessment for Russian: Report to the Field."

takers' response to the opening question: Tell me about yourself (*Расскажите о себе*). Although ACTFL certified testers rate the students' spoken and written performance in the language, the proficiency ratings assigned should not be considered formal ACTFL ratings, because of the nature of the exam.

Composite scores are provided on the College Board's AP Grade scale of one to five, and since the total grade is linked to proficiency, sub-scores for each of the four language modalities are also reported in proficiency levels. The sub-score report provides a brief summary of the exam specifications and proficiency orientation of the exam, as well as a brief descriptive summary of the proficiency level achieved on each section of the exam. The Prototype AP<sup>®</sup> Russian Language and Culture Examination is designed to provide a set of measures of functional proficiency in Russian for use as a predictive assessment for continued language study, and as a placement tool by American colleges and universities for entering freshmen with prior study of Russian. In order to maintain the comparability between the AP final grade, section proficiency levels, and performance in a comparable college course, the reading and listening multiple-choice sections of the Prototype AP<sup>®</sup> Russian Exam were administered online to college students prior to their summer Russian language study.

For thirty-five years, American Councils for International Education: ACTR/ACCELS has administered and overseen summer, semester-long, and nine-month immersion programs in Russia. The database population in this study includes participants on summer study-abroad programs administered through American Councils and on an intensive stateside summer language program. The students represent a broad demographic of colleges and universities, both large and small, and public and private. Before taking the online exam, students answered a short background questionnaire focusing on Russian language study and use.

### **Student Population and General Descriptives**

The entire Prototype AP<sup>®</sup> Russian Exam was administered online to 266 high school students across the United States, after completing an advanced program of high school Russian language study. Within the high school population, 47% were male and 53% were female. A total of 52% reported they were non-heritage speakers of Russian, and 48% reported they were heritage speakers. Whereas self-reports of factual information are reliable, students' responses to the question of whether or not they were heritage speakers of Russian were double-checked with their teachers' reports of heritage status, and further checked with additional questions about language spoken in the home and parents' native

language. A total of 73% reported that they did not study Russian prior to high school, while 27% reported that they studied Russian prior to their high school Russian language study. The years of reported high school Russian language study ranged from zero to eight years. While high school Russian language programs typically offer four years of study, it is likely that for some of the reported years of language study, students combined their middle and high school years of Russian language study, or attended a school that did not differentiate between middle and high school language study. Within the high school population, 3% of the students were freshmen, 11% were sophomores, 33% were juniors, and 53% were seniors. It is likely that students taking the Prototype AP<sup>®</sup> Russian Exam who have had fewer than four years of high school language study do so because either they are heritage speakers, they started Russian language study before entering high school, or their high school offers fewer than four years of Russian language study.

Item analyses of the exam data show the Prototype AP<sup>®</sup> Russian Exam administered to the high school student to have good psychometric characteristics. Looking at the test taker data, the reading comprehension section shows an internal consistency reliability of .85, and the listening comprehension shows a reliability of .88. Reliability provides a level of statistical certainty that repeated administrations of the exam sections would yield similar results of a test-taker's abilities. The overall group mean score for reading comprehension ( $M = 25.73$ ,  $SD = 6.187$ ) was similar to the mean score for listening comprehension ( $M = 25.23$ ,  $SD = 7.676$ ). A review of graphical representation of the distribution of scores show a relatively normal distribution. The skewness and kurtosis values are within acceptable range. Looking at the means of just the non-heritage test takers, the reading comprehension mean is lower ( $M = 21.77$ ,  $SD = 5.420$ ), and the mean for the listening comprehension is slightly lower ( $M = 18.78$ ,  $SD = 5.762$ ). The differences in means, particularly that of the listening comprehension, may be attributed to the heritage students being more proficient in their listening skills.

For the comparability study, only the multiple-choice reading and listening comprehension sections of the Prototype AP<sup>®</sup> Russian Exam were administered to 133 college students at the beginning of their summer Russian language study program. A total of 83% of the college students were beginning a study abroad summer language program, while 17% were beginning a stateside summer language program. Among the college students, 42% were male and 58% were female. A total of 98% reported they were non-heritage speakers of Russian, and 2% reported they were heritage speakers of Russian. Within this

population, 83% reported that they did not study Russian prior to college, and 17% indicated that they studied Russian before college. The college students reported between zero and twelve semesters of Russian language study. The academic levels of the college students consisted of 1% freshmen, 20% sophomores, 32% juniors, 20% seniors, and 27% graduate students and other.

Item analyses of the reading and listening sections administered to the college students show good psychometric characteristics. The reading section has a reliability of .87, and the listening comprehension shows a reliability of .79. Among the college students, the reading comprehension mean ( $M = 25.91$ ,  $SD = 6.546$ ) is slightly higher than the mean of the listening comprehension section ( $M = 21.66$ ,  $SD = 5.762$ ). Tests showed a relatively normal distribution of scores. Removing graduate and other students from the analyses, as well as those students who reported they were heritage speakers, reveals little change in the reading mean ( $N = 95$ ,  $M = 25.13$ ,  $SD = 6.622$ ) or the listening mean ( $N = 95$ ,  $M = 21.09$ ,  $SD = 5.670$ ).

### **Comparison of Means**

The standard for the Prototype AP<sup>®</sup> Russian Exam was set based on measured proficiency levels of university-level students of Russian at different points in their language study, and the target proficiency range is the intermediate range. Entering freshmen who had taken the Prototype AP<sup>®</sup> Russian Exam would have their results aligned with the typical performance levels of university students at the third, fourth, and fifth semester levels of study. In order to check on whether or not the empirically set target levels for placement correspond to high school students' performance on the Prototype AP<sup>®</sup> Exam, a sample each of both high school and college level students was selected for a comparison of their means on the reading and listening comprehension sections.

High school students who were non-heritage speakers of Russian and who reported at least four years of high school Russian language study were selected for the comparison of means. The selected high school sample consisted of 108 students. Since 87% of the non-heritage high school students received final overall AP results with suggested placement, as determined by analyses of student language performance prior to study abroad into either second semester of first-year Russian or into second-year Russian (first or second semester), the selected college student sample for comparison consisted of non-heritage undergraduate students with either one or two years of Russian language study, and who began their Russian language study at the university level. The selected university sample consisted of 57 students, which is a



sufficient sample size for providing reliable statistical results; typically, 30 is accepted as providing reliable results.

Using an independent t-test, the means for the reading and listening comprehension sections were compared for the two groups. Frequency tables of the raw scores for both groups showed the data to be evenly distributed. Normality tests showed a relatively normal distribution, and skewness and kurtosis were within acceptable range. A comparison of the reading and listening comprehension means shows higher means for the undergraduates reading comprehension (M = 23.93, SD = 6.769) and listening comprehension (M = 20.07, SD = 5.480) with one or two years of college Russian than for the high school group's reading comprehension (M = 22.16, SD = 5.197) and listening comprehension (M = 18.91, SD = 4.179). Given the unequal sample sizes, Levene's test for homogeneity of variances was examined. The results of this analysis revealed that there was no difference between the groups on the two sections of the exam: Reading comprehension [ $t(-1.726) = 91.630, p < .088$ ] and for Listening comprehension [ $t(-1.522) = 163, p < .130$ ]. That is, the high school student performance on the reading and listening sections of the Prototype AP<sup>®</sup> Russian Exam is comparable to the college students' performance on the reading and listening comprehension sections, and confirms the target university levels for placement. While the comparison of means shows there is no difference between the groups, a look at the proficiency level distributions, based on the set cut scores, shows variation in intermediate levels of proficiency that are also comparable for the target placement levels into college Russian language programs (see Table 1 and Table 2).

**Table 1 High School Non-heritage Students (N = 108)**

Proficiency Level	Reading Comprehension	Listening Comprehension
NM	5%	0%
NH	27%	7%
IL	33%	84%
IM	21%	3%
IH	9%	6%
IH*	5%	0%

IH\* Signifies that students exhibit proficiency higher than the Intermediate High level, but the test specifications do not allow for a specific rating to be established.

**Table 2: Undergraduate Students (N = 57)**

Proficiency Level	Reading Comprehension	Listening Comprehension
NM	2%	5%
NH	21%	5%
IL	26%	54%
IM	7%	2%
IH	16%	30%
IH*	28%	4%

IH\* Signifies that students exhibit proficiency higher than the Intermediate High level, but the test specifications do not allow for a specific rating to be established.

### **Discussion**

A general comparison of the mean scores of all AP high school test takers and the college test takers shows comparable mean scores, which suggests that the groups perform similarly on the multiple-choice reading and listening comprehension sections. Under further review, however, the close mean scores are most likely due to the large number of heritage speakers in the high school group. When this sub-group is removed from the analyses, the reading comprehension mean score for the total high school group decreases, and the listening comprehension mean score decreases significantly more than the reading comprehension mean score. This is not surprising, in that listening comprehension skills of heritage speakers are stronger than those of the non-heritage speakers, and reading comprehension of the non-heritage speakers is closer to the reading comprehension of heritage speakers. Since there were only two heritage speakers in the college student population, the mean scores for the reading and listening comprehension sections show negligible change, when the heritage speakers are removed.

Looking more closely at student performance for placement target levels, the analyses demonstrate that while the college level test takers, who started their Russian language study in college and have at least one or two years of college level study, have a slightly higher mean score for the reading and listening comprehension sections than the non-heritage high school population. The t-test, accounting for unequal sample size, indicates that there is no difference in the two groups' performance. These findings purport to validate the Prototype AP® Russian Exam as a measurement instrument for reading and listening proficiency, as well as for target placement into a college-level Russian

language course.

This comparability study is slightly different than the methodology used for the College Board's comparability studies, primarily in that the college students were given the full-length reading and listening comprehension sections of the Prototype AP<sup>®</sup> Russian Exam, and the results on the sections were based solely on students' performance on the multiple-choice sections. In the College Board study, college students' results on a shortened test form are estimated to scores on a full AP Foreign Language Exam. In order to obtain a composite score, the estimated scores are combined with final course grades as assigned by the participating professor or foreign language instructor. The added factor of the final course grade is based on students' performance in the course, but it is also subjective: in order to determine the final course grade, the instructor created a grading rubric in line with course expectations. With this added variable into the comparability study, the results are not fully on student performance in the language, but also on the language instructors' expectations for student performance in the course. The comparability study of the Prototype AP<sup>®</sup> Russian Exam presented here remains objective, in that it is based on student performance on the exam form and objective in the expectations for language performance, which are based on the exam's target proficiency levels.

Turning to students' proficiency levels for the reading and listening comprehension sections, the distribution for the high school group and the college-level group are similar. The high school students primarily perform at intermediate-low and -mid levels of proficiency, the target levels of the exam. The college students also perform at intermediate levels of proficiency, and more college students perform at the intermediate-high or above levels than the high school students. This can be accounted for by the inclusion of students with three or more years of college-level Russian language study. The proficiency levels assigned to student performance on the exam sections result from the application of cut scores, as decided during a standard setting exercise to the students' raw scores on the exam. Because proficiency ratings do not easily lend to statistical analyses, student raw scores were used for the comparison of means.

Tangential to the application of the results on the Prototype AP<sup>®</sup> Russian Exam for entering freshmen's college credit and/or placement is the application of the exam design, construct, administration, and scoring to Russian language proficiency tests for other purposes. Currently, some study-abroad programs require that students' language ability be assessed before and after formal study in the target culture. Thus, the model of the Prototype AP<sup>®</sup> Russian Exam can be applied to such tests as needed by the study abroad programs. Likewise, with

increased expectations for evidence of language learning, college and university departments seek their own entry and exit tests for their students. Although each Russian department has its own core curriculum expectations for language learning, a Russian language test that is based on proficiency, albeit perhaps not targeting all of the programs' expectations and outcomes, offers a uniform platform for measuring and reporting student language proficiency. With a consistency in measurement and reporting of language proficiency, those in the field of Russian language teaching will have a better means to assess and understand language performance, which in turn will inform curriculum development and language training.

### **Conclusion**

This study attests to the reliability of the online Prototype AP<sup>®</sup> Russian Exam as an instrument to measure language proficiency in terms of ACTFL Proficiency ratings for Reading and Listening, in addition to providing a final AP Grade on the College Board's 1-5 scale. Student results can be used for college placement and/or credit, depending on college or university AP grade policies. The exam content and administration to high school and university-level students showed relatively high reliabilities for both multiple-choice reading and listening sections of the exam. This is significant in that the same exam form, given in similar conditions and to a comparable student population, would yield similar results.

This study further confirms that high school students who have taken the Prototype AP<sup>®</sup> Russian Exam demonstrate comparable mean scores on the reading and listening comprehension sections to mean scores of the university-level students. These results provide evidence to the Russian language field regarding the reliability, validity, and performance outcomes of the exam, and it should be accepted with confidence that the Prototype AP<sup>®</sup> Russian Exam indeed captures students' language performance. As such, placement decisions of entering freshmen can therefore be better informed as to what the results and score reports mean, and how they match up with the expected level of college student performance and the appropriate level of language instruction at the receiving institution. The increased use of the exam for placement purposes as colleges and universities gain increasing institutional experience with the AP results, especially given the additional information of proficiency subscores, will hopefully lead to an increasing institutional use of the exams for college credit, and in some cases, even help freshmen gain sophomore standing.

## References

- Baum, David, "Establishing AP Chinese Language and Culture Exam Validity Using College Professors' Grading Standards," 2007, [http://www.apcentral.org/apc/.../collegeCompStudy\\_chineseLang\\_07.pdf](http://www.apcentral.org/apc/.../collegeCompStudy_chineseLang_07.pdf).
- Bazarova, Saodat I., Maria D. Lekic, and Camelot Marshall. "The Online Proficiency-based Reading, Listening, and Integrated Writing External Assessment Program for Russian: A Report to the Field." *Russian Language Journal* (2009) 59: 59-78.
- Bischof, Deborah Lokai, David Baum, Jodi M. Casabianca, Rick Morgan, Kathleen A. Rabiteau, and Krishna Tateneni, "Validating AP Modern Foreign Language Examinations Through College Comparability Studies." *Foreign Language Annals*, 37, 4 (2004): 616-622.
- "How AP College Comparability Studies Affect AP Grades for Foreign Language Exams," [www.apcentral.collegeboard.com/apc/public/courses/11622.html](http://www.apcentral.collegeboard.com/apc/public/courses/11622.html).
- Marshall, Camelot. 2009. The online Prototype AP® Russian Examination: An evaluation of its reliability and validity. *Mnemosynon: Studies on language and culture in the Russophone world*, eds. Richard D. Brecht, Ljudmila A. Verbitskaja, Maria D. Lekic, and William P. Rivers, 132-145. Moscow, Russia: Azbukovnik.
- "Russian-Specific Guidelines," *Standards for Foreign Language Learning: Preparing for the 21<sup>st</sup> Century*. National Standards in Foreign Language Education Project, 2006, Allen Press, Inc., 433-474.
- Standards for Educational and Psychological Testing*, 1999. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. American Educational Research Association, Washington, D.C.
- The College Board's World Languages Framework*, 2006, [www.collegeboard.com](http://www.collegeboard.com).