World Languages and Cultures Faculty Publications

Department of World Languages and Cultures

2022

Proficiency benchmarking in Spanish

Peter Swanson
US Air Force Academy, peters@tribcsp.com

Jean-Philippe Peltier

Jean W. LeLoup
SUNY College at Courtland, leloupj@snycorva.cortland.edu

Darin Earnest
United States Air Force Academy, Darin.Earnest@usafa.edu

Margaret E. Malone

Recommended Citation

Proficiency Benchmarking in Spanish

Challenge statement

The paradigm in world language teaching and learning has shifted, prioritizing proficiency testing and setting benchmarks for language learners. However, many programs either lack the funds, choose not to measure learners' proficiency, or avoid benchmarking student progress through the proficiency ladder. The following empirical research provides results about learners of Spanish and their proficiency in higher education, allowing program faculty to reflect on their own benchmarks.

Abstract

The Language Flagship programs were established at the turn of the century with the goal of creating programs that would move language learners to advanced levels of proficiency in a select number of critical languages (Winke & Gass, 2019). Later, the Flagships called for institutions of higher education to create a viable process to assess proficiency learning in high quality, well-established academic language programs. To answer that call, the present study examines outcomes via end of year proficiency testing in Spanish at the first and second levels of Spanish instruction at the United States Air Force Academy using the Adaptive Listening Tests and the Adaptive Reading Tests developed at Brigham Young University. Results indicate differences in gender, years of study of Spanish, and the number of years of Spanish study prior to attending the Academy. Additionally, the results from the present study are compared to Tschirner's (2016) comprehensive analysis of student outcomes in higher education on ACTFL reading and listening tests. The findings have implications for programs in higher education as well as those in K-12 education.

Proficiency Benchmarking in Spanish

What are reasonable expectations of language proficiency for students to attain after a specific learning sequence of language study? This question has challenged the field of language teaching and learning for decades. Since Carroll's (1967) study of language majors at graduation, instructors, students and administrators alike have struggled to establish reasonable expectations, communicate them to students and faculty and attain them in formal learning situations.

Recent focus on the importance of world languages for business, diplomacy and national security underscores the need to develop proficient speakers. In a 2019 report, ACTFL emphasized that 90% of businesses surveyed reported a need for employees with skills in languages other than English; the continued global nature of business suggests that such a need will continue to grow (ACTFL, 2019). At the same time, the recent American Academy of Arts and Sciences (2017) report shows that, despite this stated need in business, "the vast majority of American citizens remain monolingual" (p. vii). There is a strong need to set benchmarks for language proficiency and help learners develop this proficiency in many languages. Clearly, it is important to understand what can be and is attained after specific sequences of study. Without such data, students, instructors, administrators and other stakeholders cannot determine individual student and general program success, nor can they know when to intervene to improve programs and when to investigate practices that make some programs more successful than others. In addition, without benchmarks and data from other language programs, stakeholders may set goals that are too high or too low for their own groups. In the current study, the researchers investigated the baseline proficiency of cadets at the United States Air Force Academy (USAFA) in first and second year Spanish.

Literature Review

Carroll's (1967) study represented the first major investigation of student outcomes in

modern world languages. While more than 50 years old, the study is still exemplary; it investigated speaking, reading and listening outcomes in five languages (French, Italian, German, Spanish and Russian) from universities across the United States (U.S.). Carroll (1967) also examined some of the factors that were related to student outcomes, including gender, age, years of previous language study, overseas experience (or study abroad) and current year in university.

Carroll's (1967) study employed the Modern Language Association test and aligned it to the Interagency Language Roundtable (ILR) Scale. At that time, the ILR scale was newly used in government; in addition, the ILR scale was used because the study predated the development of the ACTFL Proficiency Guidelines, which are currently used in most academic and business contexts. The study is groundbreaking not only because it examined language outcomes on such a broad scale but also because it employed the relatively new ILR scale in this context. In addition, the use of the ILR scale meant that forthcoming research employing the not-yet-conceived ACTFL Guidelines could relate their results to this study in the future and thus establish benchmarks for university language majors. Carroll found the following outcomes among students studying French, German, Italian, Russian and Spanish as a major:

- Average attainment of an ILR 2+ (approximately an ACTFL Advanced-Mid or Advanced-High);
- The following factors correlated with higher levels of proficiency attained
  - Heritage language background
  - Study abroad
  - Elementary school language study
  - Language study at a large institution
- No difference between males and females.

Since Carroll (1967) was published, a few studies have examined student oral proficiency in higher education (e.g., Isbell, Winke, & Gass, 2018) or different factors shown to affect outcomes, especially study abroad (e.g., DeKeyser, 2014; Dewey et al., 2012; Freed, 1995; Hernandez, 2010; Vande Berg et al., 2009). However, there was still limited research focusing on general language proficiency outcomes in higher education world language programs for nearly 50 years. Moreover, the original languages Carroll highlighted are no longer the only focus of world language study in higher education. While Spanish, French, German, Italian and Russian are still in the top 20 languages in higher education, they have been joined and, in some cases, replaced by enrollments in American Sign Language, Japanese, Chinese and Arabic (Looney & Lusin, 2018). Therefore, Carroll's study provided essential but increasingly outdated information for decades as research in outcomes in higher education became more specialized (focusing on specific factors) and less general (examining outcomes *writ large)* for a 50-year period.

This gap was noticed and eventually acted upon. In 2014, the Flagship Initiative (The Language Flagship, 2013), a nationally funded effort to transform the way U.S. students learn languages and build their proficiency in critical languages (e.g., Arabic, Mandarin), released a request for proposals to address this gap. The program provided funding to investigate student outcomes in several languages at three state universities in the U.S. During the three-year grant period, nearly 9,000 university students took one or more language proficiency tests in the areas of reading, listening and speaking in Arabic, Chinese (Mandarin), French, German, Korean, Portuguese, Russian and Spanish (Winke & Gass, 2019). Specifically, the results of the studies showed a range of outcomes for student language learners across different institutions, in different levels of courses and with different backgrounds. For example, Isbell et al. (2018), in a study of oral proficiency outcomes, found that four semesters of language study at the university

level yielded an outcome of Intermediate-Low to -Mid in Chinese, French, Russian and Spanish among learners at large state universities.

The resulting research from this effort has been remarkable, including dozens of research articles and book chapters as well as an edited volume. At the same time, it merely scratches the surface of research that needs to be conducted, published, disseminated and replicated. As Malone (2019) pointed out, while this work is necessary and important, it is not sufficient to represent the wide array of possible outcomes at different kinds of institutions studying languages under varying conditions. For example, Carroll (1967) documented that students at large institutions outperformed students at small institutions; the Flagship-funded research was conducted at three large, public universities.

Tschirner (2016) published a comprehensive report of student outcomes in higher education on ACTFL reading and listening tests; many of the participants were part of the Flagship study. With more than 6,000 subjects who took these reading and listening tests, Tschirner was able to identify average outcomes after two, three, four, five and six semesters. Over 1,600 subjects took both tests in Spanish, and second semester learners were found to reach about Intermediate-Low in reading and just below Novice-High in listening, while fourth-semester learners reached Intermediate-Mid in reading and almost Intermediate-Low in listening (Tschirner, 2016). Although additional research is needed to determine the outcomes of students in different types of learning environments, Tschirner's data, as well as the outcomes from the Flagship project, provide benchmarks for comparison.

The present study examines the outcomes of participants at USAFA after two or four semesters of Spanish language study. USAFA's students represent one part of the higher education system and are underrepresented in language outcomes research. As frequently

highlighted in advocacy materials, world languages benefit many areas of U.S. life, including

education, business, security and diplomacy. Obviously, future leaders of the military have great

potential to influence security and even diplomacy issues; proficiency in a world language is

critical for such populations. Given the dearth of research at military service academies, the

present study was guided by the following research questions:

1. What level of proficiency in listening did participants in first-year and second-year

   Spanish attain?

2. What level of proficiency did first-year participants attain in reading?

3. What were the characteristics of participants who attained the highest and lowest levels of

   proficiency?

   a. Was there a difference in outcomes based on gender and years of study of Spanish

      prior to attending USAFA?

   b. How did participants differ at the upper and lower quartiles of proficiency?

4. How did these results compare to Tschirner's (2016) study of students enrolled in
   language study?

Methods

Background and Setting

The mission of the Department of Foreign Languages and International Programs (DFFL)

at USAFA is to produce culturally attuned and linguistically capable Airmen. Its graduates

deploy worldwide in support of the U.S. strategic interests and engagements. Simply stating that

USAFA is producing culturally and linguistically enabled officers, however, is insufficient.

There is a need to continually assess and ensure that USAFA's programs are meeting the needs

of the United States Air Force.

Faculty in DFFL teach eight languages: Arabic, Chinese, French, German, Japanese,

Portuguese, Russian, and Spanish. Prior to 2020, faculty members in each language developed a set of outcomes aligned with a modified set of the *World-Readiness Standards for Learning Languages* (The National Standards Collaborative Board, 2015)—Communication, Cultures, Connections, and Careers, which replaced Comparisons and Communities. Faculty in each language community developed and established their own desired learning outcomes tied to these standards. At the end of a typical eight-semester program, or approximately 400 hours of instruction, DFFL administered the Defense Language Proficiency Test (DLPT)—the Department of Defense standard test for all linguists across all branches of the armed forces. Throughout the years, the DLPT served as the main metric in assessing cadets' second language proficiency although it only assesses ability in the receptive skills (i.e. listening and reading).

However, at USAFA, two issues emerged regarding the assessment of cadets' second language abilities. First, it was difficult to compare stated goals with progress across all eight languages. Each language developed its own set of outcomes based on DFFL's modified national standards goal areas of the 4Cs. Starting in the 2020-2021 academic year, DFFL's eight language communities developed Language Roadmaps, which were aligned with the *ACTFL Proficiency Guidelines* (ACTFL, 2012) and the *NCSSFL- ACTFL Can-Do Statements* (ACTFL, 2017). This alignment was used to set benchmarks for cadets at each language level. The alignment allowed DFFL faculty to set a foundation for comparison across its eight programs by allowing language communities to observe how one program might aim for Novice-High after 160 hours of instruction while another might set its sights on Intermediate-Low. Fundamentally, it aligned DFFL with established national standards while allowing various languages programs to compare, gain insight, and collaborate based on a mutually accepted foundation.

The second issue is that the DLPT did not provide faculty the feedback and gradation

necessary to fine-tune DFFL programs. Because the DLPT was not aligned with the World-Readiness Standards, the faculty did not believe it could be used as a reliable measure for the each of the language community's stated objectives. The first step to bridging this gap was adopting Brigham Young University's Adaptive Reading Test and Adaptive Listening Test. These assessments are both clearly tied to *ACTFL Proficiency Guidelines* (2012). The use of these tests allowed DFFL to assess all language programs and provide individual students targeted feedback based on their results. Starting with the 2021-2022 academic year, DFFL randomly tested a subset of cadets across all levels of all eight programs to ensure that each language community was meeting its clearly defined goals as articulated in their language roadmap.

Although cadets cannot major in a language, language minors or a degree in foreign area studies (FAS) are commonplace. FAS majors can choose a language, a region, and a specific area of academic focus (e.g., Spanish, Latin America, and Political Science). Approximately 60 cadets graduate annually with a minor in Spanish. All first-year cadets are required to study a language during their initial year at USAFA. All cadets take the DFFL language placement test during basic training; they can test out of the requirement with Advanced Placement exam scores or via the placement test. Based on the results, they can validate one semester or the full year; they can also test into a higher level. Cadets who place into higher levels include those with a substantial school-based or heritage language background. Therefore, these cadets show a wide range of language backgrounds, not dissimilar to their counterparts at more traditional institutions of higher education. With respect to the present study, cadets in their first year at USAFA took both the Adaptive Reading and Adaptive Listening Tests created by Brigham Young University while cadets in the second level of Spanish took only the Adaptive Listening Test due to the testing

budget. DFFL's proficiency expectation (i.e. benchmark) for cadets finishing their first year of Spanish is Novice-Mid to Novice-High and Intermediate-Low for those completing their second year of Spanish.

Participants

Seventy-five students in first-year Spanish (Spanish 132) and second-year Spanish (Spanish 222) participated in this study. The mean age of participants in the first year of Spanish (n=33), was 18.88 (SD=0.33). Females (82%) outnumbered males, and the majority of the participants reported being either Caucasian (67%) or Latinx (33%). All participants reported that they learned most and/or all of their Spanish (M=2.5 years of study) through the U.S. educational system prior to matriculating at USAFA, while only two participants reported that some members of their family spoke Spanish at home and/or with extended family on a regular basis. The participants reported that the last Spanish class they took, on average, was two years prior to enrolling at USAFA. No participants reported having dual enrollment (college) credit for Spanish.

For participants in the second year of Spanish study (n=42), the mean age was 19.95 (SD=1.14). The number of females was equal to the number of males (50%), and the majority of the participants reported being Caucasian (69%) or Latinx (29%). Two percent of the sample reported being African American. Like the first-year Spanish group, most reported having learned most and/or all of their Spanish (M=2.5 years of study) through the U.S. educational system prior to coming to USAFA. Again, none of the participants reported having dual enrollment credit for Spanish. The cadets in the second year of Spanish were a mix of first-year cadets who had tested into the second year of Spanish and second year cadets who had passed through the first year of Spanish at USAFA.

Procedures

After obtaining Institutional Review Board approval for human subjects testing in April 2021, two DFFL Spanish professors volunteered four of their classes to participate in a baseline study of cadet proficiency in Spanish. Two of the classes were ending their first year of Spanish study at USAFA, and the other two classes were about to complete their second year of Spanish study. The Director of the DFFL Language Lab administered the listening and reading proficiency tests in the departments' language lab in late April 2021. Results from the tests were sent electronically and securely to the DFFL Director of Assessment, who forwarded the results to the two Spanish professors. Data collection ended in early May 2021 and data were analyzed using *SPSS 18*.

Instruments: Adaptive Reading Test and Adaptive Listening Test

The Adaptive Reading Test and Adaptive Listening Tests are computer adaptive, criterion-referenced tests of an individual's reading and listening proficiency, respectively. Because they are adaptive, the number of items to which individual test takers respond will vary, depending on performance. Test items are drawn from item pools at specified proficiency levels. Results can be used for multiple purposes including placement of higher education students in an appropriate course, measuring proficiency or learning gains (pre and post-tests), guiding instruction, or informing program evaluation. Results from these two assessments are reported according to the ACTFL Proficiency Guidelines for Novice, Intermediate, Advanced, and Superior (for specified tests) language abilities and are currently available in Arabic, Chinese, English, French, Russian, Spanish, and Turkish (ACTFL, 2012). Note that they are not official ACTFL tests.

In order to develop these tests, language subject matter experts and assessment

professionals aligned the texts, passages, and items with the criteria described in the ACTFL

Proficiency Guidelines (ACTFL, 2012). Item development began with the selection of authentic

texts and passages from real-world sources across a range of different fields. The item writing

process included training item writers to create items that were aligned with the ACTFL

Proficiency Guidelines for each text or passage. Upon being developed, "the test development

team reviewed the items for alignment with the targeted proficiency level and trial with a small

representative sample of examinees" (Clifford & Cox, 2013, p. 52). Poorly functioning items

were either revised and retested or removed altogether from the item development pool. The final

step in the process was empirical testing of the items to determine whether their statistical

difficulties clustered by levels on the ACTFL Proficiency Guidelines (e.g., Intermediate-Mid,

Advanced-Low). For the empirical testing portion of the development of the tests, the authors

calculated Rasch person reliability coefficients for the tests and the items because it differentiates

between people with higher abilities compared to people with lower abilities (Schumacker,

2016).

The Adaptive Reading Test includes up to 57 items: a maximum of 24 at the Intermediate

level and a maximum 33 at the Advanced level. The authors reported a 0.80 Rasch person

reliability coefficient, indicating a relatively high level of internal consistency. Item reliability is

very high (0.98), which indicates that the items function at distinctively separate levels of

difficulty. The developers of the test reported that they conducted an independent samples *t-test*

between the Intermediate and Advanced items and determined that the two groups of items

indeed differed in terms of item difficulty (Clifford & Cox, 2013).

The Adaptive Listening Test includes up to 74 items: a maximum of 35 at the

Intermediate level and a maximum of 39 at the Advanced level. Much like the Reading

Proficiency Test, a 0.85 Rasch person reliability coefficient was reported, again indicating a relatively high level of internal consistency. Item reliability measures are strong (0. 97), signifying that the items function at distinctively separate levels of difficulty. An independent samples *t-test* between the Intermediate and Advanced items and revealed that the two groups of items differed in terms of item difficulty (Cox & Clifford, 2014).

Results

The researchers collected baseline-testing data on cadets studying first and second-year Spanish at USAFA in the spring of 2021. Means to describe proficiency on the ACTFL Proficiency Scale were determined by labeling each level in a nominal sequence (e.g., Novice-Low = 1, Novice-Mid = 2).

Listening Proficiency Attained at the End of the First and Second Year of Study

With respect to the first research question about the level of proficiency attained by cadets in the first and second-year of Spanish study at USAFA using the Adaptive Listening test, as Table 1 shows, cadets in second-year Spanish showed greater listening proficiency overall than their counterparts in first-year Spanish. On average, first-year cadets earned a proficiency level of Novice-High in listening (*M*=3.06) while second-year cadets earned, on average, a score midway between Intermediate-Low and Intermediate-Mid (*M*=4.68).

Table 1

Adaptive Listening Test Results

Proficiency Rating	End of First-Year Spanish (Spanish 132)	End of Second-Year Spanish (Spanish 222)
Novice-Low	4	0
Novice-Mid	7	1
Novice-High	11	9
Intermediate-Low	5	7
Intermediate-Mid	6	11

Intermediate-High	0	10
Advanced-Low	0	1
Advanced-Mid	0	1
Advanced-High	0	0
Total (*N*)	33	40

The benchmark levels established for first-year Spanish for USAFA is Novice-Mid to

Novice- High (between 2 and 3), and the average score was at the Novice-High level; 87% of

participants earned at least the benchmark level. Similarly, the benchmark established for 222

was either met or exceeded for 39/40 (98%) at or above benchmark of Novice-High to

Intermediate-Mid. Only one cadet did not attain the benchmark level, while 68% were in the

benchmark range and 30% exceeded the benchmark range. Of the 42 participants in second-year

Spanish, two did not receive a proficiency rating.

Reading Proficiency Attained at the End of the First Year of Study

With respect to the second research question regarding the level of proficiency cadets

attained in reading near the end of the first year of Spanish study, Table 2 shows that while 33

first-year cadets took the reading test, two did not receive a score; therefore, the authors can only

report 31 participants in the results. On average, the first-year learners received a score between

Novice-Mid and Novice High (*M*=2.6), which indicated that 87% showed proficiency at or above

benchmark of Novice-Mid to Novice-High.

Table 2

Adaptive Reading Test Results at End of First Year of Spanish Study

Proficiency Rating	*N*
Novice-Low	4
Novice-Mid	10
Novice-High	10
Intermediate-Low	3
Intermediate-Mid	3
Intermediate-High	1
Advanced-Low	0

Advanced-Mid	0
Advanced-High	0
Total (*N*)	31

Characteristics of those at the Highest and Lowest Levels of Proficiency in Reading and

Listening

         Gender. Turning to the third research question about the characteristics of students who

attained the highest and lowest levels of proficiency in reading near the end of their first-year of

Spanish study at USAFA, initial data analysis showed that 33 cadets participated in study and 31

received proficiency ratings. The females in the group showed scores of a mean proficiency of

2.33 (Novice-Mid) while the scores for the males were slightly higher yet still in the Novice-Mid

range ($M=2.84$). Fifty percent of the females in the sample scored at the benchmark rating of

Novice-High and Intermediate-Low whereas slightly more (61%) of the males scored at the

benchmark rating of Novice-High rating or above (Table 3).

Table 3

Adaptive Reading Test Results at the End of First Year of Spanish Study by Gender

Proficiency Rating	Females	Males
Novice-Low	1	3
Novice-Mid	1	9
Novice-High	2	8
Intermediate-Low	1	2
Intermediate-Mid	0	3
Intermediate-High	0	1
Advanced-Low	0	0
Advanced-Mid	0	0
Advanced-High	0	0
Total (*N*)	5	26

         Results of the Adaptive Listening Test results by gender for both groups showed that both

females and males in Spanish 132 scored on average at the Novice-High level ($M=3.17$ and

M=3.04, respectively), which was again at the benchmark set by the Spanish faculty.

Years of previous study. In reviewing the proficiency ratings according to years of study

of Spanish prior to attending USAFA, it is important to note that some participants failed to

respond to the some of the requested demographic questions. Nevertheless, all of the participants

had taken either two or three years of Spanish previously. Table 4 shows that the participants at

the end of the first year of study at USAFA who reported having taken two years prior to

attending scored at the lower end of the scale (*M*=2.10, Novice-Mid). However, those who

reported having taken three years prior to attending USAFA scored higher (*M*=3.07), a rating

consistent with the lower end of the Novice-High rating. Taken collectively, the results show that

participants with both relatively low and high levels of reading proficiency had at least three

years of prior Spanish study.

Table 4

Adaptive Reading Test Results at the End of First Year of Spanish Study by Number of Years of

Spanish Study Prior to Attending USAFA

Proficiency Rating	1 year	2 years	3 years
Novice-Low	-	3	0
Novice-Mid	-	3	6
Novice-High	-	4	3
Intermediate-Low	-	0	3
Intermediate-Mid	-	0	2
Intermediate-High	-	0	0
Advanced-Low	-	0	0
Advanced-Mid	-	0	0
Advanced-High	-	0	0
Total (*N*)	-	10	14

Next, the researchers examined the relationship between the number of years studying

Spanish prior to attending USAFA for both levels as related to one's proficiency ranking on the

Adaptive Listening Tests. Table 5 shows that after two years of prior study, cadets at the first

year of study (Spanish 132) were at the higher end of the Novice-Mid benchmark rating

(*M*=2.81); yet, at the end of the second year (Spanish 222) scored at the Intermediate-Low level

(*M*=4.00). When examining the data for those cadets in first-year Spanish who reported having

studied Spanish for three years prior to matriculating at USAFA, their average rating was

Novice-High (*M*=3.35) compared to the second-year Spanish cadets who scored a rating of

Intermediate-Low level (*M*=4.85). None of the first-year cadets reported having taken four years

of Spanish prior to matriculation; however, those in the second year who studied Spanish for four

years prior to attending USAFA scored similarly to those who studied Spanish for three years

(Intermediate-Mid, *M*=4.74). Viewed collectively, the data show that most cadets had some

previous study of Spanish. Those with the highest levels of proficiency (Intermediate-Mid) in

first-year Spanish also had at least three years of prior study in Spanish. Similarly, those with the

highest levels of proficiency (Intermediate-High) in second-year Spanish had at least four years

of prior study in high school.

Table 5

Adaptive Listening Test results at the End of First and Second Years of Spanish Study by the

Number of Years of Spanish Study Prior to Attending USAFA

	1 Year		2 Years		3 Years		4 Years	
Proficiency Rating	132	222	132	222	132	222	132	222
Novice-Low	-	0	1	0	1	0	-	0
Novice-Mid	-	0	2	0	5	0	-	0
Novice-High	-	1	6	1	1	2	-	4
Intermediate-Low	-	0	2	1	2	0	-	3
Intermediate-Mid	-	0	0	1	5	2	-	5
Intermediate-High	-	1	0	0	0	3	-	5
Advanced-Low	-	0	0	0	0	0	-	0
Advanced-Mid	-	0	0	0	0	0	-	0
Advanced-High	-	0	0	0	0	0	-	0

Total (*N*)	-	2	11	3	14	7	0	17

Quartile of proficiency. The next research question focused on specific characteristics of participants in the highest and lowest proficiency quartiles of the tests. The researchers examined the data and compared two groups for those in the first year and those in the second year of study of Spanish. The participants were divided into three groups: those who demonstrate proficiency at the highest, mid and lowest levels. This section explores comparisons between those who scored on the lower end of the proficiency scale and those who performed at a higher level on the scale. The middle group was not examined for comparative purposes.

First-year Spanish results. With respect to results from the BYU Adaptive Listening Test, the lower group scored at the Novice-Low and Novice-Mid levels (*n*=11). Demographically, most (92%) self-reported as male and Caucasian (73%), while all of the participants in this group reported not having dual enrollment credit. Fifty-five percent reported taking at least three years of Spanish prior to attending USAFA while only one participant in this group reported speaking Spanish at home with family members. Ten of the 11 reported that they learned Spanish via the U.S. educational system. Similarly, those in the high achieving group in first-year Spanish scored at the Intermediate-Low and Intermediate-Mid levels (*n*=11); this group were mostly males (82%) and either Caucasian (64%) or Latinx (36%). Two reported speaking Spanish at home with family members. Nearly all (91%) learned Spanish in the U.S. educational system and 82% of the high achieving group took at least 3 or 4 years of Spanish in high school.

Next, the researchers examined the highest and lowest achieving students in first-year Spanish on the BYU Adaptive Spanish Reading Test. Data analysis showed similar results as those for the BYU Adaptive Listening Tests described above. For the lowest achieving group, which included the Novice-Low and Novice-Mid levels (*n*=14), most self-reported as males

(86%), and either Caucasian (71%) or Latinx (36%). Almost half of the participants reported having taken only two years of Spanish in high school. None reported having dual enrollment credit, being heritage speakers of the language, or having any overseas experience using Spanish. Those who scored at the higher end of the proficiency scale (Intermediate-Low and Intermediate-Mid levels, (*n*=7) were mostly males (86%) who were Caucasian (100%) and had taken at least three years of Spanish in high school via the US educational system (100%). The participant who scored the highest on the test (Intermediate-High) reported taking five years of Spanish prior to attending USAFA.

   Second-year Spanish results. Turning to the results from those in the second-year of Spanish at USAFA, similar comparisons were made for the BYU Adaptive Spanish Listening Test. The lower group consisted of those who scored at the Novice-Mid, Novice-High, and Intermediate-Low levels (*N*=17). Participants in the higher group scored at the Intermediate-High, Advanced-Low, and Advanced-Mid levels on the proficiency scale (*N*=12). The demographics for the two groups were very similar. The majority were females in both groups (59%) with all but one having learned Spanish in the U.S. educational system. Forty-one percent of the lower group had taken four years of high school Spanish. Twenty-nine percent of the same group reported having taken their last Spanish class either their sophomore or junior year of high school. In the upper group, all had completed four years of high school Spanish and all but one had taken Spanish all four years in high school. The more recently and the more courses students took, the higher their proficiency levels. In other words, participants with the highest levels of proficiency had fewer interruptions to, in addition to more, Spanish language learning experience.

Comparison to Tschirner's Findings

With respect to the final research question regarding how the present study's findings compare to Tschirner's listening and reading outcomes, it is important to note that the population of this study, cadets at a military academy, were different from Tschirner's. Tschirner (2016) conducted a large study of the proficiency levels of college students enrolled in private and public institutions, with the majority coming from large public universities. He used the ACTFL Listening Proficiency and Reading Proficiency Tests administered by Language Testing International, an official ACTFL test and not the same test used in this study. Thus, while the results can be compared, the instruments are not identical. Table 6 shows how the results of this study compared to that of Tschirner's; it shows that, on average, USAFA's cadets attained a higher level of proficiency in listening than in Tschirner's study. Regarding Tschirner's reading outcomes compared to the present study, USAFA cadets scored at the higher end of the Novice-mid level (M=2.8) compared to Tschirner's participants, who scored at the lower end of the Novice-high level (M=3.11).

Table 6

Comparison of Tschirner's Listening Outcomes to Present Findings

Tschirner Second Semester	USAFA Second Semester	Tschirner Fourth Semester	USAFA Fourth Semester
2.05	3.06	2.83	4.7
Novice-Mid	Novice-High	Novice-High	Intermediate-Mid

Discussion

Establishing both rigorous and attainable outcomes for language learning sequences is
critical to supporting programs in developing strong curricula and measuring their outcomes.
While new data related to outcomes in four-year college language programs have emerged since
2016, there are little recent data on results from other types of programs. This study provides a
first step in establishing benchmarks in second and fourth semester Spanish language courses at a
military academy. Military academies are not only post-secondary institutions but also key
players in providing language background to directly and immediately support national security
and language endeavors.

The Department of Defense continues to place a premium on language and culture
enabled military personnel, and this report provides important data for this emphasis as well as
documentation of their success in this area. Consistent with previous Department of Defense
guidelines, cadets graduating from USAFA with a major in FAS or a minor in one of the eight
languages taught at USAFA are required to take the Defense Language Proficiency Test, which,
like this study, examines cadets' ability in the receptive skills (i.e. listening and reading). The
results of this study showed proficiency attained in one language (Spanish) at two levels in
listening (second and fourth semester) and at one level in reading (second semester). The data
showed that participants with previous study of Spanish in high school had higher scores than
those who had less high school study. However, there were not many differences with respect to
attained proficiency by gender. The study has implications both locally, for the specific
institution, for other military academies and for higher education in general by documenting
these outcomes to contribute to the existing body of work on student outcomes.

As previously noted, USAFA's cadets scored similarly to the undergraduates at public

and private universities in reading across the U.S. from Tschirner's (2016) study (approximately Novice-High). Glisan and Foltz (1998) focused on secondary school learners and oral proficiency outcomes; thus, the researchers cannot compare these results. Similarly, Carroll (1967) focused on language majors with more years of study than those in the present population. Because Tschirner's study includes not only Flagship reading and listening outcomes but also outcomes from additional post-secondary programs, the discussion will focus on comparisons between Tschirner's study and the present one.

Notably, USAFA's cadets scored much higher than those in Tschirner's study in Listening. There are a number of reasons that could account for this difference. First, the BYU test is not an official ACTFL test, as is Tschirner's and there may be differences between local interpretations and official ACTFL test items. Secondly, Tschirner had a much larger sample of a more diverse audience; thus, the USAFA sample may include more motivated students than Tschirner's. Finally, because so many cadets began Spanish language study with three or more years of prior study, they may have begun with higher levels of listening proficiency than those in Tschirner's study. Tschirner did not investigate number of years of prior study, so that comparison cannot be made.

Interestingly, the cadets scored slightly lower after two semesters in reading. It is possible that reading is emphasized less in the USAFA curriculum than in the programs included in Tschirner's. In addition, classes at USAFA are capped at 24; it possible that USAFA classes are smaller and more conducive to the development of listening than at the schools included in Tschirner's samples.

It is also important to highlight that cadets enrolled in a military academy may be different in their motivations and approaches to language learning than those at a four-year public colleges. First, approaches to teaching and learning at a military academy may be more homogenous than at

a large, public institution of higher education where many introductory courses are taught by teaching assistants and part-time faculty who are responsible for teaching and learning but may not have input into course design and development. By contrast, all courses at a military academy are taught by full-time professors who are required to collectively plan, design and implement curricula. Such homogeneity may result in different teaching and learning contexts. At the same time, there are no language majors at a military academy, so no participating cadets are able to pursue the language with the intensity of a university Spanish major.

This study adds to the body of work on proficiency outcomes in higher education and introduces a new but important subgroup: cadets at a military academy. Such students in higher education are well positioned to influence security and public policy within their careers and thus their inclusion in the general outcomes data provides both information to the field and incentives to the military academies to encourage language study and to document the results. On average, the cadets scored higher in listening than the students in Tschirner's 2016 study and slightly lower in reading. While the sample size of the present study is small, it represents an important effort in noting such outcomes.

Future research can both replicate this study and add more participants to determine how cadets' outcomes compare to other students enrolled in higher education. In addition, future studies could examine qualitatively why cadets score higher in listening than their counterparts at non-military schools, if such a trend continues. Conducting benchmark studies with the USAFA population of oral proficiency outcomes will allow for comparisons to other studies, such as Isbell, Winke and Gass (2018). In conclusion, the present study can also provide important information for curriculum development and new foci for continued improvement in the program. As language professionals, it is our duty to move our learners up the proficiency ladder.

By examining proficiency benchmarks using reliable and valid tests we will know where our learners are and what we need to do to continue building their proficiency in the target language.

References

ACTFL. (2012). *ACTFL proficiency guidelines 2012*. https://www.actfl.org/resources/actfl-proficiency-guidelines-2012

ACTFL. (2017). *NCSSFL-ACTFL can-do statements*. https://www.actfl.org/publications/guidelines-andmanuals/ncssfl-actfl-can-do-statements

ACTFL. (2019) *Making languages our business.* ACTFL. https://www.leadwithlanguages.org/wp-content/uploads/MakingLanguagesOurBusiness_FullReport.pdf

American Academy of Arts and Sciences (2017). America's languages: Investing in language education for the 21st century. AAAS. https://www.amacad.org/publication/americas-languages

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

Clifford, R., & Cox, T. L. (2013). Empirical validation of reading proficiency guidelines. *Foreign Language Annals, 46*(1), 45-61. DOI: 10.1111/flan.12033

Cox, T. L., & Clifford, R. (2014). Empirical validation of listening proficiency guidelines. *Foreign Language Annals, 47*(3), 379-403. DOI: 10.1111/flan.12096

DeKeyser, R. M. (2014). Research on language development during study abroad. In C. Perez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 313-326). John Benjamins. https://doi.org/10.1075/aals.13.16ch13

Dewey, D. P., Bown, J., & Eggett, D. (2012). Japanese language proficiency, social networking, and language use during study abroad: Learners' perspectives. *Canadian Modern*

Language Review, *68*(2), 111-137. DOI:10.3138/CMLR.68.2.111

Freed, B. F. (1995). Second language acquisition in a study abroad context. *Studies in*

   Bilingualism. https://doi.org/10.1075/sibil.9

Glisan, E. W., & Foltz, D. A. (1998). Assessing students' oral proficiency in an outcome-based

   curriculum: Student performance and teacher intuitions. *The Modern Language Journal*,

   82(1), 1-18. https://doi.org/328680

Hernández, T. A. (2010). Promoting speaking proficiency through motivation and interaction:

   The study abroad and classroom learning contexts. *Foreign Language Annals*, *43*(4), 650-

   670. DOI: 10.1111/j.1944-9720.2010.01107.x

Isbell, D. R., Winke, P., & Gass, S. M. (2018). 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

The Language Flagship. (2013). *About us*. Available from,

   https://www.thelanguageflagship.org/content/about-us

Looney, D., & Lusin, N. (2018, February). *Enrollments in languages other than English in*

   United States institutions of higher education, Summer 2016 and Fall 2016: Preliminary

   Report. Modern Language Association.

   https://www.mla.org/content/download/83540/2197676/2016-Enrollments-Short-

   Report.pdf

Malone, M. E. (2019). Afterword and next steps. In P. Winke & S. M. Gass (Eds.), *Foreign*

   language proficiency in higher education (pp. 318). Springer. DOI:10.1007/978-3-030-

   01006-5_16

The National Standards Collaborative Board. (2015). *World-readiness standards for learning*

languages. ACTFL. https://www.actfl.org/resources/world-readiness-standards-learning-languages

Schumacker, R.E. (2016). *A beginner's guide to structural equation modeling*. Routledge. https://doi.org/10.1080/10705511.2017.1280798

Tschirner, E. (2016). Listening and reading proficiency levels of college students. *Foreign Language Annals*, *49*(2), 201-223. DOI: 10.1111/flan.12198

Vande Berg, M. V., Connor-Linton, J., & Paige, R. M. (2009). The Georgetown consortium project: Interventions for student learning abroad. *Frontiers: The interdisciplinary journal of study abroad*, *18*(1), 1-75. https://files.eric.ed.gov/fulltext/EJ883690.pdf

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