Second Language Research & Practice October 2020, Volume 1, Issue 1
pp. 25–64

RESEARCH ARTICLE

The proficiency profile of language students: Implications for programs

Paula Winke, Michigan State University
Xiaowan Zhang, Michigan State University
Fernando Rubio, The University of Utah
Susan Gass, Michigan State University
Dan Soneson, University of Minnesota
Jane Hacking, The University of Utah

Abstract

In this article we provide an overall picture of the level of language proficiency attained by undergraduate students learning six languages in postsecondary language programs. Specifically, we address the curricular and pedagogical implications of the proficiency assessments on the programs that emerged from this large-scale study, which took place over a three-year span at Michigan State University, the University of Utah, and the University of Minnesota. We collected speaking, reading, and listening proficiency test data from students enrolled at all levels of instruction in Arabic, Chinese, French, Portuguese, Russian and Spanish. We present the results and divide them by students' heritage status, the students' number of years of kindergarten through twelfth grade learning of the target language, and the students' postsecondary curricular level (first, second, third, or fourth) of instruction. We also investigate junior and senior upper-division students' proficiency attainments by their language major or minor status. We build on issues and questions raised by Teschner in 1991 and the Modern Language Association in 2007.

Keywords: *proficiency, heritage language learners, K-12 language learning, postsecondary language programs, majors*

APA Citation: Winke, P., Zhang, X., Rubio, F., Gass, S., Sonenson, D., & Hacking, J. (2020). The proficiency profiles of language students: Implications for programs. *Second Language Research & Practice*, *1*(1), 25–64. http://hdl.handle.net/10125/69840

Introduction

Large-scale assessments of proficiency attainment in U.S. language programs are rare. John Carroll's influential 1967 study was the first to report the proficiency levels of foreign language majors at U.S. colleges and universities. He gave the Modern Language Association (MLA) Foreign Language Proficiency Tests for Teachers and Advanced Students to 2,782 seniors majoring in French, German, Italian, Russian, or Spanish at 203 institutions. He found that majors varied widely in the levels of language skills they attained, with some obtaining superior performance (they tested at the upper limits of the proficiency tests), and some performing "abysmally poor[ly]" (Carroll, 1967, p. 134) on the assessments. The median scores from Carroll's 1967 report are in Table 1.[1] Note that Carroll did not report the Italian results because only 28 tested in Italian. In addition to the median score reports, Carroll reported that the following factors affected median scores significantly and positively: professional preparation (i.e., those seeking teacher certification to teach the language), time abroad, an elementary or high-school (pre-college) start in learning the language, and parents' use of the language at home. He opined that the Russian learners' lower median

attainments could be because Russian is more challenging, or because the Russian learners did not start learning in elementary or high school, as individuals studying the other languages did.

Table 1

The Median Learning Attainments of Language Majors (adapted from Carroll, 1967, pp. 141-145).

		Median Score on 1967 FSI Scale and Modern ACTFL Scale Conversion							
Language	Number	Listening		Speaking		Reading		Writing	
French	1270	almost 2+	AM	2	AL	3	S	almost 3	AH
German	413	3	S	2+	AH	3	S	almost 3	AH
Russian	105	2	AL	1+	IH	almost 2+	AM	2+	AH
Spanish	968	2+	AH	2+	AH	almost 3+	S	3	S

Note. The Foreign Service Institute (FSI) scale is from 0, no proficiency, to 5, educated native speaker proficiency (see Hertzog, n.d.). The ACTFL 2012 scale is from Novice Low (NL) to Distinguished (D) (see ACTFL, 2012).

Carroll's study, although significant and widely cited, may have limited application to today's proficiency attainments in college-level programs because (a) Carroll's focus was on 1965 language majors, who may not correspond to today's undergraduate majors and the 21st century's typical college-level language learner; and (b) he limited his reporting to four languages (French, German, Russian, and Spanish).

It was also not clear in Carroll's study if the learners had significant amounts of pre-college language learning, perhaps more than students today may have. Rifkin (2005) followed up with a study that investigated the proficiency attainment of 352 college-level Russian language learners participating in a summer Russian immersion program. The learners were at various levels of instruction in the program, but by using the pre- and post-program test data, Rifkin demonstrated that only upper-level learners with more than 600 hours of prior instruction were able to reach advanced levels of proficiency. Rifkin suggested that the number of class hours in college-level programs matter, so much so that "the college or university intent on helping students achieve that goal [of advanced proficiency] would need to offer a Russian program with, on average, five hours of instruction per week, 30 weeks per year, for four years" (2005, p. 12), even though none of the 80 Russian programs he surveyed offered that many hours of instruction.

To illustrate potential further differences between the current proficiency attainments of majors contemporarily and that of the majors in Carroll's 1967 study, Winke, et al. (2019), using a subset of the data used for the current analysis, followed up on Carroll's work and asked a similar question: What is the status of foreign language proficiency amongst university undergraduate majors? They investigated the proficiency levels of French, Russian, and Spanish majors in upper division courses (3rd and 4th year) at Michigan State University. They found that the solo language major was rare: Most students minored instead of majored in foreign languages, and when they did major in a language, they paired it with another major and/or one or more minors. Like Carroll's work, Winke et al.'s (2019) was limited in terms of the number of foreign languages investigated, but the study demonstrated that new, large-scale research on language learners and their proficiency attainments is needed, in part because the major of today is qualitatively different from the major in the 1960s.

Since Carroll's work, newer but smaller-scale studies have reported on college-level language learning, with most focusing on the first two years of college-level language programs. Researchers investigating German acquisition (Goertler et al., 2016; Norris & Pfeiffer, 2003; Tschirner, 1992; Tschirner & Heilenman, 1998) generally found that after four semesters (two years) of German, the average college student attained an Intermediate Low to Intermediate Mid speaking proficiency level. Freed (1987), Magnan (1986), and Schmitt (2014) investigated French learners' oral proficiency levels in lower-level

college programs and found similar results as the German studies. That is, after two years of college-level French, students on average obtained an Intermediate Low to Mid level of proficiency. Hernández (2006), Sanders (2005), and Schmitt (2014) investigated Spanish learners' proficiency attainment. They too found that after three to four semesters of college-level Spanish, the average student obtained Intermediate Low to Mid in Spanish speaking proficiency. Meanwhile, Thompson (1996) found Russian students on average obtained Novice High to Intermediate Low oral proficiency levels after two years of college Russian. However, in these studies, proficiency attainment was not qualified by the factors Carroll (1967) originally identified as important influencers: when the learner started language study (in college, or in elementary, middle, or high school); whether the learner spent time abroad; or whether the language was spoken in the home when the learner was growing up. Such information is needed to explain proficiency attainment. As a reviewer of this manuscript pointed out, there is some comparison of apples to oranges as well, in part because Carroll researched only those who were seniors and majoring in the language, whereas participation in the other studies was not restricted to majors, and measured proficiency at lower curricular points in four-year, college-level programming.

In the current article, we take these issues into consideration, and we take a broad view of proficiency in U.S.-based language programs by investigating proficiency levels across the full curriculum, somewhat as Rifkin (2005) did before, who investigated the proficiencies of all Russian learners in an intensive language program, but here on a larger scale; that is with more students, more than one language, and across a four-year program. We include six languages (Arabic, Chinese, French, Portuguese, Russian, and Spanish) in our analysis. The results are from a large assessment project conducted at Michigan State University, the University of Utah, and the University of Minnesota. Funding for The Language Flagship Proficiency Initiative was provided by the National Security Education Program, a part of the Department of Defense.[2] The Proficiency Initiative is intended to:

> integrate Flagship proficiency assessment practices and processes within existing high-quality academic language programs. The purpose of this initiative is to introduce the Flagship proficiency assessment process to established academic foreign language programs to measure teaching and learning, and to evaluate the impact of such testing practices on teaching and learning. (Request for Proposals, The Language Proficiency Flagship Initiative, p. 1)

The Language Flagship Initiative was established nearly two decades ago with the express goal of creating programs that would move students to advanced language proficiency in a select number of critical languages. Initially, the program served the graduate student population, but in 2006 shifted its focus to undergraduate education with the aim of creating global professionals with high levels of proficiency. As explained on their website:

> The Language Flagship is a national initiative to change the way Americans learn languages through a groundbreaking approach to language education…The Language Flagship graduates students who will take their place among the next generation of global professionals, commanding a superior level of proficiency in one of ten languages critical to U.S. national security and economic competitiveness. (The Language Flagship, n.d.)

Assessment is, of course, an important part of any language program and is a way of: (a) understanding curricular needs and (b) determining successes and shortcomings of language programs in meeting their proficiency-based goals.

It was against this backdrop that the Language Flagship Program issued a call for institutions of higher education to partner with the Defense Language and National Security Education Office (DLNSEO) "to create a viable process to assess proficiency learning in high quality, well-established academic language programs and to document the impact of introducing rigorous proficiency assessment on language pedagogy practice and outcomes" (Request for Proposals, The Language Proficiency Flagship Initiative, p. 3).[3] The data set on which we report in this article departs from previous work on assessment conducted within the context of the Language Flagship Program (e.g., Davidson, 2010; Davidson, et al., 2017) in that

student proficiency levels reported are those from language programs without specific Flagship goals and, therefore, more closely reflect the proficiency levels attained in the majority of U.S.-based language programs.

The goals of this initiative were to (a) evaluate proficiency levels currently being attained, (b) document the impact of introducing proficiency assessment on language pedagogy practice and outcomes, and (c) share best practices in proficiency-based curricula with other institutions. Our specific mandates for this project were the following:

- institutionalize proficiency assessment practices that align student placement, course goals;
- document ways in which assessment results are integrated into foreign language programs (curriculum and teaching); and
- share practices with others in the foreign language community.

The current project began in the fall of 2014 and data collection continued through the spring semester of 2017. With over 20,000 tests administered over three years, this project constitutes the first major effort to ascertain the level of proficiency of post-secondary students since Carroll (1967). As noted above, unlike Carroll's study, which focused on language majors, this project took a broader view, testing the language proficiency of both majors and non-majors who were enrolled in elementary (first and second year) and intermediate to advanced level (third and fourth year) language classes. We also collected data from students enrolled in some less commonly taught languages for which only limited large-scale assessment data have previously been available. In addition to proficiency data, we collected background information from test takers in order to establish potential correlations between proficiency ratings and other factors such as attitudes, motivation, pre-university exposure to the language, and heritage status. In this article, we focus only on the latter two.

The analyses reported in this article come from one semester of cross-sectional data collected in spring 2017. We selected this semester because it reflected the largest single-semester cohort across all three institutions. Furthermore, collaboration on the background questionnaire resulted in a revised instrument in 2017 used by all three institutions in the final year of testing.[4]

Research questions for this article are:

1. What is the current status of proficiency within academic language programs at the post-secondary level?
2. What roles do factors such as pre-university experience and language heritage background play in contributing to language proficiency?
3. Do language majors outperform language minors and students neither majoring nor minoring in the language?

Methodology

In the spring of 2017, at all three universities, intact language classes were brought to computer labs by their instructors to take part in this study. The number of classes that could test at a given time was limited by the number of labs available for testing because the assessments were computer-based. Thus, classes that participated were selected based on class time as well as the individual instructors' willingness to have their classes participate. In the labs, the students first took a language background questionnaire and an oral proficiency test. The listening and reading assessments were taken by the intact classes on another day, or the individual students were asked to come into the language laboratories on the campuses outside of their regularly scheduled language classes to take the listening and reading assessments on their own, with proctors present.

Participants

For the current study, we include data from 5,241 scored proficiency tests, with 2,036 being scored tests of speaking, 1,641 of reading, and 1,564 of listening. In total, 2,121 students (learners of Arabic = 146, Chinese = 226, French = 396, Portuguese = 71, Russian = 173, Spanish = 1,109) produced these 5,241 scored assessments. At the time of testing, the students who took these scored tests were in the second, fourth, sixth, or eighth semester of their respective foreign language programs. The numbers of scored tests taken for this study and the genders of the individuals who took them are listed in Table 2 and results reflect the combined total across all three institutions. Because individuals were not required to take all three tests, the numbers of listening, speaking, and reading tests within a language are not equal. Because individual classes mostly took the speaking tests first, and students were sometimes asked to take the listening and reading tests as volunteers on a later date (in the proctored, walk-in language laboratories of the respective universities) on their own time, the number of students who took the reading and listening tests is lower than the number of speaking test takers.

Table 2

Number of Scored Assessments by Language, Skill, and Gender

Skill and Gender	Language					
	Arabic	Chinese	French	Portuguese	Russian	Spanish
Listening	136	145	345	68	139	731
Females	73	69	260	36	66	480
Males	63	76	85	32	73	251
Other	0	0	0	0	0	0
Speaking	139	215	385	71	169	1054
Females	79	102	288	37	75	721
Males	62	113	97	34	94	333
Other	1	0	0	0	0	0
Reading	130	148	329	67	143	824
Females	74	74	244	37	66	551
Males	55	74	85	30	77	273
Other	1	0	0	0	0	0

A number of students tested but did not receive viable ratings from their assessments. We excluded those non-scored assessments from this study, and these numbers are not included in Table 2. There were three reasons for this type of outcome: (a) the student took a test form that was too difficult, and the student was given a rating of below-range (BR) from the testing company; (b) the student took a test form that was too easy, and the student was given a rating of above-range (AR) from the testing company; or (c) the sample provided by the student was un-ratable for any one of a variety of reasons (e.g., no response, technology failure), and the student was given a rating of un-ratable (UR). In Table 3, for transparency purposes, we report the numbers of tests given but not scored by skill and language.

Table 3

Number of Non-Scored (Non-Used) Assessments by Language, Skill, and Non-Score Type

Skill and Non-score Type	Language					
	Arabic	Chinese	French	Portuguese	Russian	Spanish
Listening	2	0	10	1	15	51
AR	0	0	0	0	0	0
BR	0	0	10	1	15	51
UR	2	0	0	0	0	0
Speaking	0	0	0	0	3	0
AR	0	0	0	0	2	0
BR	0	0	0	0	0	0
UR	0	0	0	0	1	0
Reading	12	0	5	0	8	3
AR	0	0	0	0	0	0
BR	12	0	5	0	8	3
UR	0	0	0	0	0	0

Materials

There are two sources of data covered in this article: actual proficiency scores and background data.

Proficiency in speaking, reading, and listening was assessed using on-demand, computer-delivered tests administered by Language Testing International (https://www.languagetesting.com/), which are based on the American Council on the Teaching of Foreign Language (ACTFL) Proficiency Guidelines (ACTFL, 2012). In particular, we used the Oral Proficiency Interview [by] computer (OPIc), the Reading Proficiency Test (RPT) and the Listening Proficiency Test (LPT).[5] Whereas the OPIc is rated by official ACTFL-certified raters, the other two are automatically scored multiple-choice tests, and scores are generated immediately once the student has completed the test. In the OPIc, students responded to questions by a virtual partner. The speaking test results were emailed to the students and to the universities approximately two weeks after the speaking tests were taken. Ratings for each test can be given on the Interagency Language Roundtable (ILR) scale (Herzog, n.d.) or the ACTFL (2012) scale, based on the test purchaser's discretion. At the time of testing, two universities had the test scores reported out on the ACTFL scale, and one had the test scores reported out on the ILR scale. For this article, we converted all ILR scores to the ACTFL scale, using the conversion table similar to the one in Swender (2012). The ACTFL levels have four broad categories (Novice, Intermediate, Advanced, Superior), with the first three further divided into three sublevels (Low, Mid, High), resulting in 10 categories ranging from Novice Low to Superior (see ACTFL, 2012).

The second source of data comes from our internally developed questionnaire. The three universities worked closely to develop a single questionnaire that addressed, among other things, heritage status and K–12 language experience, the two foci of this article. The questionnaire had 28 items that covered such issues as previous language exposure (at home, at school), cultural exposure (immersion, living/studying abroad), current language practices, and motivation for learning the language.

Heritage learners were identified as those who answered "yes" to all the following questions:

1. Was (the target language) spoken in your home when you were growing up?
2. Did you yourself speak (the target language) at home growing up?
3. Do you have family members who speak the language?

With regard to K–12 experience, participants were asked to indicate whether they had studied the language or taken classes in the language in school prior to enrolling in university language classes. If they answered in the affirmative, they were asked to check a box for each year in school in which they took a class in the target language, from kindergarten through twelfth grade. The number of checkmarks was tallied to determine the extent of formal study of the language prior to university study.

In 2017 on the survey, at one university test takers were asked to indicate if they were majoring or minoring in the language being learned. At the other two universities, the researchers obtained each test taker's major and minor information from the registrar's office.

Results

In this results section, we provide descriptions of the data we collected in two ways: tabular and visually through plots. We provide proficiency level attainment by language and skill and by student division in three ways. First, we describe these data in light of the students' heritage status. Second, we describe them based on the students' K–12 (pre-college) language study. And third, we describe upper-division, junior, and senior students' proficiency level attainment by language, skill, and their major/minor status.

The Proficiency Profiles of Language Students by Language, Division, Skill, and Heritage Status

In Tables 4, 5, and 6, we list the average proficiency attainments language learners achieved by language, division, skill, and their heritage status. Note that under "division," Lower 1 is the first year in a four year language program, (Lower division, year 1), and Lower 2 is the second year in a four year program (Lower division, year 2), with "lower division" denoting the first two years of a four year program. "Upper" is the upper division: Upper 1 is the third year in a four-year program (Upper division, year 3 in the program), and Upper 2 is the fourth year in a four-year program (Upper division, year 4 in the program). Twelve of the Russian courses were designated as equally Upper 1 and Upper 2 courses, primarily due to lower enrollments in the Russian program, meaning that the upper division courses were for both 3rd and 4th year Russian students. Thus, some of the Russian courses do not have a 1 or 2 behind their Upper division distinction. Portuguese was not taught beyond the third year; thus, it does not have a fourth-year course distinction. We would like to note that students may take multiple courses at a time, or they may take a course based on the timing and subject level rather than based on its course division and level. Thus, course division/level is a gross categorization for learners, and one that may have more blurred lines in the upper division than in the lower division. Students were allowed to skip questions on the questionnaire; thus, heritage status is marked as NA (not applicable) if the student skipped (did not respond to) the questions that we used as heritage status indicators. In other words, NAs in the tables in this article indicate missing data. Table 4 has the speaking scores by heritage status, language, and division, whereas Table 5 has reading, and Table 6 has listening. Note that for Tables 4, 5, and 6, *n* = number, M = mean ACTFL rating with Novice Low being 1 and Superior being 10 (the tests do not test at the Distinguished level), SD = standard deviation, Min = minimum score, Max = maximum score. We also report the median scores to match Carroll's (1967) reporting method.

Table 4

Speaking Scores by Heritage Status, Language, and Division

Language	Division	Heritage Status	*n*	M	SD	Median	Min	Max
Arabic speaking	Lower1	Non-heritage	76	3.95	1.06	4	1	7
		Heritage	2	5.00	0.00	5	5	5
		NA	1	7.00	NA	7	7	7
	Lower2	Non-heritage	38	4.82	1.14	5	2	7
		Heritage	3	6.33	1.15	7	5	7
	Upper1	Non-heritage	18	5.22	1.59	5	2	8
		NA	1	5.00	NA	5	5	5
	Upper2	Non-heritage	3	4.33	3.51	4	1	8
Chinese speaking	Lower1	Non-heritage	68	2.68	1.15	2	1	6
		Heritage	4	3.75	0.96	3.5	3	5
	Lower2	Non-heritage	52	3.67	1.08	4	2	8
		Heritage	4	5.50	1.00	5	5	7
		NA	1	2.00	NA	2	2	2
	Upper1	Non-heritage	50	4.42	1.85	4	1	10
		Heritage	8	6.00	1.93	6	4	9
	Upper2	Non-heritage	23	5.61	1.78	5	3	9
		Heritage	4	6.00	2.45	6.5	3	8
		NA	1	2.00	NA	2	2	2
French speaking	Lower1	Non-heritage	74	2.78	0.97	3	1	5
		Heritage	1	4.00	NA	4	4	4
	Lower2	Non-heritage	154	4.68	0.89	5	2	7
	Upper1	Non-heritage	105	5.22	1.04	5	3	7
		Heritage	1	8.00	NA	8	8	8
		NA	1	4.00	NA	4	4	4
	Upper2	Non-heritage	49	6.33	1.68	6	1	9
Portuguese speaking	Lower1	Non-heritage	5	4.00	1.22	4	2	5
	Lower2	Non-heritage	46	5.02	0.83	5	2	7
		Heritage	1	5.00	NA	5	5	5
		NA	2	4.50	3.54	4.5	2	7
	Upper1	Non-heritage	16	5.75	0.86	5.5	5	7
		NA	1	7.00	NA	7	7	7
Russian speaking	Lower1	Non-heritage	73	3.36	1.06	4	1	5
		Heritage	4	5.00	0.82	5	4	6
		NA	1	4.00	NA	4	4	4
	Lower2	Non-heritage	54	4.11	1.11	4	2	7
		Heritage	4	7.75	1.50	7	7	10
		NA	1	4.00	NA	4	4	4

Language	Division	Heritage Status	*n*	M	SD	Median	Min	Max
	Upper	Non-heritage	12	4.33	0.65	4	3	5
	Upper1	Non-heritage	19	5.79	1.44	6	4	8
	Upper2	Non-heritage	1	8.00	NA	8	8	8
Spanish speaking	Lower1	Non-heritage	175	2.92	0.98	3	1	5
		NA	1	2.00	NA	2	2	2
	Lower2	Non-heritage	341	4.28	0.97	4	1	7
		Heritage	3	7.00	0.00	7	7	7
		NA	2	4.00	1.41	4	3	5
	Upper1	Non-heritage	397	4.87	0.88	5	3	8
		Heritage	14	6.79	1.25	7	4	9
		NA	4	5.50	1.00	5	5	7
	Upper2	Non-heritage	106	5.57	1.04	6	3	9
		Heritage	11	7.36	0.67	7	6	8

Table 5

Reading Scores by Heritage Status, Language, and Division

Language	Division	Heritage Status	*n*	M	SD	Median	Min	Max
Arabic reading	Lower1	Non-heritage	75	2.00	1.23	2	1	5
		Heritage	2	2.50	0.71	2.5	2	3
		NA	1	5.00	NA	5	5	5
	Lower2	Non-heritage	28	3.68	1.28	4	1	6
		Heritage	3	4.67	0.58	5	4	5
		NA	2	3.00	2.83	3	1	5
	Upper1	Non-heritage	15	4.07	1.58	4	1	6
		NA	1	7.00	NA	7	7	7
	Upper2	Non-heritage	3	3.33	3.21	2	1	7
Chinese reading	Lower1	Non-heritage	53	1.36	0.90	1	1	4
		Heritage	1	4.00	NA	4	4	4
		NA	1	1.00	NA	1	1	1
	Lower2	Non-heritage	40	1.98	1.19	2	1	5
		Heritage	4	5.00	3.74	4.5	1	10
		NA	2	1.50	0.71	1.5	1	2
	Upper1	Non-heritage	24	4.04	2.40	4	1	7
		Heritage	5	4.80	1.48	5	3	7
		NA	1	5.00	NA	5	5	5
	Upper2	Non-heritage	12	5.58	1.56	6	3	7
		Heritage	3	5.67	1.53	6	4	7
		NA	2	5.50	2.12	5.5	4	7
French	Lower1	Non-heritage	74	2.62	0.99	3	1	5

Language	Division	Heritage Status	*n*	M	SD	Median	Min	Max
reading		Heritage	1	4.00	NA	4	4	4
		NA	2	2.00	1.41	2	1	3
	Lower2	Non-heritage	148	4.23	1.28	4	1	7
	Upper1	Non-heritage	68	6.72	0.75	7	5	8
		NA	2	6.50	0.71	6.5	6	7
	Upper2	Non-heritage	33	7.15	0.83	7	6	9
		NA	1	7.00	NA	7	7	7
Portuguese reading	Lower1	Non-heritage	5	5.40	1.52	5	4	7
	Lower2	Non-heritage	43	5.58	1.38	6	1	7
		Heritage	1	5.00	NA	5	5	5
		NA	2	6.00	2.83	6	4	8
	Upper1	Non-heritage	15	6.33	0.98	7	5	8
		NA	1	7.00	NA	7	7	7
Russian reading	Lower1	Non-heritage	65	2.57	1.22	2	1	5
		Heritage	4	4.75	0.50	5	4	5
		NA	2	2.00	0.00	2	2	2
	Lower2	Non-heritage	37	3.70	1.56	4	1	7
		Heritage	4	6.25	0.96	6.5	5	7
		NA	1	4.00	NA	4	4	4
	Upper	Non-heritage	8	4.62	1.30	5	3	7
	Upper1	Non-heritage	19	6.00	1.53	7	3	8
		Heritage	2	7.00	1.41	7	6	8
	Upper2	Non-heritage	1	7.00	NA	7	7	7
Spanish reading	Lower1	Non-heritage	168	2.76	1.33	3	1	5
		NA	1	3.00	NA	3	3	3
	Lower2	Non-heritage	334	4.30	1.48	5	1	7
		Heritage	3	4.67	0.58	5	4	5
		NA	5	3.80	1.79	4	1	6
	Upper1	Non-heritage	222	6.97	0.99	7	5	10
		Heritage	9	6.89	0.78	7	6	8
		NA	5	6.60	0.89	6	6	8
	Upper2	Non-heritage	67	7.24	1.28	7	1	10
		Heritage	8	8.00	0.76	8	7	9
		NA	2	8.50	2.12	8.5	7	10

Table 6

Listening Scores by Heritage Status, Language, and Division

Language	Division	Heritage Status	*n*	M	SD	Median	Min	Max
Arabic listening	Lower1	Non-heritage	70	2.20	1.23	2	1	5
		Heritage	2	4.50	0.71	4.5	4	5
		NA	1	6.00	NA	6	6	6
	Lower2	Non-heritage	37	3.11	1.37	3	1	6
		Heritage	3	4.33	0.58	4	4	5
		NA	2	2.50	2.12	2.5	1	4
	Upper1	Non-heritage	17	2.65	1.90	1	1	6
		NA	1	5.00	NA	5	5	5
	Upper2	Non-heritage	3	2.33	2.31	1	1	5
Chinese listening	Lower1	Non-heritage	47	1.36	0.92	1	1	5
		Heritage	2	4.00	0.00	4	4	4
		NA	1	1.00	NA	1	1	1
	Lower2	Non-heritage	36	2.00	1.49	1	1	7
		Heritage	3	6.00	3.46	4	4	10
		NA	2	1.00	0.00	1	1	1
	Upper1	Non-heritage	28	3.79	2.20	4	1	7
		Heritage	6	5.33	1.37	5	4	7
		NA	1	4.00	NA	4	4	4
	Upper2	Non-heritage	15	4.93	1.28	5	3	7
		Heritage	2	7.00	0.00	7	7	7
		NA	2	5.50	2.12	5.5	4	7
French listening	Lower1	Non-heritage	73	2.59	1.23	3	1	5
		Heritage	1	3.00	NA	3	3	3
		NA	2	1.00	0.00	1	1	1
	Lower2	Non-heritage	143	4.06	1.06	4	1	7
	Upper1	Non-heritage	82	6.12	0.81	6	5	7
		Heritage	1	7.00	NA	7	7	7
		NA	3	6.00	1.00	6	5	7
	Upper2	Non-heritage	39	6.46	0.85	7	5	8
		NA	1	6.00	NA	6	6	6
Portuguese	Lower1	Non-heritage	5	4.60	3.29	7	1	7

Language	Division	Heritage Status	*n*	M	SD	Median	Min	Max
listening	Lower2	Non-heritage	43	5.14	1.44	5	1	7
		Heritage	1	6.00	NA	6	6	6
		NA	2	5.50	2.12	5.5	4	7
	Upper1	Non-heritage	16	6.06	1.34	7	3	7
		NA	1	7.00	NA	7	7	7
Russian listening	Lower1	Non-heritage	69	1.70	1.02	1	1	4
		Heritage	4	4.25	0.50	4	4	5
		NA	2	1.00	0.00	1	1	1
	Lower2	Non-heritage	36	2.53	1.34	3	1	5
		Heritage	4	5.75	0.96	5.5	5	7
		NA	1	4.00	NA	4	4	4
	Upper	Non-heritage	5	3.40	0.55	3	3	4
	Upper1	Non-heritage	15	4.67	1.11	4	3	7
		Heritage	2	7.00	1.41	7	6	8
	Upper2	Non-heritage	1	7.00	NA	7	7	7
Spanish listening	Lower1	Non-heritage	171	1.70	0.93	1	1	5
		NA	1	2.00	NA	2	2	2
	Lower2	Non-heritage	312	3.14	1.31	3	1	7
		Heritage	3	5.00	2.00	5	3	7
		NA	5	2.60	1.14	3	1	4
	Upper1	Non-heritage	153	5.58	0.82	5	5	8
		Heritage	13	6.77	0.60	7	5	7
		NA	3	6.00	1.73	5	5	8
	Upper2	Non-heritage	57	6.04	1.13	6	1	8
		Heritage	11	6.73	0.79	7	6	8
		NA	2	7.50	0.71	7.5	7	8

In Figures 1, 2, 3, 4, 5 and 6, the data that contributed to the calculations in Tables 4, 5, and 6 are displayed as violin plots so that readers can visually inspect the shape of the data at the individual level. Figures 1 through 6 each have data from one of the six languages (Arabic in Figure 1, Chinese in Figure 2, etc.). Each figure has three panels, with visualizations of the speaking, reading, and listening scores charted within each panel, respectively. Within each panel, the language program's division offerings each have their own violin plots. For example, in Figure 1, in the first panel of Arabic speaking scores by division and heritage status, each individual dot is an Arabic speaking-test (OPIc test) taker. The person's dot is positioned along the y axis in coordination with the ACTFL rating that the person received. If the person is a heritage speaker as

identified in this article (with a code of 1 in the data spreadsheet), their dot is blue. Non-heritage test takers (coded as a 0) are red. Black dots are students who did not respond to the questions on heritage status (those whose code on heritage is not available, that is, NA). The dots are jiggered (slightly moved apart by our request within the statistical system *R*) so that the individual dots can be better seen, although in dense data areas, dots may appear on top of other dots. The density of the data at a given ACTFL level is indicated by the width of the violin plot. Thus, in Figure 1, in the first Arabic speaking panel, within the Lower 1 courses (first-year Arabic courses), the most frequent ACTFL rating obtained was Intermediate Low. This is also indicated in Table 4, where it is listed that the 72 non-heritage Lower 1 Arabic students obtained an average score of 3.95 (approximately Intermediate Low) on the ACTFL OPIc speaking test. It is evident from Figure 1 that heritage Arabic learners have an advantage in proficiency assessment, especially in the skills of speaking and listening, as one would expect based on the literature concerning heritage language learners (for reviews see Benmamoun, et al., 2013 and Swender, et al., 2014).

The heritage advantage is prominent in all six languages, but most visually prominent in Spanish (see Figure 6), where the number of heritage learners is rather large and where their advantage is relatively clear.

Figure 1

Arabic Speaking, Reading, and Listening Scores by Division and Heritage Status

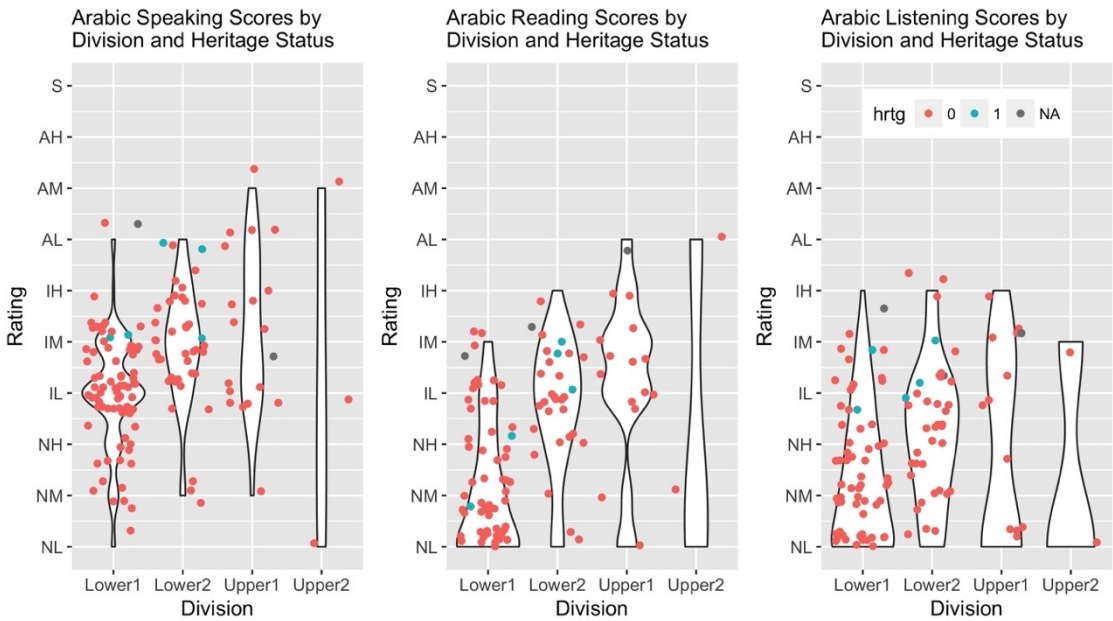


Figure 2

Chinese Speaking, Reading, and Listening Scores by Division and Heritage Status

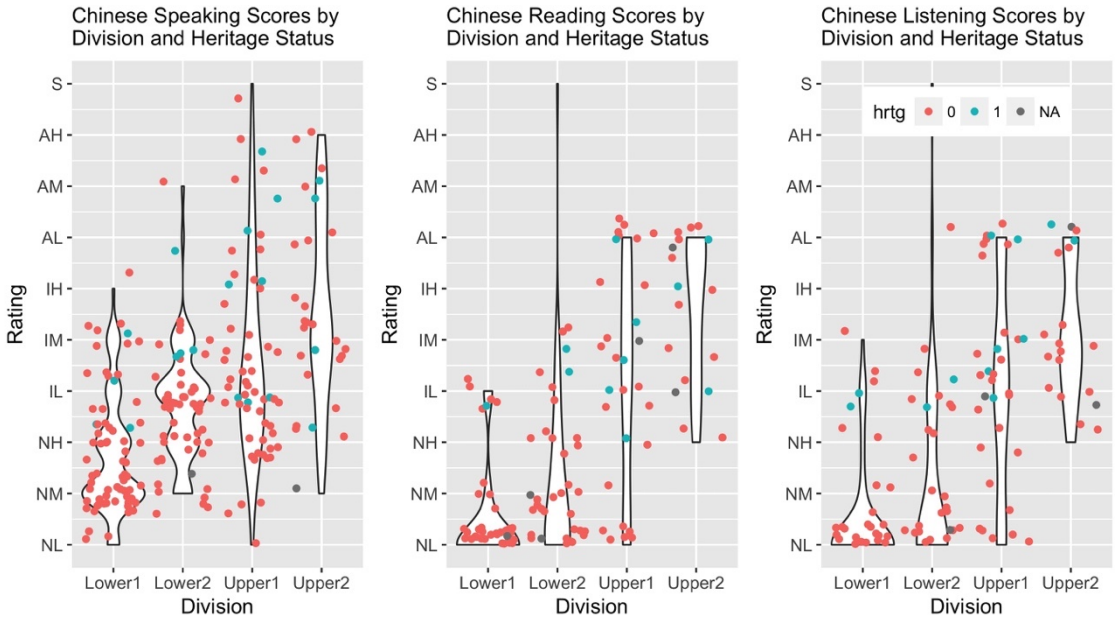


Figure 3

French Speaking, Reading, and Listening Scores by Division and Heritage Status

Figure 4

Portuguese Speaking, Reading, and Listening Scores by Division and Heritage Status

Figure 5

Russian Speaking, Reading, and Listening Scores by Division and Heritage Status

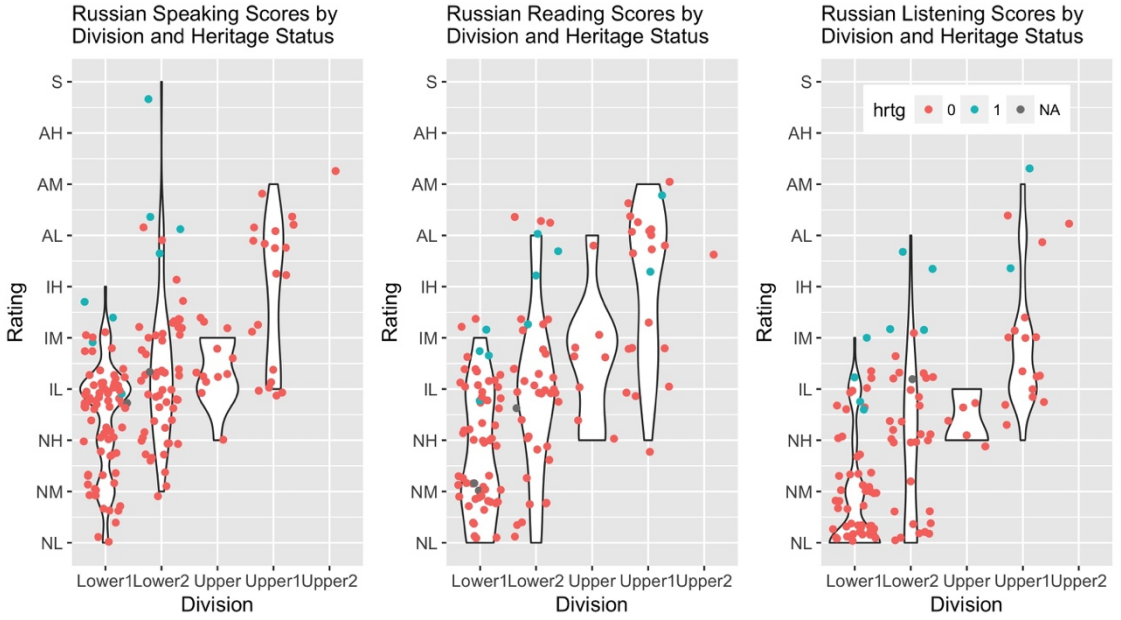
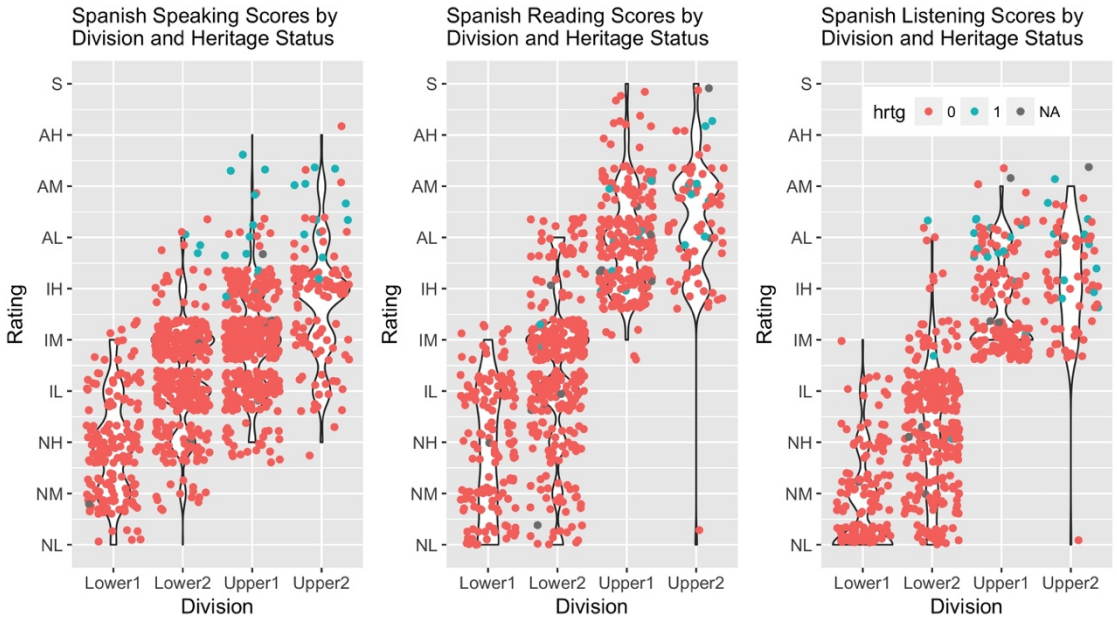


Figure 6

Spanish Speaking, Reading, and Listening Scores by Division and Heritage Status



The Proficiency Profiles of Language Students by Language, Division, Skill, and K–12 Experience

In Tables 7, 8, and 9, we list the average proficiency attainments language learners achieved by language, division, skill, and their experience learning the language in kindergarten through 12th grade. In Figures 7, 8, 9, 10, 11, and 12, we plot the calculations from Tables 7, 8, and 9 by language (one figure for each language). For clarity, we report K–12 experience as none, one year, two years, three years, and four years or more because the number of students who had more than four years was negligible, and because color coding (heat mapping) more than five categories of K–12 experience was visually difficult, especially considering space limitations. More K–12 experience in Figures 7 through 12 is represented by lighter shading of blue. In Figures 7 through 12, a person with missing data on K–12 experience is represented by a grey dot.

There are many different ways one could look at coding and investigating K–12 experience in relation to proficiency attainment, especially given the theories in second language acquisition (SLA) that purport that earlier-in-life learning may have advantages over later-in-life learning (DeKeyser, et al., 2010; Hartshorne, et al., 2018). We do not propose that the way we visually plot K–12 experience in this article is the only way to do it; we simply suggest this is one way researchers can do it to clearly see the effects of K–12 learning on college-level proficiency outcomes, and how that experience differentially contributes to the various foreign language programs at colleges and universities in the United States. For example, upper-division French and Spanish courses at the three universities that participated in this study are almost uniquely comprised of students with two or more years of K–12 French or Spanish learning experience, respectively.

Table 7

Speaking Scores by Language, Division, and K–12 Experience

Language	Division	K12	*n*	M	SD	Median	Min	Max
Arabic speaking	Lower1	None	72	3.89	1.06	4	1	7
		1 Year	2	5.00	0.00	5	5	5
		>=4 Years	4	5.00	0.00	5	5	5
		NA	1	7.00	NA	7	7	7
	Lower2	None	33	5.03	1.19	5	2	7
		1 Year	2	4.50	0.71	4.5	4	5
		2 Years	1	4.00	NA	4	4	4
		>=4 Years	5	4.60	1.52	5	2	6
	Upper1	None	15	5.27	1.44	5	4	8
		1 Year	2	4.50	3.54	4.5	2	7
		>=4 Years	1	6.00	NA	6	6	6
		NA	1	5.00	NA	5	5	5
	Upper2	None	2	2.50	2.12	2.5	1	4
		3 Years	1	8.00	NA	8	8	8
Chinese speaking	Lower1	None	41	2.54	1.10	2	1	5
		1 Year	8	3.00	1.41	2.5	2	6
		2 Years	5	2.80	1.30	2	2	5
		3 Years	7	2.71	1.11	2	2	5
		>=4 Years	11	3.27	1.19	3	2	5
	Lower2	None	20	4.10	1.55	4	2	8
		1 Year	6	4.17	0.75	4	3	5
		2 Years	4	3.25	0.96	3.5	2	4
		3 Years	7	3.43	0.98	3	2	5
		>=4 Years	19	3.63	0.83	4	2	5
		NA	1	2.00	NA	2	2	2
	Upper1	None	16	4.69	2.47	4	1	9
		1 Year	5	4.60	1.14	5	3	6
		2 Years	6	5.17	1.72	5	3	7
		3 Years	6	4.00	0.63	4	3	5
		>=4 Years	25	4.64	1.98	4	3	10
	Upper2	None	7	6.14	2.04	6	3	9
		1 Year	2	4.50	0.71	4.5	4	5
		2 Years	2	5.00	0.00	5	5	5
		3 Years	4	5.75	0.96	5.5	5	7
		>=4 Years	12	5.67	2.23	5	3	9
		NA	1	2.00	NA	2	2	2

Language	Division	K12	*n*	M	SD	Median	Min	Max
French speaking	Lower1	None	31	2.29	0.78	2	1	4
		1 Year	6	2.83	0.75	3	2	4
		2 Years	10	2.90	0.99	2.5	2	4
		3 Years	12	3.08	0.67	3	2	4
		>=4 Years	16	3.50	1.10	3.5	1	5
	Lower2	None	35	4.29	0.99	4	2	6
		1 Year	4	4.25	0.96	4.5	3	5
		2 Years	15	4.13	0.83	4	3	5
		3 Years	17	4.47	0.72	5	3	5
		>=4 Years	83	5.01	0.76	5	3	7
	Upper1	None	6	6.33	1.21	6.5	5	8
		1 Year	9	4.22	1.20	4	3	6
		2 Years	3	5.33	0.58	5	5	6
		3 Years	5	5.20	0.84	5	4	6
		>=4 Years	84	5.26	1.00	5	3	7
	Upper2	None	5	7.00	1.41	7	5	9
		1 Year	4	5.00	1.41	5.5	3	6
		2 Years	5	6.40	1.52	7	4	8
		3 Years	1	9.00	NA	9	9	9
		>=4 Years	34	6.29	1.70	6	1	9
Portuguese speaking	Lower1	None	5	4.00	1.22	4	2	5
	Lower2	None	47	5.02	0.82	5	2	7
		NA	2	4.50	3.54	4.5	2	7
	Upper1	None	13	5.77	0.93	5	5	7
		1 Year	3	5.67	0.58	6	5	6
		NA	1	7.00	NA	7	7	7
Russian speaking	Lower1	None	72	3.43	1.12	4	1	6
		1 Year	3	3.33	1.15	4	2	4
		2 Years	2	4.00	0.00	4	4	4
		NA	1	4.00	NA	4	4	4
	Lower2	None	50	4.24	1.46	4	2	10
		1 Year	1	5.00	NA	5	5	5
		3 Years	5	5.00	1.58	5	3	7
		>=4 Years	2	5.50	0.71	5.5	5	6
		NA	1	4.00	NA	4	4	4
	Upper	None	12	4.33	0.65	4	3	5
	Upper1	None	18	5.89	1.41	6.5	4	8
		1 Year	1	4.00	NA	4	4	4
	Upper2	None	1	8.00	NA	8	8	8

Language	Division	K12	*n*	M	SD	Median	Min	Max
Spanish speaking	Lower1	None	66	3.08	1.06	3	1	5
		1 Year	26	2.73	1.12	3	1	5
		2 Years	29	2.62	0.78	3	1	4
		3 Years	19	3.00	0.94	3	1	5
		>=4 Years	35	2.97	0.86	3	1	5
		NA	1	2.00	NA	2	2	2
	Lower2	None	20	3.75	1.21	4	2	5
		1 Year	11	3.45	1.44	4	1	6
		2 Years	30	4.00	1.31	4	2	7
		3 Years	36	4.25	1.08	4	2	7
		>=4 Years	248	4.43	0.86	4	2	7
		NA	1	3.00	NA	3	3	3
	Upper1	None	12	5.33	1.44	5.5	3	8
		1 Year	13	5.46	0.88	5	4	7
		2 Years	11	4.73	0.91	4	4	6
		3 Years	18	4.61	1.04	5	3	7
		>=4 Years	357	4.93	0.93	5	3	9
		NA	4	5.50	1.00	5	5	7
	Upper2	None	4	6.50	1.29	6.5	5	8
		1 Year	2	5.50	0.71	5.5	5	6
		2 Years	4	6.25	1.50	6	5	8
		3 Years	4	6.25	0.96	6.5	5	7
		>=4 Years	103	5.67	1.13	6	3	9

Table 8

Reading Scores by Language, Division, and K–12 Experience

Language	Division	K12	*n*	M	SD	Median	Min	Max
Arabic reading	Lower1	None	71	1.96	1.18	2	1	5
		1 Years	2	1.50	0.71	1.5	1	2
		>=4 Years	4	3.25	1.71	3.5	1	5
		NA	1	5.00	NA	5	5	5
	Lower2	None	24	3.71	1.37	4	1	6
		1 Years	1	3.00	NA	3	3	3
		2 Years	1	4.00	NA	4	4	4
		>=4 Years	5	4.20	0.84	4	3	5
		NA	2	3.00	2.83	3	1	5
	Upper1	None	12	4.00	1.54	4	1	6
		1 Years	2	4.00	2.83	4	2	6
		>=4 Years	1	5.00	NA	5	5	5

Language	Division	K12	*n*	M	SD	Median	Min	Max
		NA	1	7.00	NA	7	7	7
	Upper2	None	2	1.50	0.71	1.5	1	2
		3 Years	1	7.00	NA	7	7	7
Chinese reading	Lower1	None	30	1.40	0.93	1	1	4
		1 Years	6	1.00	0.00	1	1	1
		2 Years	5	2.20	1.64	1	1	4
		3 Years	4	1.75	1.50	1	1	4
		>=4 Years	9	1.11	0.33	1	1	2
		NA	1	1.00	NA	1	1	1
	Lower2	None	18	2.72	2.30	2	1	10
		1 Years	5	2.60	1.67	3	1	5
		2 Years	5	1.20	0.45	1	1	2
		3 Years	5	1.80	1.30	1	1	4
		>=4 Years	11	2.00	1.00	2	1	4
		NA	2	1.50	0.71	1.5	1	2
	Upper1	None	9	4.67	2.18	5	1	7
		1 Years	2	5.00	2.83	5	3	7
		2 Years	1	7.00	NA	7	7	7
		3 Years	4	2.75	2.06	2.5	1	5
		>=4 Years	13	3.92	2.33	4	1	7
		NA	1	5.00	NA	5	5	5
	Upper2	None	5	6.20	1.30	7	4	7
		2 Years	1	5.00	NA	5	5	5
		3 Years	3	4.67	2.08	4	3	7
		>=4 Years	6	5.67	1.51	6	3	7
		NA	2	5.50	2.12	5.5	4	7
French reading	Lower1	None	31	2.55	1.06	2	1	4
		1 Years	6	2.33	0.52	2	2	3
		2 Years	11	2.27	1.10	2	1	4
		3 Years	12	2.83	0.94	3	1	4
		>=4 Years	15	3.07	0.88	3	2	5
		NA	2	2.00	1.41	2	1	3
	Lower2	None	32	3.78	1.48	4	1	7
		1 Years	4	4.25	1.26	4	3	6
		2 Years	16	3.81	0.75	4	2	5
		3 Years	17	3.94	1.14	4	2	6
		>=4 Years	79	4.56	1.24	5	1	7
	Upper1	None	4	6.75	0.50	7	6	7
		1 Years	6	6.17	0.41	6	6	7

Language	Division	K12	*n*	M	SD	Median	Min	Max
		2 Years	1	5.00	NA	5	5	5
		3 Years	3	7.00	0.00	7	7	7
		>=4 Years	54	6.80	0.76	7	5	8
		NA	2	6.50	0.71	6.5	6	7
	Upper2	None	5	7.00	0.71	7	6	8
		1 Years	2	8.50	0.71	8.5	8	9
		2 Years	4	7.75	0.50	8	7	8
		3 Years	1	7.00	NA	7	7	7
		>=4 Years	21	6.95	0.81	7	6	8
		NA	1	7.00	NA	7	7	7
Portuguese reading	Lower1	None	5	5.40	1.52	5	4	7
	Lower2	None	44	5.57	1.37	5.5	1	7
		NA	2	6.00	2.83	6	4	8
	Upper1	None	12	6.33	0.99	6.5	5	8
		1 Years	3	6.33	1.15	7	5	7
		NA	1	7.00	NA	7	7	7
Russian reading	Lower1	None	64	2.66	1.30	2	1	5
		1 Years	3	2.67	1.53	3	1	4
		2 Years	2	4.00	0.00	4	4	4
		NA	2	2.00	0.00	2	2	2
	Lower2	None	36	3.92	1.78	4	1	7
		1 Years	1	5.00	NA	5	5	5
		3 Years	2	4.00	0.00	4	4	4
		>=4 Years	2	4.00	1.41	4	3	5
		NA	1	4.00	NA	4	4	4
	Upper	None	8	4.62	1.30	5	3	7
	Upper1	None	20	6.15	1.53	7	3	8
		1 Years	1	5.00	NA	5	5	5
	Upper2	None	1	7.00	NA	7	7	7
Spanish reading	Lower1	None	65	2.86	1.39	3	1	5
		1 Years	20	2.85	1.35	3	1	5
		2 Years	28	2.29	1.33	2	1	5
		3 Years	18	2.67	1.19	2.5	1	5
		>=4 Years	37	2.95	1.25	3	1	5
		NA	1	3.00	NA	3	3	3
	Lower2	None	19	4.05	1.93	4	1	7
		1 Years	13	3.46	1.76	4	1	5
		2 Years	28	3.61	1.62	4	1	6
		3 Years	31	4.16	1.13	4	1	7

Language	Division	K12	*n*	M	SD	Median	Min	Max
		>=4 Years	247	4.46	1.41	5	1	7
		NA	4	3.75	2.06	4	1	6
	Upper1	None	7	6.71	1.11	7	5	8
		1 Years	6	6.67	0.82	6.5	6	8
		2 Years	5	7.00	1.00	7	6	8
		3 Years	7	6.71	0.76	7	6	8
		>=4 Years	206	6.99	0.99	7	5	10
		NA	5	6.60	0.89	6	6	8
	Upper2	None	3	8.67	1.15	8	8	10
		2 Years	3	7.67	0.58	8	7	8
		3 Years	2	8.50	0.71	8.5	8	9
		>=4 Years	67	7.21	1.25	7	1	10
		NA	2	8.50	2.12	8.5	7	10

Table 9

Listening Scores by Language, Division, and K–12 Experience

Language	Division	K12	*n*	M	SD	Median	Min	Max
Arabic listening	Lower1	None	66	2.12	1.22	2	1	5
		1 Year	2	3.50	0.71	3.50	3	4
		>=4 Years	4	4.00	0.82	4	3	5
		NA	1	6.00	NA	6	6	6
	Lower2	None	33	3.24	1.35	3	1	6
		1 Year	2	2.00	1.41	2	1	3
		>=4 Years	5	3.40	1.52	4	1	5
		NA	2	2.50	2.12	2.50	1	4
	Upper1	None	14	2.43	1.83	1	1	6
		1 Year	2	3.00	2.83	3	1	5
		>=4 Years	1	5.00	NA	5	5	5
		NA	1	5.00	NA	5	5	5
	Upper2	None	2	1.00	0.00	1	1	1
		3 Years	1	5.00	NA	5	5	5
Chinese listening	Lower1	None	28	1.36	0.87	1	1	4
		1 Year	7	1.43	1.13	1	1	4
		2 Years	4	2.00	2.00	1	1	5
		3 Years	2	2.50	2.12	2.50	1	4
		>=4 Years	8	1.38	0.74	1	1	3
		NA	1	1.00	NA	1	1	1
	Lower2	None	15	2.73	2.71	1	1	10
		1 Year	4	3.00	1.41	2.50	2	5

Language	Division	K12	*n*	M	SD	Median	Min	Max
		2 Years	4	2.50	1.73	2.50	1	4
		3 Years	5	2.00	1.41	1	1	4
		>=4 Years	11	1.55	0.93	1	1	4
		NA	2	1.00	0.00	1	1	1
	Upper1	None	8	4.50	2.51	4.50	1	7
		1 Year	4	3.75	2.50	3.50	1	7
		2 Years	4	5.00	1.63	5	3	7
		3 Years	4	2.75	2.06	2.50	1	5
		>=4 Years	14	4.00	2.08	4	1	7
		NA	1	4.00	NA	4	4	4
	Upper2	None	5	5.80	1.10	5	5	7
		1 Year	1	4.00	NA	4	4	4
		2 Years	2	4.00	1.41	3	3	5
		3 Years	2	5.50	2.12	4	4	7
		>=4 Years	7	5.14	1.46	3	3	7
		NA	2	5.50	2.12	4	4	7
French listening	Lower1	None	31	2.32	1.05	2	1	4
		1 Year	6	2.00	0.89	2	1	3
		2 Years	11	2.36	0.92	2	1	4
		3 Years	12	2.83	1.27	3	1	5
		>=4 Years	14	3.43	1.55	4	1	5
		NA	2	1.00	0.00	1	1	1
	Lower2	None	29	3.62	1.27	4	1	7
		1 Year	4	3.75	0.50	4	3	4
		2 Years	15	3.87	0.83	4	2	5
		3 Years	16	3.69	1.30	4	1	5
		>=4 Years	79	4.35	0.89	4	1	7
	Upper1	None	5	6.40	0.89	7	5	7
		1 Year	6	6.00	0.63	6	5	7
		2 Years	1	5.00	NA	5	5	5
		3 Years	3	6.67	0.58	7	6	7
		>=4 Years	69	6.10	0.83	6	5	7
		NA	2	6.50	0.71	6.5	6	7
	Upper2	None	4	6.25	0.96	6.5	5	7
		1 Year	2	7.00	0.00	7	7	7
		2 Years	4	7.00	0.00	7	7	7
		3 Years	1	7.00	NA	7	7	7
		>=4 Years	28	6.36	0.91	6.5	5	8
		NA	1	6.00	NA	6	6	6

Language	Division	K12	*n*	M	SD	Median	Min	Max
Portuguese listening	Lower1	None	5	4.60	3.29	7	1	7
	Lower2	None	44	5.16	1.43	5	1	7
		NA	2	5.50	2.12	5.5	4	7
	Upper1	None	13	6.16	1.14	7	4	7
		1 Year	3	5.67	2.31	7	3	7
		NA	1	7.00	NA	7	7	7
Russian listening	Lower1	None	68	1.82	1.17	1	1	5
		1 Year	3	1.67	1.15	1	1	3
		2 Years	2	2.50	0.71	2.5	2.5	3
		NA	2	1.00	0.00	1	1	1
	Lower2	None	34	2.82	1.64	1	3	7
		1 Year	1	3.00	NA	3	3	3
		3 Years	4	2.75	2.06	1	2.5	5
		>=4 Years	1	4.00	NA	4	4	4
		NA	1	4.00	NA	4	4	4
	Upper	None	5	3.40	0.55	3	3	4
	Upper1	None	16	4.94	1.39	4.5	3	8
		1 Year	1	5.00	NA	5	5	5
	Upper 2	None	1	7.00	NA	7	7	7
Spanish listening	Lower1	None	65	1.77	1.01	1	1	5
		1 Year	22	1.64	0.85	1	1	4
		2 Years	29	1.38	0.78	1	1	4
		3 Years	18	1.83	0.99	2	1	4
		>=4 Years	37	1.81	0.88	2	1	4
		NA	1	2.00	NA	2	2	2
	Lower2	None	18	2.39	1.42	2.5	1	5
		1 Year	11	2.45	1.57	2	1	5
		2 Years	23	2.78	1.48	3	1	7
		3 Years	26	3.27	1.37	3	1	7
		>=4 Years	238	3.28	1.26	3	1	7
		NA	4	2.25	0.96	7	5	7
	Upper1	None	7	6.43	0.98	7	5	7
		1 Year	6	5.67	1.03	5	5	7
		2 Years	1	5.00	NA	5	5	5
		3 Years	5	5.40	0.89	5	5	7
		>=4 Years	147	5.65	0.84	5	5	8
	Upper2	None	3	7.33	0.58	7	7	8
		1 Year	1	5.00	NA	5	5	5
		2 Years	2	7.50	0.71	7.5	7	8

Language	Division	K12	*n*	M	SD	Median	Min	Max
		3 Years	4	6.00	0.82	6	5	7
		>=4 Years	58	6.07	1.11	6	1	8
		NA	2	7.50	0.71	7.5	7	8

Figure 7

Arabic Speaking, Reading, and Listening Scores by Division and K–12 Experience

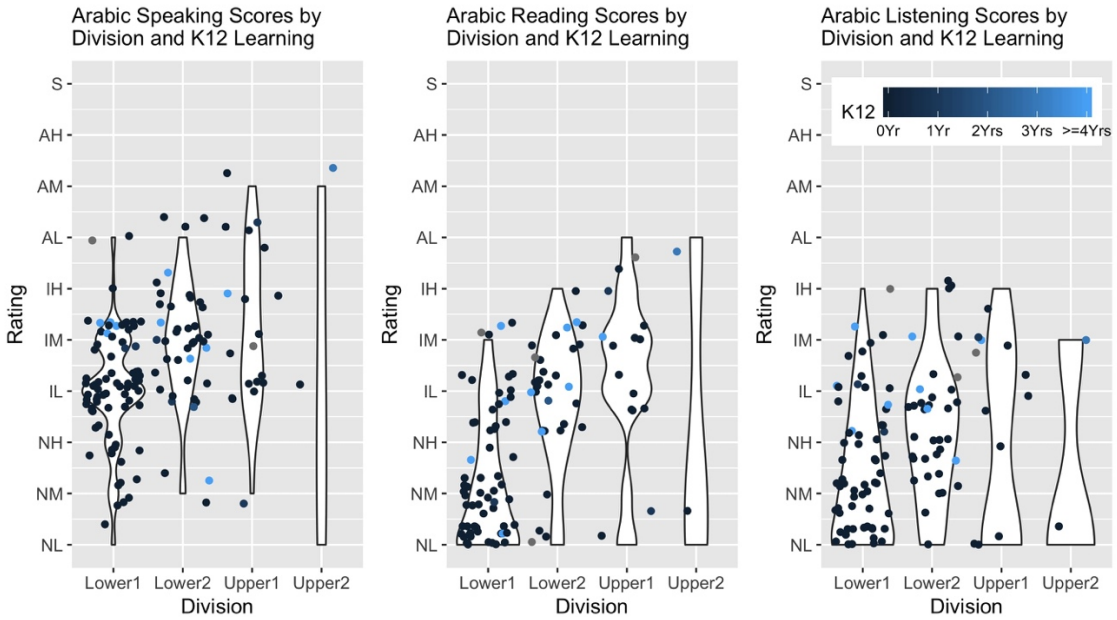


Figure 8

Chinese Speaking, Reading, and Listening Scores by Division and K–12 Experience

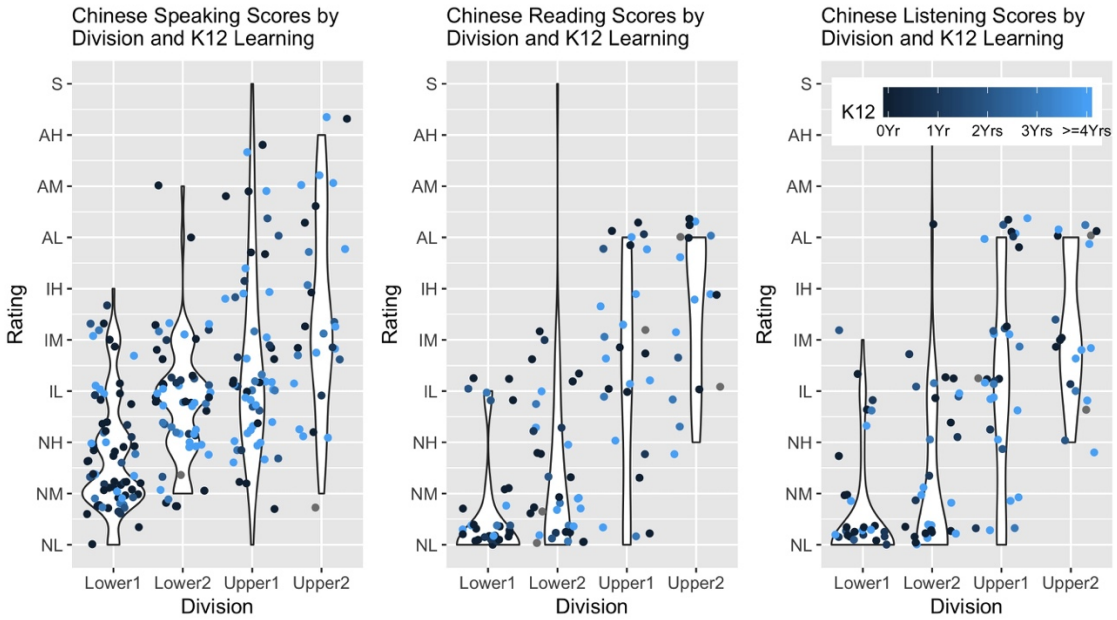


Figure 9

French Speaking, Reading, and Listening Scores by Division and K–12 Experience

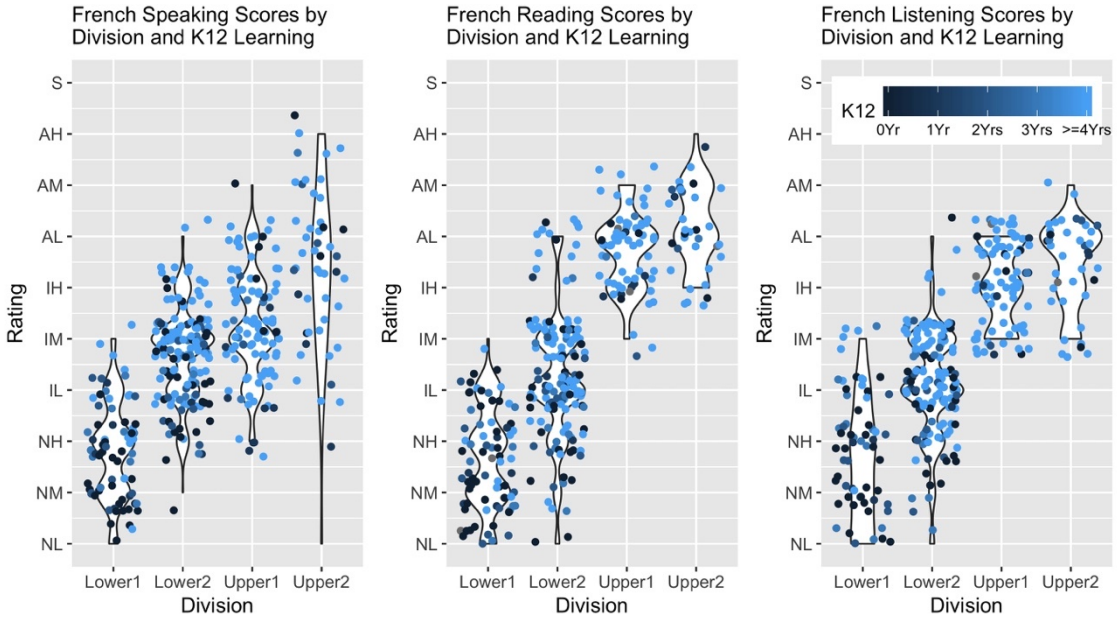


Figure 10

Portuguese Speaking, Reading, and Listening Scores by Division and K–12 Experience

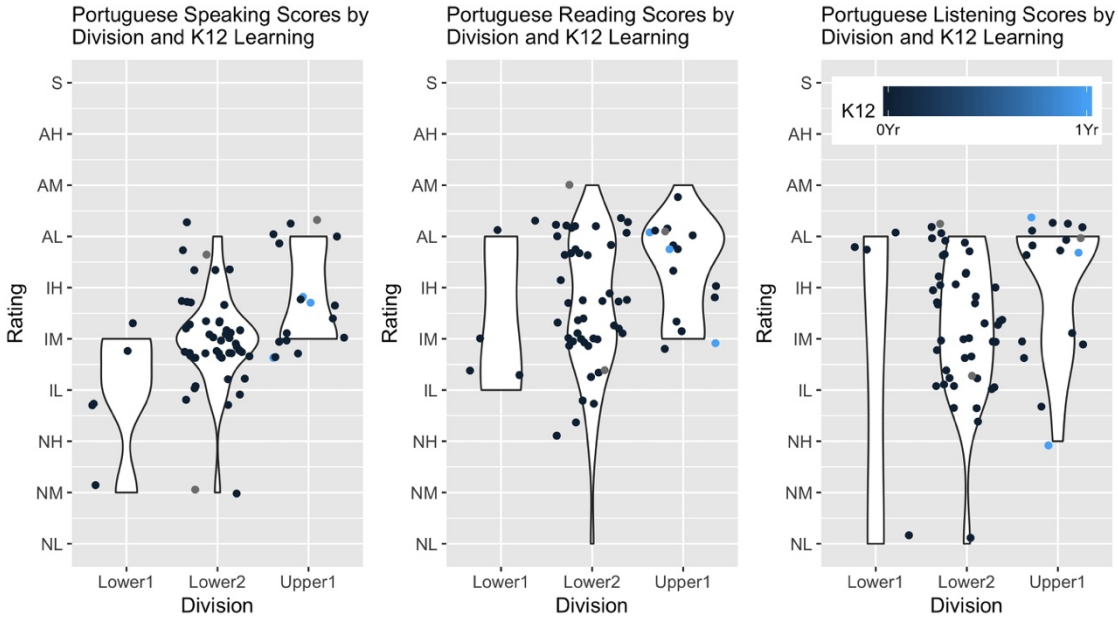


Figure 11

Russian Speaking, Reading, and Listening Scores by Division and K–12 Experience

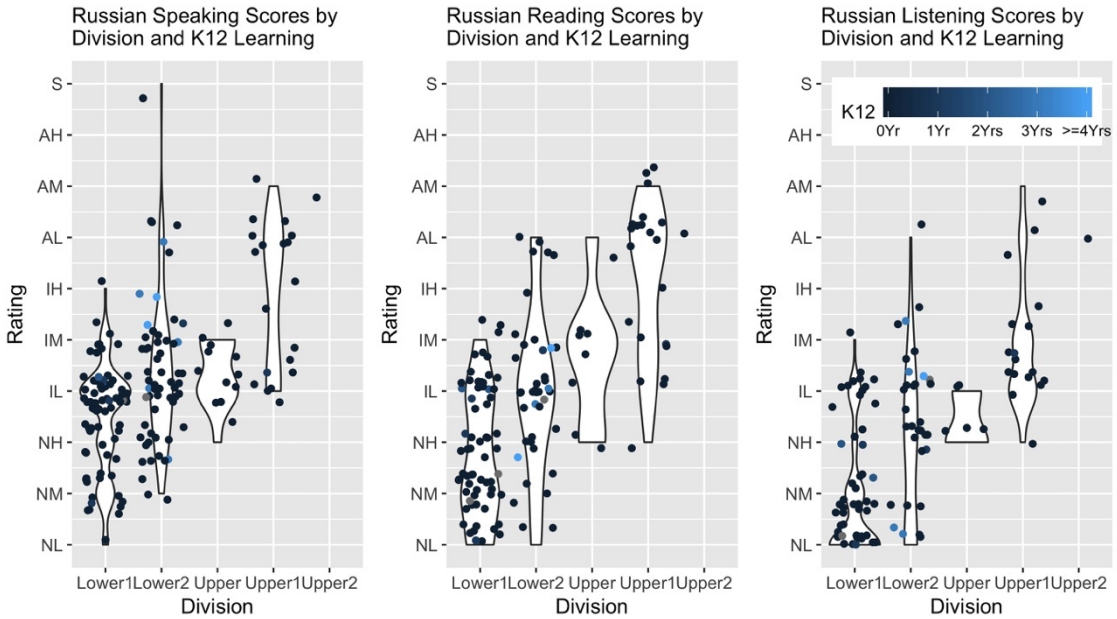
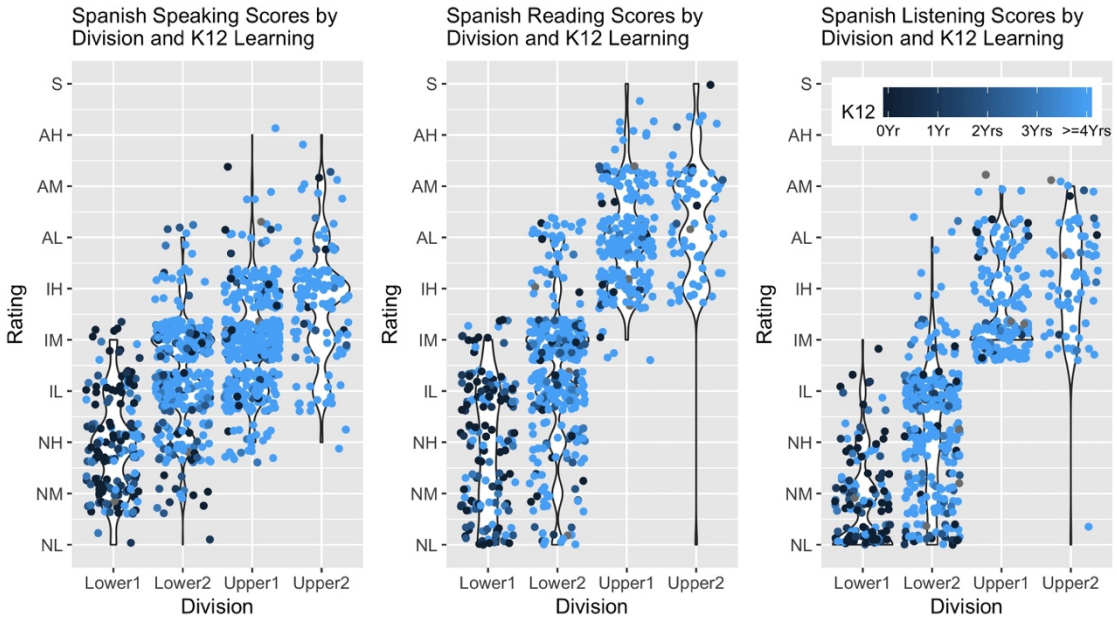


Figure 12

Spanish Speaking, Reading, and Listening Scores by Division and K–12 Experience



The Proficiency Profiles of Upper Division Junior and Senior Language Students by Language, Skill, and their Major or Minor Status

In Tables 10, 11, and 12, we report the ACTFL scores of upper-division juniors and seniors by language and their major/minor status to best replicate findings from Carroll's (1967) study of foreign language majors (see Table 1).

Using the median scores as Carroll (1967) did, 2017 French, Russian, and Spanish foreign language majors are not performing drastically differently from the 1965 majors represented in Carroll's study. Carroll found French majors in 1965 to have median speaking, reading, and listening scores as follows: Advanced Low, Superior, and Advanced Mid. In 2017, French majors obtained speaking, reading, and listening median scores that were Advanced Low across the board, which may be quite comparable to Carroll's findings given that the 2017 sample of majors includes juniors as well as seniors (Carroll's sample included only seniors) and juniors and seniors who most likely have more double majors and additional minors than the students in Carroll's sample had in 1965.

Russian scores are not quite as high as the 1965 cohort, which had reached Intermediate High in speaking, Advanced Mid in reading, and Advanced Low in listening. The median scores for the 2017 Russian cohort were close to Intermediate High in speaking, Intermediate High in reading, and Intermediate Low in listening.

Spanish median scores in 1965 were Advanced High in speaking, Superior in reading, and Advanced High in listening. In 2017, they were Intermediate High, Advanced Low, and Advanced Low, respectively.

Caveats in the comparisons abound, of course. For example, the tests used in 1965 are not the same tests as those used today. Test content across the two test eras most certainly is not directly comparable. The 1965 MLA tests and the 2017 ACTFL test versions were not equated. Equating could theoretically be done if a sample of students today took both tests with little time in between test administrations, but the issues of time-bound content appropriateness and unequal content familiarity would pose true equating difficulties. Proficiency level definitions, how raters are trained, and how raters interpret proficiency bands may have experienced drift over the years, so it is difficult to fully understand if the scores from 1965 testing are directly comparable to 2017 testing. Additionally, it is important to remember that Carroll's data included only majors, whereas our study describes the proficiency attainment of majors, but also minors and students who were not pursuing a degree in the language that they were studying. Nonetheless, the comparisons are interesting, and they call into question the lack of national, annual assessments of foreign language proficiency in K–16 educational programming in the United States. Indeed, it may be time to reinstate the groundwork on the foreign language portion of the U.S. National Assessment of Educational Progress (NAEP) that began in the early 2000s, but that was abandoned shortly thereafter. College-level language programs could institute laboratory fees or other methods to ensure the standardized testing of exiting majors and minors as a degree requirement. Such targeted, annual testing could prove extremely informative for program administrators, educators, and students alike.

Tables 10, 11, and 12 display data from language majors/minors for the six languages discussed in this article. In Figures 13, 14, and 15, we plotted a subset of the data presented in those tables so that readers could visually inspect the ACTFL proficiency scores of upper-division juniors and seniors by language and major/minor status. We did not include data from Arabic, Portuguese, and Russian given the small number of majors/minors and non-majors/non-minors.

In Figures 13, 14, and 15, each small, light-colored dot in the background represents a test taker of Chinese, French, or Spanish, whose speaking/reading/listening rating can be found on the y axis. The dots are jiggered (slightly moved apart by our request within the statistical system *R*) so that the individual dots can be better seen. The large, solid dots in the foreground represent the average speaking/reading/listening ratings achieved by majors, minors, and non-major/non-minors studying Chinese, French, and Spanish. Whiskers around the mean are error bars that represent the 95% confidence intervals (CIs) for the mean.

In general, one might expect language majors to outperform language minors, and language majors and/or minors to outperform their classroom peers who are neither majoring nor minoring in the language. However, we found these expectations did not always appear to materialize in the data. Only for French, and only for speaking and reading, is the expected pattern seen. This can be seen in Figures 13 and 14, where French learners appear to conform to the expectations of majors doing best, minors doing second best, and non-majors/non-minors falling behind majors and minors. But visual inspection of Chinese and

Spanish data in Figures 13–15 appears to show different patterns. The Spanish data suggest that non-majors/non-minors outperform minors in all modalities and, in the case of Reading (Figure 14), even outperform or are at least equal to majors. The Chinese data also show a different pattern with non-majors/non-minors doing better, in terms of proficiency outcomes, than majors and minors in reading and listening (Figures 14 and 15) and better than minors in speaking (Figure 13). Further statistical analyses will help in determining the validity of these visual inspections, especially in light of the fact that the error bars (the 95% CIs, which indicate the variation in the data) are quite large and overlap across categories. Why these patterns occur is a matter of speculation, and qualitative data may help shed light on this issue. It may also help in future studies to control for learner factors such as heritage status.

Certainly, these data pose more questions. For example, most educators may think of upper-division language learners as primarily majors or minors. But if they are not, and if this third category of non-majors and non-minors is large and outperforming the minors, questions arise: Who are they? Why are they taking upper-division courses in the language? Why are they neither majoring nor minoring in the language? If language programs test the proficiency of only majors and minors upon exiting or graduation, are the language programs going to miss a large piece of acquisition and programmatic information?

Table 10

Speaking Scores of Upper-Division Juniors and Seniors by Language and Major/Minor Status

Language	Major/Minor	*n*	M	SD	Median	Min	Max
Arabic speaking	Major	2	2.50	2.12	2.5	1	4
	Minor	3	4.33	0.58	4	4	5
	Not major nor minor	13	5.85	1.82	6	2	8
Chinese speaking	Major	30	5.17	1.46	5	3	8
	Minor	24	4.29	2.18	4	1	10
	Not major nor minor	11	4.45	2.66	4	2	9
French speaking	Major	26	6.50	1.92	7	1	9
	Minor	54	5.44	1.09	5	3	8
	Not major nor minor	35	5.31	1.32	5	3	9
Portuguese speaking	Major	7	5.57	0.79	5	5	7
	Minor	1	7.00	NA	7	7	7
	Not major nor minor	4	5.75	0.96	5.5	5	7
Russian speaking	Major	18	5.00	1.41	4.5	3	8
	Minor	8	5.38	1.19	5	4	7
	Not major nor minor	1	4.00	NA	4	4	4
Spanish speaking	Major	81	6.00	1.02	6	3	9
	Minor	161	4.99	1.02	5	3	8
	Not major nor minor	70	5.20	1.04	5	4	8

Table 11

Reading Scores of Upper-Division Juniors and Seniors by Language and Major/Minor Status

Language	Major/Minor	*n*	M	SD	Median	Min	Max
Arabic reading	Major	2	1.00	0.00	1	1	1
	Minor	3	3.67	1.53	4	2	5
	Not major nor minor	10	4.90	1.37	5	2	7
Chinese reading	Major	16	4.62	2.22	5	1	7
	Minor	11	4.82	2.27	5	1	7
	Not major nor minor	6	5.67	2.16	7	2	7
French reading	Major	25	7.16	0.75	7	6	8
	Minor	28	6.89	0.88	7	5	9
	Not major nor minor	22	6.82	0.80	7	5	8
Portuguese reading	Major	7	6.86	0.69	7	6	8
	Minor	1	6.00	NA	6	6	6
	Not major nor minor	4	6.00	1.15	6	5	7
Russian reading	Major	16	5.31	1.49	5	3	7
	Minor	6	5.67	1.63	6	3	7
	Not major nor minor	2	5.50	0.71	5.5	5	6
Spanish reading	Major	64	7.22	1.28	7	1	10
	Minor	95	7.12	1.00	7	6	10
	Not major nor minor	34	7.24	1.07	7	6	10

Table 12

Listening Scores of Upper-Division Juniors and Seniors by Language and Major/Minor Status

Language	Major/Minor	*n*	M	SD	Median	Min	Max
Arabic listening	Major	2	1.00	0.00	1	1	1
	Minor	3	1.00	0.00	1	1	1
	Not major nor minor	13	3.46	1.85	4	1	6
Chinese listening	Major	22	4.23	1.66	4	1	7
	Minor	11	4.36	2.16	4	1	7
	Not major nor minor	6	5.50	2.51	7	1	7
French listening	Major	24	6.50	0.78	7	5	7
	Minor	43	5.98	0.83	6	5	8
	Not major nor minor	24	6.29	0.91	6.5	5	8
Portuguese listening	Major	7	5.71	1.25	5	4	7
	Minor	1	7.00	NA	7	7	7
	Not major nor minor	4	6.00	2.00	7	3	7
Russian listening	Major	12	4.17	1.11	4	3	7
	Minor	4	4.25	0.96	4.5	3	5
	Not major nor minor	2	6.00	2.83	6	4	8

Language	Major/Minor	*n*	M	SD	Median	Min	Max
Spanish listening	Major	54	6.46	1.18	7	1	8
	Minor	75	5.75	0.82	6	5	8
	Not major nor minor	27	5.89	0.89	6	5	7

Figure 13

Speaking Scores of Upper-Division Juniors and Seniors by Language and Major/Minor Status

Figure 14

Reading Scores of Upper-Division Juniors and Seniors by Language and Major/Minor Status

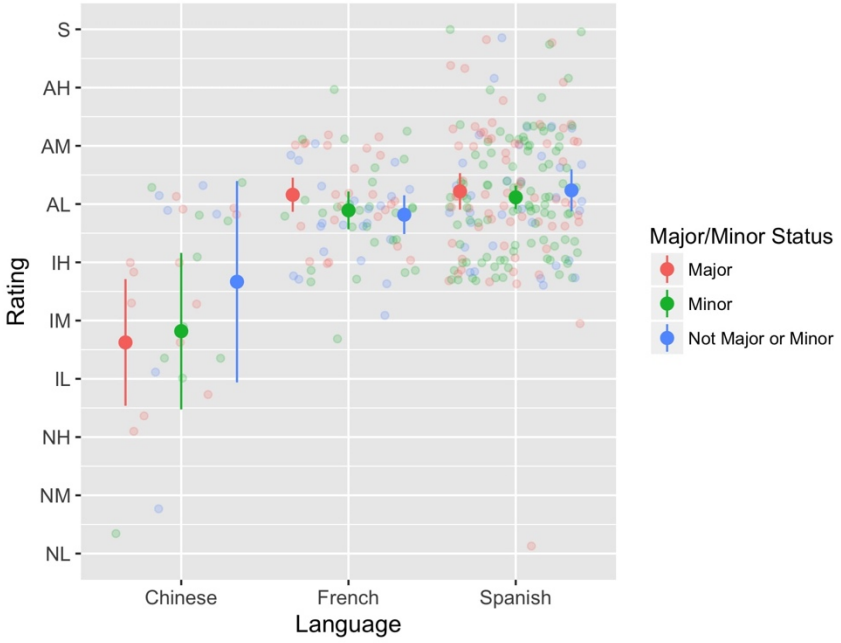
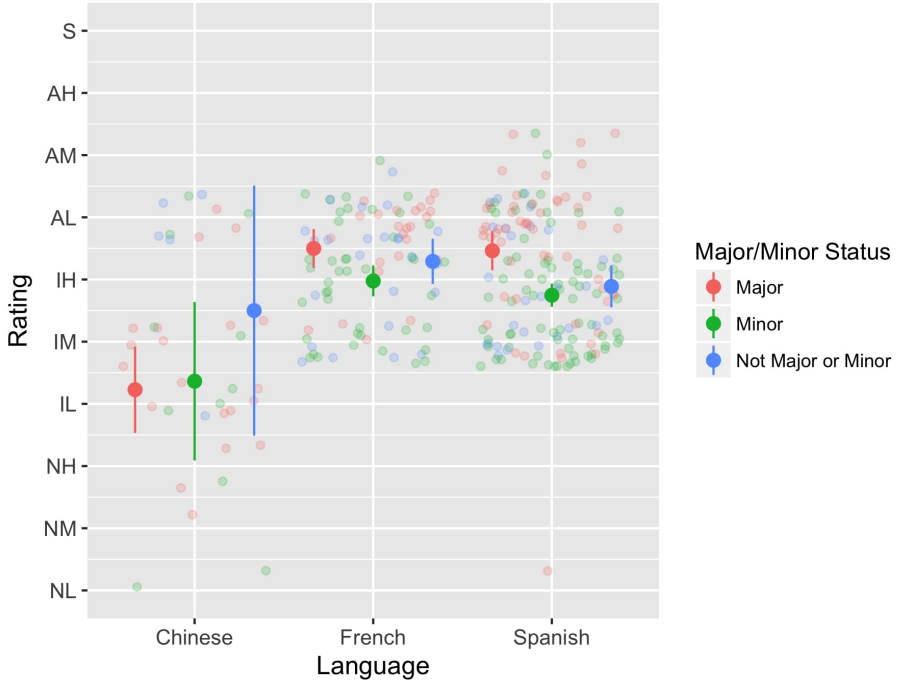


Figure 15

Listening Scores of Upper-Division Juniors and Seniors by Language and Major/Minor Status



Discussion

In the United States, college-level foreign language educators have opportunities for reflection, and the ability to critically assess their goals and mission, when they take advantage of outcomes assessment to

monitor proficiency achievement in their foreign language programs (Norris, 2006). Such work is needed, as called for by the MLA (2007), because Americans must, in larger numbers, study languages to the Advanced level of proficiency. As described in the 2007 MLA report, U.S. college language learners need to obtain "deep translingual and transcultural competence" so that they can better engage within a global, post-9/11 world. The MLA recommended that foreign language programs set clear standards of achievement for language students based on their major or primary field of study, be it the target language, international studies, music, or a STEM (science, technology, engineering, or math) field. The MLA also encouraged programs to liaise with K–12 educators, develop programs for gifted and heritage learners, increase opportunities for coursework in translation and interpretation services, and broaden the scope of language education on college campuses to attract and maintain a diverse, significantly and sustainably large, and committed language learner pool. At the center of these calls for improvement is a review of who the students in the programs are and how their proficiency is linked to their coursework and their background characteristics, such as their heritage status, their pre-college language learning experiences, and their decisions to major or minor in the language.

In this study, we sought to understand better one large cohort of foreign language learners: students studying six different languages at three large public universities in the spring of 2017. By taking a cross-sectional look at the students' language proficiency achievement, and by tagging those outcomes by select learner background characteristics, we could more clearly ask ourselves about our foreign language programs' goals, and the overall foreign language program missions at our universities.

The latest MLA enrollment report showed a significant disparity in enrollments between lower- and upper-division language courses, with approximately five students enrolled in lower division for every student in upper division courses (Looney & Lusin, 2018). Strawbridge et al. (2019), also, using the data generated by The Language Flagship Proficiency Initiative, indicated that the students who populate advanced (upper division) language courses do not tend to come from the programs' introductory (lower division) courses, and that the students in foreign language programs may constitute two separate populations: one group taking language courses mostly to satisfy a requirement, and the other to pursue a degree, which was noted more than two decades ago by Tschirner and Heilenman (1998). In the data in this article, however, a third group seems to be present in the upper division: those interested in taking the language not to satisfy a requirement and not to pursue a degree. Therefore, in order to answer our first research question (What is the current status of proficiency within academic language programs at the post-secondary level?) we looked at the data collected for this study by separating the discussion of upper- and lower-division courses.

Considering upper-division courses in general, the highest levels of proficiency were achieved in reading, where students who had completed four years of instruction at the post-secondary level often reached the Advanced Low level of proficiency. In listening, non-heritage learners frequently reached the Intermediate High level and, in some languages, heritage learners reached Advanced Low. The weakest modality in terms of ultimate attainment tended to be speaking. With the exception of Spanish heritage learners, who consistently reached Advanced Low, the majority of the students tested after four years of university-level coursework only attained between Intermediate Mid and Intermediate High proficiency in speaking. Overall, the picture that emerges from upper-division courses indicates that language graduates are typically strongest in reading and weakest in speaking, and that there is a large range of proficiency levels within the upper-division students.

The picture is quite different in the case of lower-division courses, where the ranking of the three modalities in terms of proficiency achieved is almost the opposite of what we found in the upper division. Speaking proficiency is strongest on average in all languages but Portuguese for non-heritage learners in the lower division, with the Intermediate level an achievable target for non-heritage learners, and with heritage learners stronger still. Arabic non-heritage learners, in fact, approached Intermediate Mid in speaking at the end of Lower 2. In reading, students of Spanish, French, and Russian showed similar levels to those of speaking, reaching or approaching Intermediate Low. In Spanish, reading and speaking are at the same levels in Lower 1 and Lower 2. In French, reading levels mirror listening levels, slightly below speaking

ratings at Lower 2. Speaking is consistently at least one sublevel higher than reading in Arabic throughout and in Chinese through Lower 2. Reading in Chinese lags further behind where, on average, students are still at Novice Mid at Lower 2. Students of Arabic, Spanish, and Russian achieved or approached Novice High in listening, although heritage learners could get as high as Intermediate Mid. Listening proficiency for French learners at Lower 2 was the equivalent of their reading proficiency, at Intermediate Low, and Portuguese learners demonstrated Intermediate Mid listening proficiency. Surprisingly, the median listening proficiency level of Chinese non-heritage learners when they complete the lower division sequence is only Novice Low. In sum, the proficiency profile of lower-division students shows stronger levels in speaking and reading with slightly lower levels in listening.

Overall, as expected, students' proficiency attainment is generally higher in the Romance languages than in those that are typologically most distant from English. For example, whereas non-heritage learners in Chinese (n = 40) completed the lower-division sequence with a median rating of Novice Mid in reading, the same group reached reading levels of Intermediate Low in French (n = 148) and Intermediate Mid in Spanish (n = 334). Of note is the high degree of uniformity in the proficiency attainment of students in the three Romance languages despite presumably different teaching approaches and student populations.

Our second research question concerns the role of previous experience and heritage status on proficiency attainment: What roles do factors such as pre-university experience and language heritage background play in contributing to language proficiency? The median scores of heritage and non-heritage learners show that in general, heritage status conveys the expected benefit of higher proficiency in all the languages tested across levels and skills, but the differences between the two groups are not particularly large when considered overall, and are often no more than one sub-level on average. However, a closer look at the distribution of the data in the figures shows that the heritage learners tend to concentrate around the top of the score distribution both in lower and upper-division courses. This is particularly true for listening and, in particular, for speaking, but not necessarily in reading, as can be seen in Figures 2 and 6 (Chinese and Spanish, respectively). This is not surprising considering the predominance of orality over literacy in the exposure to the language of many heritage learners. Also, as expected, the diversity of proficiency profiles among heritage learners explains the large distribution of scores reflected in the data. For example, Table 4 and Figure 2 show that speaking proficiency levels for Chinese heritage students in upper-division courses ranges from Novice High to Advanced High.

The effect of previous language experience is perhaps even more significant for ultimate proficiency attainment than heritage status. As we mentioned earlier, few students who start language learning at the college level decide to continue beyond lower-division coursework. The majority of the students who populate the upper-division courses in Spanish (91%) and French (79%) have three or more years of previous language experience in school. Even in Chinese, the majority of students in upper-division courses have three or more years of previous Chinese instruction (55%). When students with little or no previous experience decide to continue beyond the second year, they tend to attain lower levels of proficiency than their peers who had more extensive previous language experience. This observation is borne out by our data. Figures 9 and 12, for example, show that the majority of the students in upper-division courses in French and Spanish, respectively, had significant language learning experience at the secondary level. Again, there are important differences among languages. Clearly, the role that previous experience plays in proficiency attainment is more obvious in those languages that are commonly taught in the K–12 system than in those that are rarely offered in the public schools (e.g., Portuguese and Russian), as shown in Figures 10 and 11.

Our third research question (Do language majors outperform language minors and students neither majoring nor minoring in the language, as one might expect?) generated perhaps the most unexpected results. When looking at the performance of juniors and seniors enrolled in upper-division courses, one might expect that majors would do better than minors, and minors would outperform those who are not seeking a degree in the language. However, this distribution occurred only in one language and in one skill: French speaking. In all other cases there were many instances where the group of upper-division non-majors/non-minors

outperformed the minors and, sometimes even the majors, or at least was on par with them in the skill. In listening (Arabic, Chinese, and Russian), reading (Arabic, Chinese, and Spanish) and speaking (Portuguese) the group of upper-division non-degree seekers performed better than the other two. This finding is significant if we consider that, with the exception of Russian, the number of non-degree seekers represents a sizable percentage of the total number of juniors and seniors in the upper-division courses. However, we need to be cautious given that the numbers for Arabic, Portuguese, and Russian majors and minors are small. Replications of these findings that include statistical testing are needed before we can be confident about comparing upper-division students who are majors, minors, and non-majors/non-minors.

Implications

Our data present findings that can inform discussions about the curriculum and pedagogy of language programs. For example, consider the data on the speaking proficiency of students at the lower and upper levels. The strength of speaking proficiency in the lower division, particularly among non-heritage learners, is evidence of instructional emphasis on speaking within a communicative framework. Students have an opportunity to develop speaking proficiency through classroom instruction and can reach the Intermediate level, even in non-typologically similar languages. The data show that there is comparatively lower attainment at the lower instructional level in listening and reading proficiency. Language program directors might consider how to buttress listening and reading skills within a communicatively focused approach to language teaching and learning. In contrast, language majors do not typically reach the Advanced level of speaking proficiency. The slowdown in speaking proficiency gains in the upper division may be due a greater emphasis in upper-division courses on reading and listening, with less time to develop speaking toward the Advanced level within a classroom setting.

Overall, the data also suggest that language program directors will benefit from more research on skill development other than speaking. The cross-sectional displays of the data show that neither reading nor listening proficiency is as strong as one would like, and that in Spanish, Portuguese, and Russian, listening proficiency lags behind other skills at the lower levels of the curriculum. Whether this is natural or not is a matter of debate and could prompt researchers to ponder a number of questions: Is the skill of listening under-calibrated on proficiency scales in relation to the other skills? Does listening follow a growth pattern different from the other skills? Does the relationship of listening to reading differ among languages?

Data from the background questionnaire also provides useful information for language program directors. For example, our study sheds light on the effect of heritage status and K–12 study on attainment by language. Proficiency attainment in college-level programs must be qualified by such background characteristics, or a major portion of the process of SLA is unaccounted for and may lead to biased results. Additionally, a review of the background questionnaires of language majors at Michigan State University revealed that the majority of language majors are in fact double majors and that there were many more language minors than majors (see Winke, et al., 2019). This is a reality that differs starkly from the situation encountered by both Carroll (1967) and Teschner (1991) and that has important implications for language programs. For example, are programs addressing the needs of all their students? If the goals of a program include retention, is enough being done to retain and support the continued and sustained learning of all subgroups within the program? We show in this article that identifying students in terms of proficiency attainment and background characteristics can help programs answer such questions.

Conclusion

With the publication of this article, we provide foreign language educators with a glimpse of how informative proficiency test data coupled with learners' background characteristics can be. The role other background characteristics play in proficiency development could and should be investigated. The impact of such factors as learners' study abroad experiences (Davidson, 2010), learners' effort in their foreign language classes, and how much they use the language outside of class are also important to understanding

levels of proficiency. We collected such additional data, but due to space considerations, were not able to report on those data in this article. Instead, we encourage researchers and applied linguists to take our data and to use them to investigate the effects of such variables on proficiency achievement and attainment. To that end, we have published the data used in this study, and other data (including additional test data collected during the 2014–2015 and 2015–2016 academic years) gathered as part of the larger Language Proficiency Flagship Initiatives at Michigan State University, the University of Utah, and the University of Minnesota, in a public data repository. The data are stored in freely accessible and downloadable files at the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan (Winke, et al., 2020). We will use the data, and we hope others will also, to answer questions such as:

- How much does inside- and outside-of-class learning effort impact proficiency attainment across and within college-level foreign language programs? (Based on Rifkin's 2005 work, if outside-of-class learning counts as learning time, it may indeed have an impact.)
- How does the timing of high-school language study affect college-learning outcomes? Specifically, does it matter (for college-learning) when a person studies a foreign language during high school (for example, during the first two years or the last two years, if only two years of high-school language study are available)? Could the timing of the two years of high-school study affect the person's ultimate post-secondary attainment? Does when a person studies a foreign language during high school affect their likelihood of majoring or minoring in the language in college?
- Using longitudinal test data (same student tested at different points during undergraduate study), how does proficiency grow over time at the college level, and what background characteristics affect the growth? (See Isbell et al., 2019, and Zhang, Winke, & Clark, 2020, for two preliminary studies on proficiency growth using Michigan State University subsets of these Flagship data.)
- How does study abroad and the length of study abroad impact proficiency attainment and growth within and across language programs?

Acknowledgements

The Language Proficiency Flagship Initiative grants for this data collection project came from the Language Flagship within the National Security Education Program (NSEP) and the Defense Language and National Security Education Office (DLNSEO).

The grants were administered through the Institute of International Education (IIE) in Washington, DC. We would like to thank Erwin Tschirner, Margaret E. Malone, and Helen Hamlyn for their valuable assistance with data collection, data information, and testing assistance. We also thank Samuel D. Eisen and Kaveri Advani at DLNSEO, and Carrie Reynolds and Chelsea Sypher at IIE, for their grant guidance and overall project support.

Notes

1. Throughout this article, we use the following abbreviations to refer to ratings from the American Council on the Teaching of Foreign Languages (ACTFL) proficiency scale: Novice Low (NL), Novice Mid (NM), Novice High (NH), Intermediate Low (IL), Intermediate Mid (IM), Intermediate High (IH), Advanced Low (AL), Advanced Mid (AM), Advanced High (AH), and Superior (S). For our statistical analyses, we transformed these into a scale of 1 to 10, with 1 representing Novice Low, and 10 representing Superior.

2. In 2012 NSEP merged with the Defense Language Office to form the Defense Language & National Security Education Office (DLNSEO).

3. cf. Murphy & Evans-Romaine, 2017, and Nugent & Slater, 2017, for in-depth discussion of how the Language Flagship Program came into existence and how it was intended to interface with K–16 foreign

language instruction.

4. For additional discussions of the results from this database, see Gass, et al., 2016; Hacking & Tschirner, 2017; Isbell, et al., 2019; Ma & Winke, 2019; Soneson, in press; Strawbridge, et al., 2019; Tigchelaar, et al., 2017; Tschirner, et al., 2018; Winke & Gass, 2018a, 2018b, 2019; Winke, Heidrich Uebel, & Gass, in press; Zhang, et al., 2020.

5. Students of Arabic at the University of Utah took the computer-adaptive listening test from Brigham Young University, administered by Language Testing International, that returned the same range of ratings as the LPT.

References

ACTFL. (2012). *ACTFL Proficiency Guidelines 2012*. http://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics, 39*(3–4), 129–181. https://doi.org/10.1515/tl-2013-0009

Carroll, J. B. (1967). Foreign language proficiency levels attained by language majors near graduation from college. *Foreign Language Annals*, *1*(2) 131–151. https://doi.org/10.1111/j.1944-9720.1967.tb00127.x

Davidson, D. E. (2010). Study abroad: When, how long, and with what results? New data from the Russian front. *Foreign Language Annals*, *43*(1), 6–26. https://doi.org/10.1111/j.1944-9720.2010.01057.x

Davidson, D. E., Garas, N., & Lekic, M. D. (2017). Assessing language proficiency and intercultural development in the overseas immersion context. In D. Murphy & K. Evans-Romaine (Eds.), *Exploring the US Language Flagship Program: Professional competence in a second language by graduation* (pp. 156–176). Multilingual Matters.

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics, 31*(3), 413–438. https://doi.org/10.1017/S0142716410000056

Freed, B. F. (1987). Preliminary impressions of the effects of a proficiency-based language requirement. *Foreign Language Annals*, *20*(2), 139–146. https://doi.org/10.1111/j.1944-9720.1987.tb02932.x

Gass, S., Winke, P., & Van Gorp, K. (2016). The Language Flagship Proficiency Initiative. *Language Teaching*, *49*(4), 592–595. https://doi.org/10.1017/S0261444816000215

Goertler, S., Kraemer, A., & Schenker, T. (2016). Setting evidence-based language goals. *Foreign Language Annals*, *49*(3), 434–454. https://doi.org/10.1111/flan.12214

Hacking, J. F., & Tschirner, E. (2017). The contribution of vocabulary knowledge to reading proficiency: The case of college Russian. *Foreign Language Annals*, *50*(3), 500–518. https://doi.org/10.1111/flan.12282

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition, 177*, 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Hernández, T. (2006). Integrative motivation as a predictor of success in the intermediate foreign language classroom. *Foreign Language Annals*, *39*(4), 605–617. https://doi.org/10.1111/j.1944-9720.2006.tb02279.x

Herzog, M. (n.d.). *An overview of the history of the ILR language proficiency skill level descriptions and scale*. Interagency Language Roundtable. https://www.govtilr.org/Skills/IRL%20Scale%20History.htm

Isbell, D. R., Winke, P., & Gass, S. (2019). Using the ACTFL OPIc to assess proficiency and monitor progress in a tertiary foreign languages program. *Language Testing*, *36*(3), 439–465. https://doi.org/10.1177/0265532218798139

Kang, S., & Lim, B. (2020). Implications of ILR/ACTFL Proficiency Guidelines and text typology in teaching and testing college level Korean. *The Korean Language in America, 9*, 126–143.

Looney, D., & Lusin, N. (2018). *Enrollments in languages other than English in United States institutions of higher education, Summer 2016 and Fall 2016: Preliminary report*. Modern Language Association of America. https://www.mla.org/content/download/83540/2197676/2016-Enrollments-Short-Report.pdf

Ma, W., & Winke P. (2019). Self-assessment: How reliable is it in assessing oral proficiency over time? *Foreign Language Annals*, *52*(1), 66–86. https://doi.org/10.1111/flan.12379

Magnan, S. S. (1986). Assessing speaking proficiency in the undergraduate curriculum: Data from French. *Foreign Language Annals*, *19*(5), 429–438. https://doi.org/10.1111/j.1944-9720.1986.tb01031.x

MLA. (2007). *Foreign languages and higher education: New structures for a change world.* https://www.mla.org/Resources/Research/Surveys-Reports-and-Other-Documents/Teaching-Enrollments-and-Programs/Foreign-Languages-and-Higher-Education-New-Structures-for-a-Changed-World

Murphy, D., & Evans-Romaine, K. (Eds.). (2017). *Exploring the US Language Flagship Program*. Multilingual Matters.

Norris, J. M. (2006). The issue: The why (and how) of assessing student learning outcomes in college foreign language programs. *The Modern Language Journal*, *90*(4), 576–583. https://www.jstor.org/stable/4127045

Norris, J. M., & Pfeiffer, P. C. (2003). Exploring the uses and usefulness of ACTFL oral proficiency ratings and standards in college foreign language departments. *Foreign Language Annals*, *36*(4), 572–581. https://doi.org/10.1111/j.1944-9720.2003.tb02147.x

Nugent, M., & Slater, R. (2017). The Language Flagship: Creating expectations and opportunities for professional-level language learning in undergraduate education. In D. Murphy & K. Evans-Romaine (Eds.), *Exploring the US Language Flagship Program: Professional competence in second language by graduation* (pp. 9–28). Multilingual Matters.

Rifkin, B. (2005). A ceiling effect in traditional classroom foreign language instruction: Data from Russian. *The Modern Language Journal, 89*(1), 3–18. https://doi.org/10.1111/j.0026-7902.2005.00262.x

Sanders, R. F. (2005). Redesigning introductory Spanish: Increased enrollment, online management, cost reduction, and effects on student learning. *Foreign Language Annals*, *38*(4), 523–532. https://doi.org/10.1111/j.1944-9720.2005.tb02519.x

Schmitt, E. (2014). Seat time versus proficiency: Assessment of language development in undergraduate students. In J. M. Norris & N. Mills (Eds.), *Innovation and accountability in language program evaluation* (pp. 110–130). Cengage.

Soneson, D. (in press.). Advanced users of Spanish: Relationships among modalities of speaking, reading, and listening. In M. Menke & P. Malovrh (Eds.), *Advancedness in second language Spanish:*

Definitions, challenges, and possibilities. John Benjamins.

Strawbridge, T., Soneson, D., & Griffith, C. (2019). Lasting effects of pre-university language exposure on undergraduate proficiency. *Foreign Language Annals*, *52*(4), 776–797. https://doi.org/10.1111/flan.12430

Swender, E. (2012). *Introducing the ACTFL Proficiency Guidelines 2012* [Power Point Slides]. Interagency Language Roundtable. https://www.govtilr.org/Publications/ACTFL%20Guidelines%202012%20ILR%20Presentation%201.6.pdf

Swender, E., Martin, C. L., Rivera-Martinez, M., & Kagan, O. E. (2014). Exploring oral proficiency profiles of heritage speakers of Russian and Spanish. *Foreign Language Annals, 47*(3), 423–446. https://doi.org/10.1111/flan.12098

Teschner, R. V. (Ed.). (1991). *Assessing foreign language proficiency of undergraduates*. Heinle & Heinle.

The Language Flagship. (n.d.). *The next generation of global professionals.* https://www.thelanguageflagship.org

Thompson, I. (1996). Assessing foreign language skills: Data from Russian. *The Modern Language Journal*, *80*(1), 47–65. https://doi.org/10.1111/j.1540-4781.1996.tb01137.x

Tigchelaar, M., Bowles, R. P., Winke, P., & Gass, S. (2017). Assessing the validity of ACTFL Can-Do Statements for spoken proficiency: A Rasch analysis. *Foreign Language Annals*, *50*(3), 584–600. https://doi.org/10.1111/flan.12286

Tschirner, E. (1992). Oral proficiency base lines for first- and second-year college German. *Die Unterrichtspraxis/Teaching German*, *25*(1), 10–14. https://www.jstor.org/stable/3531404

Tschirner, E., Hacking, J. F., & Rubio, F. (2018). Reading proficiency and vocabulary size: An empirical investigation. In P. Ecke & S. Rott (Eds.), *Understanding vocabulary learning and teaching: Implications for language program development* (pp. 49–68). Heinle & Heinle.

Tschirner, E., & Heilenman, L. K. (1998). Reasonable expectations: Oral proficiency goals for intermediate-level students of German. *The Modern Language Journal*, *82*(2), 147–158. https://doi.org/10.1111/j.1540-4781.1998.tb01187.x

Winke, P., Heidrich Uebel, E. S., & Gass, S. (in press). Individual differences in advanced Spanish proficiency: Cluster analysis and case-matching analyses on 127 advanced learners. In M. Menke & P. Malovrh (Eds.), *Advancedness in second language Spanish: Definitions, challenges, and possibilities*. John Benjamins.

Winke, P., & Gass, S. M. (2018a). Individual differences in advanced proficiency. In P. A. Malovrh & A. Benati (Eds.), *The Wiley handbook of advanced proficiency in second language acquisition* (pp. 157–178). John Wiley & Sons.

Winke, P., & Gass, S. (2018b). When some study abroad: How returning students realign with the curriculum and impact learning. In C. Sanz & A. Morales-Front (Eds.), *The Routledge handbook of study abroad research and practice* (pp. 527–544). Routledge.

Winke, P., & Gass, S. M. (Eds.). (2019). *Foreign language proficiency in higher education: Curricular and assessment issues*. Springer. https://doi.org/10.1007/978-3-030-01006-5

Winke, P., Gass, S., & Heidrich, E. S. (2019). Foreign language proficiency in higher education. In P. Winke & S. M. Gass (Eds.), *Foreign language proficiency in higher education: Curricular and assessment issues* (pp. 117–136). Springer. https://doi.org/10.1007/978-3-030-01006-5_6

Winke, P. M., Gass, S. M., Soneson, D., Rubio, F., & Hacking, J. F. (2020). *Foreign language proficiency test data 2014-2017 from three U.S. universities*. Inter-university Consortium for Political and Social Research (ICPSR). https://doi.org/10.3886/ICPSR37499.v1

Zhang, X., Winke, P., & Clark, S. (2020). Background characteristics and oral proficiency development over time in lower-division college foreign language programs. *Language Learning*, *70*(3), 807–847. https://doi.org/10.1111/lang.12396

About the Authors

Paula Winke is a Professor in the Department of Linguistics and Languages and Director of the Second Language Studies Program at Michigan State University. She is the recipient of the 2012 TESOL International Distinguished Researcher Award, and the 2020 American Association of Applied Linguistics Article of the Year Award.

Xiaowan Zhang is currently a fourth-year PhD student (ABD) in the Second Language Program at Michigan State University. Her research interests include language testing and assessment, language policy, instructed SLA, and quantitative research methods. She received her master's degree in TESOL in 2016 from the University of Illinois at Urbana-Champaign.

Fernando Rubio is Professor of Spanish Linguistics and Director of the Second Language Teaching and Research Center at the University of Utah. He publishes in the field of second language acquisition, language teaching methodology and computer-assisted language learning.

Susan Gass is University Distinguished Professor at Michigan State University. She has published widely in the field of second language acquisition and is the recipient of numerous awards and honors. She is the current editor of Studies in Second Language Acquisition and co-editor of the Routledge Second Language Acquisition Research Series.

Dan Soneson, Director of the Language Center in the College of Liberal Arts, University of Minnesota Twin Cities, was the Principal Investigator of the Proficiency Assessment for Curricular Enhancement (PACE) project at the University of Minnesota, a four-year federal grant within the proficiency initiative sponsored by the Language Flagship.

Jane F. Hacking is Professor of Russian at the University of Utah. Her research areas are Second Language Phonology and Second Language Proficiency. She co-founded and co-directed the Second Language Teaching and Research Center at the University of Utah, now a Department of Education Title VI Language Resource Center.

E-mail: fernando.rubio@utah.edu