

ScholarWorks@GSU

Student Proficiency Growth in Spanish: Advancements at the U.S. Air Force Academy's World Language Program

Item Type	Book chapter
Authors	Swanson, Peter B.;LeLoup, Jean W.;Peltier, Jean-Philippe
Citation	Swanson, P., LeLoup, J.W., & Peltier, J.P. (2025). Student proficiency growth in Spanish: Advancements at the U.S. air force academy's world language program. In J.R. Watson, R.L. Wolfel, & A.J. Kalkstein (Eds.), Language, regional expertise, and culture (LREC): State of the science (pp. 13-35). New York: West Point Press.
Download date	2026-04-11 17:11:25
Link to Item	https://hdl.handle.net/20.500.14694/15813

CHAPTER 2

Student Proficiency Growth in Spanish

Advancements at the U.S. Air Force Academy's World Language Program

Pete Swanson, PhD, U.S. Air Force Academy

Jean W. LeLoup, PhD, U.S. Air Force Academy

Jean-Philippe Peltier, PhD, Colonel, U.S. Air Force Academy

Abstract

The paradigm of the teaching and learning of world languages has been significantly impacted by Communicative Language Teaching approaches. An increased emphasis on the proficiency testing of language learners is now prioritized where program coordinators and administrators have set proficiency benchmarks for language learners to achieve at all levels of instruction. Established at the turn of the twenty-first century, the Language Flagship program—a national effort to change the way Americans learn languages—calls for institutions of higher education to create a “viable process to assess proficiency learning in high quality, well-established academic language programs” (Swanson et al. 2). Perhaps in line with the national movement, the Secretary of the United States Air Force stated that the Language Enabled Airman Program—a congressionally mandated program—serves to “[I]ncrease the language inventory from within the force . . . we can’t contract this ability out to non-warfighters” (Chesser 3). In 2020, leadership at the United States Air Force Academy directed faculty to create Language Roadmaps to Proficiency, where proficiency benchmarks were developed for each of the eight languages taught at the Academy. Afterward, a comprehensive proficiency testing regime was set into place. In 2021, students enrolled in first- and second-year Spanish took the ACTFL Proficiency Placement Tests in fall 2021 and again in spring 2022 to gauge potential gains in Spanish proficiency. Students showed impressive gains in both reading and listening modalities. Results have implications for instructors, program directors, language learners, and language curricula.

KEYWORDS: communicative language teaching, language assessment, language benchmarks, Language Enabled Airman Program, language flagship program, foreign language proficiency, language proficiency testing, second language acquisition

Introduction

The question regarding reasonable expectations of language proficiency for students to attain after a specific learning sequence of language study has challenged the field of language teaching and learning for decades (Swanson et al. 2). In the 1960s, Carroll reported that educational stakeholders (e.g., instructors, program directors, administrators) have struggled to establish reasonable proficiency benchmarks, communicate them to language learners and faculty, and attain them after various sequences of study (e.g., first year, second year) (Carroll 131–132).

Today in the United States of America, having world language (WL) skills is crucial for America’s diplomatic, business, and national security interests (La Corte and Voisine 3). The American Council on the Teaching of Foreign Languages (ACTFL) stated that 90% of businesses surveyed reported a need for employees with skills in languages other than English (ACTFL, “Making Languages” 15). Additionally, the ACTFL report indicated that the global economy would continue to grow, further emphasizing the vital importance of language proficiency in the public sphere (15). Unfortunately, despite this specified need in business, “the vast majority of American citizens remain monolingual” (American Academy of Arts and Sciences vii), and unfortunately, there has been a serious decrease in the enrollment of students taking WLs other than English. Prior to 2009, surveys showed sustained growth; however, there was a 29.3% decrease in enrollment between 2009 and 2021 (Lusin et al. 4).

While students studying WLs do so for a variety of reasons (e.g., employment opportunities, travel), it is important to note that not all language learners acquire language at the same rate (McLaughlin 7). Thus, there is a need to set benchmarks for WL proficiency and help learners move along their interlanguage continuum, gaining increasingly higher levels of proficiency as they progress through learning sequences (e.g., first year, second year). Setting proficiency benchmarks allows for the understanding of what can be and is attained after specific sequences of study. Without such data, educational stakeholders (e.g., students, instructors, administrators) “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” (Swanson et al. 2). Moreover, the lack of such knowledge may cause programs to set benchmarks that are either too ambitious or too low for their language learners. With respect to the current study, the researchers investigated the baseline proficiency of students of Spanish at the United States Air Force Academy (USAFA) during the first two days of instruction and again near the end of the two-semester sequence for both first-year and second-year language learners.

Literature Review

The seminal investigation by Carroll marked the first major exploration of student outcomes in contemporary WLs. Despite its more than five-decade vintage, Carroll’s study remains noteworthy, delving into the realms of speaking, reading, and listening proficiency across five languages (French, Italian, German, Spanish, and Russian) from various U.S. universities. The study’s scope extended beyond mere language assessment, incorporating an analysis of factors influencing student outcomes such as gender, age, prior language study duration, overseas experience, and current university year.

Carroll employed the Modern Language Association test, aligning it with the then-novel Interagency Language Roundtable (ILR) scale, a choice dictated by the scale’s recent adoption in government circles. Significantly, the study predated the ACTFL Proficiency Guidelines, currently ubiquitous in academic and business contexts. Carroll’s groundbreaking work not only surveyed language outcomes expansively but also introduced the use of the ILR scale, enabling future research aligning with the subsequently developed ACTFL Guidelines. See Appendix A for a comparison of the ILR and ACTFL scales.

Findings from Carroll’s investigation indicated that students majoring in French, German, Italian, Russian, and Spanish achieved an average ILR 2+ proficiency level (141–51), approximately equivalent to ACTFL Advanced-Mid or Advanced-High. Factors correlating with higher proficiency levels included heritage language background, study abroad experiences, elementary school language study, language study at larger institutions, and no discernible difference between genders.

Post-Carroll, comprehensive research on general language proficiency outcomes in higher education was notably limited for nearly half a century, exacerbated by the evolving landscape of WL study in universities. The original languages highlighted in Carroll’s study were no longer the only focus of WL study in higher education.

While these languages were still in the top 20 languages in higher education, they were subsequently joined and, in some cases, replaced by enrollments in American Sign Language, Japanese, Chinese, and Arabic (Looney and Lusin 4–5). Carroll’s initial study provided essential but increasingly outdated information for decades as research on outcomes in higher education became more specialized (focusing on specific factors) and less general.

Recognizing this gap, the Flagship Initiative in 2014 launched a call for proposals to investigate student outcomes in various languages across three state universities in the United States. The resulting three-year grant period witnessed approximately 9,000 students undergoing language proficiency tests, offering diverse insights across reading, listening, and speaking in languages such as Arabic, Chinese (Mandarin), French, German, Korean, Portuguese, Russian, and Spanish (Winke et al. 94–95). As a result, subsequent studies began to zero in on specific language skills and particular factors affecting proficiency outcomes (such as language learning venue). For example, listening was identified as one of the least-researched or most under-researched aspects of assessment (Harding et al. 326). More specifically, listening fluency and extensive listening have suffered a lack of attention in research studies (Chang et al. 423). Other research suggests that listening skills can be improved by focusing on reading (Jiang et al. 1160), and extensive reading and listening may have a reciprocal positive effect (Renandya and Jacobs 12). Because of its important source of target language input, listening is a primary skill needed to develop language proficiency. But assessments must focus on the learner’s ability to comprehend the specific type of oral input/domain being evaluated (Wagner 231).

Isbell and Winke advocated for using the “ACTFL Speaking Assessment: Oral Proficiency Interview—Computer” (ACTFL Speaking Assessment: OPIc) to assess proficiency and monitor language learning progress at the tertiary level. Several studies have delved into specific factors affecting outcomes, particularly in the context of study abroad (see, for instance, DeKeyser; Dewey et al.; Freed; Hernández; and Vande Berg et al.). At times these studies show conflicting results, underscoring the challenges to research in this area: the differences in measurement of quantity and quality of target language use, the target language proficiency level of learners entering study abroad programs, lack of longitudinal accounts of target language spoken while abroad, and the great differentiation/variation among study abroad programs (see, for instance, Di Silvio et al.; Isabelli-García et al.; McManus et al.; and Tullock and Ortega).

Despite these advances, the body of research emerging from the Flagship Initiative, while substantial, represents just a fraction of the comprehensive studies needed. Malone emphasized the necessity of further research that encompasses a wider spectrum of outcomes under varying institutional conditions, highlighting the contrast between Carroll's findings from large institutions and the Flagship-funded research conducted in three large public universities (317). Tschirner added depth to the discourse with a comprehensive report on ACTFL reading and listening tests, incorporating participants from the Flagship study. With a participant pool exceeding 6,000, Tschirner delineated average outcomes after two, three, four, five, and six semesters of study (Tschirner 201–23). Despite this progress, more research is essential to gauge outcomes in diverse learning environments, with Tschirner's data and the Flagship project outcomes serving as valuable benchmarks.

While research studies that assess student language proficiency outcomes are important contributions to the overall second language (L2) acquisition database, a further—and perhaps more germane—step needs to be taken to give meaning to these research results. Greatly missing in the quest to evaluate proficiency and ascertain if benchmarking goals are being met is the delineation of the underlying reason(s) for said progress. If the goal is proficiency, it is necessary not only to assess student progress but also to evaluate program curricula to ascertain that the most effective teaching practices are being integrated into classroom instruction to achieve student progress on the proficiency continuum (Soneson and Tarone 51). In other words, student progress toward proficiency goals can be assessed, but this does not verify the underlying reasons/causes of said progression. One way is to set initial proficiency benchmarks for courses and then assess student progress toward them. The best approach would be to develop and apply a systematic plan of assessment taking into consideration student progress across several semesters using demographic data. Such data could be used to determine students' baseline proficiency and then to ascertain how the students have increased their proficiency ratings over time. The next step, then, is to analyze and observe programmatic components that have contributed in a meaningful way. Proficiency benchmarking is an ongoing process, fluid by nature given the final goal of developing the optimal WL curriculum and teaching practices to promote movement along the WL continuum.

In this context, the present study endeavors to contribute to the expanding body of knowledge by examining outcomes after two or four semesters of Spanish language study at USAFA. The cadets, representing a distinct subset of postsecondary students, have been underrepresented in language outcomes research. Given the broader implications of language proficiency in U.S. life, particularly in areas

like education, business, security, and diplomacy, understanding the language outcomes of future military leaders is paramount. The following research questions guide the present study:

1. What levels of proficiency in listening and in reading did the first-year Spanish participants attain?
2. What levels of proficiency in listening and in reading did the second-year Spanish participants attain?
3. Based on testing results, how accurate are the previously set proficiency benchmarks for first- and second-year Spanish?

Methods

The mission of the Department of Languages and Cultures (DFLC) at USAFA is to produce culturally attuned and linguistically capable Airmen. Its graduates deploy worldwide in support of U.S. strategic interests and engagements. The focus of the program is squarely on oral proficiency by employing Communicative Language Teaching approaches—a signature pedagogy in WL instruction. It is critical for U.S. Airmen to work with partners and allies in their respective languages and to have sufficient cultural knowledge to meaningfully engage and create lasting relationships. To accomplish this critical mission, it is imperative to continuously assess and ensure that USAFA’s language programs are meeting the needs of the U.S. Air Force. Faculty in DFCL teach eight languages: Arabic, Chinese, French, German, Japanese, Portuguese, Russian, and Spanish. Starting in Academic Year 2020–2021, the department began to standardize the language used for assessment. The eight DFCL language communities developed their individual Language Roadmaps for Proficiency, which are aligned with the “ACTFL Proficiency Guidelines” and the “NCSSFL-ACTFL Can-Do Statements.” This language is codified in the department’s roadmaps, with each language community having its own roadmap and desired proficiency outcomes.

This pedagogical alignment was used to set proficiency benchmarks for students at each level, which allowed faculty to set a foundation for comparison across its eight programs. Respective language communities are able to see how one program might aim for Novice-High after 160 hours of instruction, while another might set its sights on Intermediate-Low. Fundamentally, DFCL’s roadmaps were aligned with established national standards while enabling various language programs to compare, gain insight, and collaborate based on a mutually accepted foundation.

Prior to the development of the language roadmaps, the department relied on the Defense Language Proficiency Test “DLPT Relevant Information”—an online listening and reading assessment that examines one’s receptive skills in the target language—to assess learners’ L2 proficiency. The test was not without its deficiencies. First, The DLPT provided inadequate feedback to both faculty and learners, and it was not aligned with current or past standards. The faculty, rightfully, did not believe that the DLPT could be used as a dependable measure for each of the language community’s stated objectives. For example, learners who score a 2 on the ILR scale and feel disheartened do not understand that, given the number of instructional contact hours, a 2 in Japanese is laudable. There is no insight regarding scoring, and no feedback in terms of potential strengths and weakness. The learner gets a proficiency rating with little to no explanation.

Second, and not unique to the DLPT, this assessment only measures ability in the receptive skills. DFCL’s gold standard is oral proficiency—the ability to communicate orally with partners and allies. The long-term focus on receptive skills assessment impacted DFCL’s pedagogical approach and associated curricula, and consequently, oral assessment was not heavily weighted in course syllabi.

The first step to bridging this gap was adopting the Adaptive Reading Test and Adaptive Listening Test “ACTFL Proficiency Placement Test.” These assessments are both clearly tied to the ACTFL Proficiency Guidelines. The use of these tests allowed DFCL to assess all language programs and provide individual language learners with targeted feedback based on their results. Starting with the 2021–2022 academic year, through random sampling, DFCL implemented a regular testing cycle across all levels of the eight language programs to ensure that each language community was meeting its clearly defined goals as articulated in its roadmap. These tests, however, while a marked improvement over the DLPT, shared a similar deficiency: they lacked assessment of the productive skill of speaking.

To address the need for greater emphasis on oral proficiency, the department developed a new policy. Any learner that scored the equivalent of Intermediate in one or both of the reading or listening portion of either the DLPT or Adaptive tests was eligible to take the Oral Proficiency Interview (OPI), which is “a live, 15- to 30-minute telephone conversation between a certified ACTFL Tester and the candidate,” “ACTFL Speaking Assessment: The Oral Proficiency Interview” [OPI] 1. Ideally, the OPI would be administered to each graduating student, but the cost of testing restricts wider dissemination. The administration of the OPI serves as an effective reminder that oral

production is the department's focus while also acting as an incentive among learners who want to qualify to take the OPI.

While USAFA does not offer a language major, it does offer language minors and a Bachelor of Science in Foreign Area Studies. Cadets selecting the major can choose a language, a world region, and a specific academic discipline (e.g., French, Africa or, History). In terms of minors, approximately 240 students graduate annually with a WL minor (Carriedo). All first-year students are required to study a foreign language during their initial year at USAFA—two semesters, or 160 hours. All students take the DFLC language placement test during basic training. They can test out of the language requirement with Advanced Placement exam scores or via the department's placement test. Based on the results, they can validate one semester or the full year; they can also test into a higher level. Students who place into higher levels include those with substantial school-based or heritage language experience. Therefore, these students show a wide range of language backgrounds, not dissimilar to their counterparts at more traditional institutions of higher education.

Study Context

USAFA is a unique institution of higher education. It is a four-year military academy that prepares the next generation of Air Force officers. As part of the core curriculum, cadets must take two semesters of the same language or validate the credit (see above). At the first year or beginning level, cadets take WL classes every day Monday through Friday instead of a typical Monday-Wednesday-Friday or Tuesday-Thursday collegiate offering. First-year cadets studying WLs must attend classes each day in order to complete an 80-day class semester. WL classes at the second-year level and above meet every other day for a total of 40 days per semester. In addition to attending academic classes, the cadets must participate in military and physical training activities and are evaluated on performance each semester. Additionally, cadets are excused from classes for a variety of purposes (e.g., flight physicals). The first six days of class cadets may drop or add classes. At the end of each academic year, cadets are randomly selected to participate in proficiency tests to gauge progress.

USAFA Spanish instructors set proficiency benchmarks in the Spanish Language Roadmap to Proficiency for each level to attain by the end of the academic year. With respect to the present study, the benchmark proficiency level for first-year

Spanish for USAFA is set as a band from Novice-Mid to Novice-High. Similarly, the benchmark proficiency band for second-year Spanish is from Novice-High to Intermediate-Mid.

Procedures

After obtaining Institutional Review Board approval for human subjects testing, 58 cadets enrolled in first- and second-year Spanish classes were randomly selected to take two adaptive tests (see below for a detailed description) in the department's language lab the second and third day of classes of Fall 2021. Approximately nine months later, the same 58 cadets took the two adaptive tests at the end of the second semester in late April 2022. The Director of the Language Lab proctored both test administrations and sent the results electronically to the researchers. Data collection ended in early May 2022 and data were analyzed using IBM SPSS Statistics 29 (SPSS 29).

Instruments

The ACTFL Proficiency Placement Test (APPTI) consists of two parts: reading and listening. These two inexpensive assessments represent the interpretive mode of communication, as described in the *World-Readiness Standards for Learning Languages* (National Standards Collaborative Board). Both tests assess test takers' proficiency between Novice-Low and Advanced-Low. Each item consists of either a genuine reading text or audio passage and one multiple-choice question with one correct answer associated with the text or passage. Each test can last up to 60 minutes. However, instead of a time limit to take the entire test, a time limit has been set for each test item. Test items targeted at the Intermediate and Advanced levels have a time limit of 75 and 120 seconds, respectively. The test is computer-adaptive, and each skill (i.e., reading and listening) presents 10 to 25 items depending on the ability of the test taker. Both tests are computer adaptive, which means the number of items individual test takers respond to will vary, depending on performance. Test items are drawn from item pools at specified proficiency levels assessing a broad range of topics including everyday life, current events, and education, among others. As test takers begin to fail at a certain proficiency level, the test concludes.

Once a test is completed, a floor rating (the level at which the test taker has demonstrated sustained performance) and a ceiling rating (the level at which the examinee has demonstrated patterns of breakdown) is computed "ACTFL

Proficiency Placement Test.” Test takers receive separate ratings for reading and for listening. Results from the APPT can be helpful for a multitude of purposes such as the placement of higher education students in an appropriate course, measuring proficiency at certain points of the curricula, and informing program evaluation. The APPT ratings are from the ACTFL Proficiency Guidelines as language subject matter experts and assessment professionals align the texts, passages, and items with the criteria described in the ACTFL Proficiency Guidelines. With respect to the Spanish tests, the Adaptive Reading Test includes 57 items: 24 at the Intermediate level and 33 at the Advanced level. The developers reported a 0.80 Rasch person reliability coefficient, indicating a relatively high level of internal consistency (Clifford and Cox 57, 390). Item reliability is strong (0.98), indicating that the items function at distinctive levels of difficulty (390). Clifford and Cox 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. The Adaptive Listening Test includes 74 items: 35 at the Intermediate level and 39 at the Advanced level (390). Like the Reading Proficiency Test, a 0.85 Rasch person reliability coefficient was reported, indicating a relatively high level of internal consistency. Item reliability analysis revealed a strong coefficient (0.97) (390), which signifies that the items function at separate levels of difficulty. Cox and Clifford conducted independent samples t-test between the Intermediate and Advanced items and reported that the two groups of items differed in terms of item difficulty (53).

Participants

As mentioned earlier, a total of 58 participants in the first two years of Spanish offered at USAFA were randomly selected for the present study. All of the participants took the APPT in both modalities (reading and listening) in August 2022 on the second and third day of the academic semester and then again in May 2023 about one week before the semester ended.

With respect to participants in first-year Spanish ($n = 44$), the majority self-reported as males ($n = 33$), white/Caucasian ($n = 30$) with 6 Latino/a, 3 African American, and 5 Asian, and the mean age was 19.5 years at the second administration of the tests. Turning to the 14 participants in second year Spanish, the majority self-reported as males ($n = 9$), white/Caucasian ($n = 10$) with 2 Latinos, and 2 Asian, and the average age was 20.2 years at the second administration of the tests. The sample’s demographics are representative of USAFA in general (USAFA).

Viewed collectively, all of the participants reported previous experience in Spanish, obtaining all or most of their language ability through the U.S. educational system ($M = 2.40$ years of study) prior to matriculating at USAFA. The participants indicated that the last Spanish class they took, on average, was two years prior to enrolling at USAFA. None of the participants reported having dual enrollment (college) credit for Spanish.

Findings

After each of the two administrations of the APPT for both first- and second-year Spanish participants, the first author received the APPT data, which contained both floor and ceiling ratings, and entered them in SPSS 29 for data analysis. The APPT proficiency ratings are categorical variables, sometimes referred to as nominal variables (a variable that has two or more categories). As described in the ACTFL Proficiency Guidelines, the proficiency scale ranges from Novice to Distinguished. The first three levels, Novice, Intermediate, and Advanced, have three sublevels each: low, mid, and high. For example, a person taking the APPT may receive a proficiency rating of Novice-High, Intermediate-Low, and so forth, which are categorical variables. Given that they are not continuous variables (variables that can be assigned a value within a range), and considering the small number of participants in the study, inferential statistics (e.g., correlation, regression) were not calculated. Thus, frequencies of proficiency ratings are reported here for the floor rating only because it is the proficiency level at which the test taker demonstrated sustained performance.

With respect to the first research question about the levels of proficiency attained in the two modalities by the first-year Spanish participants, Table 2.1 reflects the findings at the pretest (start of the academic year) and posttest (end of the academic year) for both listening and reading. Regarding proficiency ratings for listening comprehension at the onset of the academic year, almost three quarters (73%, $n = 32$) of the participants received proficiency ratings in the Novice range. Twenty-five percent ($n = 11$) garnered proficiency ratings in the Intermediate range, and one student received a rating of Advanced-Low. However, at the end of the two semesters of Spanish classes in listening comprehension, the number of participants testing at the Novice range decreased by over half to 30% ($n = 13$), and the number of participants who received a rating in the Intermediate range almost tripled (70%, $n = 31$). There were not any participants at the Advanced-Low level. The data suggest that participants increased their proficiency in listening comprehension over the academic year.

TABLE 2.1
Pretest and Posttest Proficiency Ratings for First-Year Spanish Cadets in Listening and Reading

	LISTENING		READING	
	Pretest	Posttest	Pretest	Posttest
Novice-Low	4	1	12	3
Novice-Mid	8	1	6	3
Novice-High	20	11	17	18
Intermediate-Low	6	11	4	6
Intermediate-Mid	5	16	2	2
Intermediate-High	0	4	2	8
Advanced-Low	1	0	0	3
Advanced-Mid	0	0	1	1
Advanced-High	0	0	0	0
n	44	44	44	44

Source: U.S. Military Academy, Department of Languages and Culture

With respect to the modality of reading for the first-year Spanish participants at the start of the academic year, Table 2.1 shows that 80% of the participants received a rating in the Novice range ($n = 35$) while 18% were in the Intermediate range ($n = 8$). One participant received a rating of Advanced-Mid in reading. At the end of the two semesters of Spanish classes with respect to listening comprehension, 24 participants (55%) tested at the Novice range, 16 participants (36%) received a rating in the Intermediate range, and 4 participants (9%) received a rating at the Advanced levels, indicating that participants moved up the proficiency ladder in reading comprehension over the course of the two semesters.

Next, the researchers examined participants' proficiency growth individually. Inspection of the listening comprehension data revealed that 68% of the participants ($n = 30$) increased in proficiency by at least one proficiency sublevel or higher during the academic year, while 25% of the participants ($n = 11$) remained at the same sublevel. Three of the participants (7%) decreased two sublevels. Analysis of the reading comprehension data showed that 64% of the participants ($n = 28$) increased at least one proficiency level higher during the academic year, 32% ($n = 14$) remained at the same sublevel, and two participants (5%) decreased one and three sublevels respectively from the Intermediate level to the Novice level, suggesting that one of the participants may not have taken the APPT seriously during the second administration at the end of the academic year.

Turning to the second research question about the levels of proficiency in listening and in reading the second-year Spanish participants attained, Table 2.2 shows that 29% of the participants were in the Novice range ($n = 4$) and 71% were in the Intermediate range ($n = 10$) in listening at the start of the academic year. However, at the end of the two semesters of Spanish classes, one participant (7%) tested at the Novice range, 12 participants (86%) received a rating in the Intermediate range, and one participant (7%) received a rating of Advanced-Low.

TABLE 2.2
Pretest and Posttest Proficiency Ratings for Second-Year Spanish Cadets in Listening and Reading

	LISTENING		READING	
	Pretest	Posttest	Pretest	Posttest
Novice-Low	0	0	0	0
Novice-Mid	1	0	1	0
Novice-High	3	1	6	5
Intermediate-Low	6	1	1	0
Intermediate-Mid	3	5	1	1
Intermediate-High	1	6	4	5
Advanced-Low	0	1	1	1
Advanced-Mid	0	0	0	2
Advanced-High	0	0	0	0
n	14	14	14	14

With respect to the modality of reading for the second-year Spanish participants, Table 2.2 shows that exactly half of the participants (50%) were in the Novice range ($n = 7$), 43% were in the Intermediate range ($n = 6$), and one participant received a rating of Advanced-Low in reading at the start of the academic year. At the end of the two semesters of Spanish classes, five participants (36%) tested at the Novice range, 6 participants ($n = 43%$) received a rating in the Intermediate range, and three participants (21%) received a rating of Advanced-Low.

Closer inspection of the listening data indicated that nine participants (64%) increased at least one proficiency sublevel or higher during the academic year, while three participants (21%) remained at the same sublevel and two participants (14%) decreased only one sublevel. Analysis of the reading data indicated that 50% of the participants increased at least one proficiency level or higher during the academic year, 43% ($n = 6$) remained at the same sublevel, and one first-year participant decreased three sublevels from Intermediate-Low to Novice-Low (see

Table 2.1),, again suggesting that this single participant may not have taken the APPT seriously during the second administration at the end of the academic year.

Finally, with respect to the third research question about accuracy of proficiency benchmarks set by the department, it is important to note that the benchmark proficiency level for first-year Spanish for USAFA was set as a band from Novice-Mid to Novice-High, and the benchmark proficiency band for second-year Spanish was from Novice-High to Intermediate-Mid. With respect to first-year Spanish listening comprehension, only one participant did not reach the lower end of the proficiency benchmark, Novice-Low, at the end of the academic year. Thus, 97% of the cadets met or exceeded the benchmark. Similar findings were found for reading comprehension, where three participants did not reach the Novice-Mid benchmark, yet 41 of the 44 total participants either met or exceeded the benchmark.

Turning to the second-year participants and their proficiency ratings compared to the proficiency benchmark band, Novice-High to Intermediate-Low, 100% of the participants reached at least the lower end of the band, Novice-High, for listening comprehension, and many exceeded the upper end of the band, Intermediate-Low. Seven participants received proficiency ratings of Intermediate-High and Advanced-Low. Analogous findings were found for reading comprehension, where six participants had ratings of Intermediate-High and Advanced-Low.

Discussion

The establishment of both “rigorous and attainable outcomes for WL learning sequences is critical to supporting programs in developing strong curricula and measuring their outcomes” (Swanson et al. 13). While there is a dearth of literature regarding WL learner proficiency in institutions of higher education, the present study aims to provide a much-needed first step in establishing proficiency benchmarks in second and fourth semester Spanish language courses at a military academy. These findings not only serve post-secondary institutions but also play an integral role in providing WL understanding in order to support national security and WL endeavors.

The results of this study are important for several reasons. First and foremost, the data from this study provide valuable evidence of the accuracy of the benchmarks initially set by the language roadmaps in the Spanish curricula at the first- and second-year levels. Determining more precise benchmarks at each language level enables course developers to create effective syllabi that can help learners progress

more efficiently on their language learning continuum. For first-year Spanish, the data show that the benchmarks appear to be too conservative. For second-year Spanish, however, the benchmarks appear to be accurate thus far. Nevertheless, caution dictates that information from additional testing will provide further answers to determine the correctness of these benchmarks. While the data are intriguing, results from more than one year of testing are necessary to provide more reliable information leading to more precise benchmarking.

Second, particularly for the subject population in this study, use of the DLPT to measure language proficiency is a given in the military arena. The DLPT is used by most federal government agencies for testing language ability of Department of Defense personnel worldwide (“DLPT Relevant Information”). Given this reliance on the DLPT to gauge language proficiency, it is advantageous for this study’s subjects to understand their personal results from this testing in order to aim for further improvement in their language proficiency. Language proficiency for those serving in the military has both career implications and financial ramifications. In addition, demonstration of progress in language proficiency can provide strong motivation for further study.

Recognizing that the DLPT only assesses receptive skills, a word about receptive skills versus productive skills is in order here. In simplified terms, reading and listening are classified as receptive skills, while speaking and writing are classified as productive skills. In general, most language students express a desire to be able to speak the language they are studying; oral communication is their primary focus. Yet productive skills in an L2 generally lag behind the receptive skills, according to studies by Davies (“Receptive” 441), Van Parreren (251), and Yuzar and Rejeki (101–102).. There is, however, evidence that concentration on the development of the receptive skills in L2 learning lays a foundation for growth in the productive skills. Reading has been shown to improve oral comprehension and production in an L2 (Rodrigo 59). Other research indicates that listening is a primary skill that will transfer to and support speaking ability (Davies, “Receptive” 441; Sreena and Hankumaran 669). The results from this study directly related to receptive skills can certainly impact the productive skills in the L2 eventually. To that end, the researchers are also engaged in a study that evaluates the benchmarking of oral skills by learners of Spanish, using the OPI, also used by federal government agencies in general and the military in particular (see the study by Swanson and LeLoup). Analyses from studies such as these will offer additional material on which to draw for further benchmarking activity.

The U.S. Department of Defense continues to hold language and culture enabled military personnel in high regard. Findings from this study provide important information regarding expectations for language learners of Spanish and the curricula used to prepare these future military officers. Per U.S. Department of Defense guidelines, cadets graduating from USAFA with language expertise (i.e., a major in Foreign Area Studies or a minor in one of the eight languages taught at USAFA) must take the DLPT. This assessment reports language proficiency using the ILR scale, which assigns a numerical rating (e.g., 1+, 2) instead of using the ACTFL proficiency descriptors (e.g., Novice-High). The latter are more meaningful to language learners for feedback purposes. Thus, results from the APPT are more helpful in terms of feedback to the learner about his or her performance on the reading and listening tests.

Given the importance of the findings, this study is not without its limitations. More participants and more demographic information about them would be helpful. Such information as years of study of Spanish and the primary approach to WL instruction used in those classes prior to attending USAFA would be useful. Additionally, having confirmation of Communicative Language Teaching as the principal approach to teach Spanish at USAFA by instructors would corroborate the aims of the present benchmarking. For example, it would be informative to learn if all instructors are teaching in the target language 90%+ of the class time as recommended by the ACTFL “Facilitate” or if they are teaching in English primarily where participants learn about the language instead of acquiring it (Krashen 2). Finally, as indicated above, the subject pool at USAFA is potentially quite different from that of other institutions of higher education, certainly civilian institutions. It would be interesting to conduct a similar study in a civilian institution of higher education with language minors and majors.

Finally, these results make clear that future research is needed to corroborate such findings, include additional subject variables, and thus broaden the database on proficiency research. For example, information about the prior language learning experiences of the students would be helpful to be able to parse results and perhaps determine cause and effect more clearly. Also, a thorough examination of teaching practices that embrace Communicative Language Teaching as the basis for the curricula needs to be substantiated. It would be informative to know if these positive results are due to the curricular development and instructional methods used at USAFA in Spanish or if there are other variables that might explain some of the variance in outcomes.

Proficiency testing of language learners holds promise for moving language learners along their interlanguage continuum. By prioritizing such assessment, instructors and program coordinators as well as administrators can set achievable proficiency benchmarks for language learners at all levels of instruction. While proficiency testing can be impeded by cost due to WL programs lacking internal funds, educational leaders need to embrace the Language Flagship program call for K-20 programs to create and fund a sustainable process to assess language learner proficiency (Swanson et al. 2). WL skills are critical for America's diplomatic, business, and national security interests (La Corte and Voisine) and are in need today ("ACTFL Proficiency Guidelines").

APPENDIX A

A Comparison of the Proficiency Skill Levels of the ACTFL Scale
and the Interagency Language Roundtable (ILR)

Novice (Low/Mid/High)	0/0+
Intermediate (Low/Mid/High)	1/1+
Advanced-Low	2
Advanced-Mid	2
Advanced-High	2+
Superior	3/3+/4/4+

Works Cited

- “About—What Is The Language Flagship.” *The Language Flagship*, 2013, <https://www.thelanguageflagship.org/About>. Accessed 10 May 2024.
- “ACTFL Proficiency Placement Test.” *Language Testing International*, 2024, <https://www.languagetesting.com/actfl-proficiency-placement-test-appt>. Accessed 10 May 2024.
- “ACTFL Speaking Assessment: The Oral Proficiency Interview (OPI).” *Language Testing International*, 2024, <https://www.languagetesting.com/oral-proficiency-interview-opi>. Accessed 10 May 2024.
- “ACTFL Speaking Assessment: The Oral Proficiency Interview—Computer (OPIC).” *Language Testing International*, 2024, <https://www.languagetesting.com/oral-proficiency-interview-by-computer-opic>. Accessed 10 May 2024.
- . “Facilitate Target Language Use.” *American Council on the Teaching of Foreign Languages (ACTFL)*, 2010, <https://www.actfl.org/educator-resources/guiding-principles-for-language-learning/facilitate-target-language-use>. Accessed 10 May 2024.
- American Council on the Teaching of Foreign Languages (ACTFL), “ACTFL Proficiency Guidelines 2024.” *Language Testing International*, 2024, https://www.actfl.org/uploads/files/general/Resources-Publications/ACTFL_Proficiency_Guidelines_2024.pdf. Accessed 10 May 2024.
- . “Making Languages Our Business: Addressing Foreign Language Demand Among U.S. Employers.” *American Council on the Teaching of Foreign Languages (ACTFL)*, 2019, https://www.actfl.org/uploads/files/general/MakingLanguagesOurBusiness_FullReport.pdf. Accessed 10 May 2024.
- American Academy of Arts and Sciences. “America’s Languages: Investing in Language Education for the 21st Century.” *American Academy of Arts and Sciences*, Feb. 2017, <https://www.amacad.org/publication/americas-languages>. Accessed 10 May 2024.
- Carriedo, Robert. Personal communication. 7 January 2025.
- Carroll, John B. “Foreign Language Proficiency Levels Attained by Language Majors Near Graduation from College.” *Foreign Language Annals*, vol. 1, no. 2, 1967, pp. 131–51.
- Chang, Anna, et al. “Developing Listening Fluency Through Supported Extensive Listening Practice.” *RELC Journal*, vol. 50, no. 3, 2019, pp. 422–38, <https://doi.org/10.1177/0033688217751468>.
- Chesser, Christopher R. “Language Enabled Airman Program.” *Air Force Culture and Language Center*, 2022, <https://zlibrary.to/dl/language-enabled-airman-program>. Accessed 10 May 2024.

Clifford, Ray, and Troy L. Cox. "Empirical Validation of Reading Proficiency Guidelines." *Foreign Language Annals*, vol. 46, no. 1, 2013, pp. 45–61, <https://doi.org/10.1111/flan.12033>.

Cox, Troy L., and Ray Clifford. "Empirical Validation of Listening Proficiency Guidelines." *Foreign Language Annals*, vol. 47, no. 3, 2014, pp. 379–403, <https://doi.org/10.1111/flan.12096>.

Davies, Norman F. "Receptive Versus Productive Skills in Foreign Language Learning." *Modern Language Journal*, vol. 60, no. 8, 1976, pp. 440–43.

---. "Putting Receptive Skills First: An Experiment in Sequencing." *Canadian Modern Language Review*, vol. 36, no. 3, 1980, pp. 461–67.

DeKeyser, Robert M. "Research on Language Development During Study Abroad." *Language Acquisition in Study Abroad and Formal Instruction Contexts*, edited by Carmen Perez-Vidal, John Benjamins Publishing Company, 2014, pp. 313–26. <http://doi.org/10.1075/aals.13.16ch13>.

Dewey, Dan P., et al. "Japanese Language Proficiency, Social Networking, and Language Use During Study Abroad: Learners' Perspectives." *Canadian Modern Language Review*, vol. 68, no. 2, 2012, pp. 111–37.

Di Silvio, Francesca, et al. "The Development of L2 Fluency During Study Abroad: A Cross-Language Study." *Modern Language Journal*, vol. 100, no. 3, 2016, pp. 610–24.

"DLPT Relevant Information and Guides." *Defense Language Institute Foreign Language Center (DLIFLC)*, www.dliflc.edu/resources/dlpt-guides.

Freed, Barbara F., editor. *Second Language Acquisition in a Study Abroad Context*. John Benjamins Publishing Company, 1995. <https://doi.org/10.1075/sibil>. Accessed 10 May 2024.

Harding, Luke, et al. "Diagnostic Assessment of Reading and Listening in a Second or Foreign Language: Elaborating on Diagnostic Principles." *Language Testing*, vol. 32, no. 3, 2015, pp. 317–36.

Hernández, Todd A. "Promoting Speaking Proficiency Through Motivation and Interaction: The Study Abroad and Classroom Learning Contexts." *Foreign Language Annals*, vol. 43, no. 4, 2010, pp. 650–70.

Isabelli-García, Christina, et al. "Language Learning and Study Abroad." *Language Teaching*, vol. 51, no. 4, 2018, pp. 439–84.

- Isbell, Dan, and Paula Winke. "ACTFL Oral Proficiency Interview—Computer (OPIc)." *Language Testing*, vol. 36, no. 3, 2019, pp. 467–77, <https://doi.org/10.1177/0265532219828253>.
- Jiang, Dayu, et al. "The Curious Case of Improving Foreign Language Listening Skills by Reading Rather Than Listening: An Expertise Reversal Effect." *Educational Psychology Review*, vol. 30, no. 3, 2018, pp. 1139–65, <https://doi.org/10.1007/s10648-017-9427-1>.
- Krashen, Stephen. *Second Language Acquisition and Second Language Learning*. Pergamon Press, 1981.
- La Corte, Matthew, and Alexander Voisine. "Foreign Language Skills Are Crucial for America's Diplomatic, Business, and Security Interests." *Niskanen Center*, 27 May 2020, <https://www.niskanencenter.org/foreign-language-skills-are-crucial-for-americas-diplomatic-business-security-interests>. Accessed 10 May 2024.
- Looney, Dennis, and Natalia Lusin. "Enrollments in Languages Other Than English in United States Institutions of Higher Education, Summer 2016 and Fall 2016: Preliminary Report." *Modern Language Association*, Feb. 2018, <https://www.mla.org/Resources/Guidelines-and-Data/Reports-and-Professional-Guidelines/Enrollments-in-Languages-Other-Than-English-in-United-States-Institutions-of-Higher-Education>. Accessed 10 May 2024.
- Lusin, Natalia, et al. "Enrollments in Languages Other Than English in US Institutions of Higher Education, Fall 2021." *Modern Language Association*, 2023. <https://www.mla.org/content/download/191324/file/Enrollments-in-Languages-Other-Than-English-in-US-Institutions-of-Higher-Education-Fall-2021.pdf>. Accessed 10 May 2024.
- Malone, Margaret. "Afterword and Next Steps." *Foreign Language Proficiency in Higher Education*, edited by Paula Winke and Susan Gass, Springer, 2019, pp. 309–18.
- McLaughlin, Barry. *Theories of Second-Language Learning*. Arnold, 1995.
- McManus, Kevin, et al. "A Longitudinal Study of Advanced Learners' Linguistic Development Before, During, and After Study Abroad." *Applied Linguistics*, vol. 42, no. 1, 2021, pp. 136–63.
- National Council of State Supervisors for Languages (NCSSFL) and the American Council on the Teaching of Foreign Languages (ACTFL). "How to Use the NCSSFL-ACTFL Can-Do Statements, 2014." *NCSSFL-ACTFL Can-Do Statements*. ACTFL Language Connects, 2017, <https://www.actfl.org/educator-resources/ncssfl-actfl-can-do-statements>. Accessed 10 May 2024.

- National Standards Collaborative Board. “World-Readiness Standards for Learning Languages.” *ACTFL*, 2015, <https://www.actfl.org/resources/world-readiness-standards-learning-languages>.
- Renandya, Willy A., and George M. Jacobs. “Extensive Reading and Listening in the L2 Classroom.” *English Language Teaching Today*, edited by Willy A. Renandya and Puji W. Handoyo, Springer, 2016, pp. 97–110.
- Rodrigo, Victoria. *La comprensión lectora en la enseñanza del español LE/L2: De la teoría a la práctica*. London: Routledge, 2018.
- Soneson, Dan, and Elaine E. Tarone. “Picking Up the PACE: Proficiency Assessment for Curricular Enhancement.” *Foreign Language Proficiency in Higher Education*, edited by Paula Winke and Susan Gass, Springer, 2019, pp. 45–70.
- Sreena, S., and M. Ilankumaran. “Developing Productive Skills Through Receptive Skills—A Cognitive Approach.” *International Journal of Engineering and Technology*, vol. 7, no. 4, 2018, pp. 669–73.
- Swanson, Pete, and Jean W. LeLoup. “Validating Proficiency Benchmarks at a US Military Academy.” *Language Testing International*, 2022, <https://www.languagetesting.com/blog/validating-proficiency-benchmarks-at-a-us-military-academy>. Accessed 10 May 2024.
- Swanson, Pete, et al. *Proficiency Benchmarking in Spanish*. Central States Report, 2022, pp. 1–16.
- Tschirner, Erwin. “Listening and Reading Proficiency Levels of College Students.” *Foreign Language Annals*, vol. 49, no. 2, 2016, pp. 201–23.
- Tulloch, Brandon, and Lourdes Ortega. “Fluency and Multilingualism in Study Abroad: Lessons from a Scoping Review.” *System*, vol. 71, 2017, pp. 7–21.
- United States Air Force Academy (USAFA). “Demographic Profile of the Incoming Class of 2025.” 8 July 2021, <https://www.usafa.edu/app/uploads/CL2025-Class-Profile.pdf>.
- Van Parreren, Carel. “Primacy of Receptive Skills in Foreign Language Learning.” *System*, vol. 11, no. 3, 1983, pp. 249–53.
- Vande Berg, et al. “The Georgetown Consortium Project: Interventions for Student Learning Abroad.” *Frontiers: The Interdisciplinary Journal of Study Abroad*, vol. 18, no. 1, 2009, pp. 1–75.

Wagner, Elvis. "Assessing Listening." *The Routledge Handbook of Language Testing*, 2nd ed., edited by Glenn Fulcher and Luke Harding, Routledge, 2019, pp. 223–35.

"When Do Children Learn to Talk?" Linguistic Society of America. https://old.linguisticsociety.org/sites/default/files/Language_Acquisition.pdf. Accessed 10 May 2024.

Winke, Paula., et al. "Modern-Day Foreign Language Majors: Their Goals, Attainment, and Fit Within a Twenty-First Century Curriculum." *Foreign Language Proficiency in Higher Education*, edited by Paula Winke and Susan Gass, Springer, 2019, pp. 117–36.

Yuzar, Ella, and Siti Rejeki. "Correlation Between Productive and Receptive Language Skills: An Examination on ADFELPS Test Scores." *Study of Applied Linguistics and English Education*, vol. 1, no. 2, 2021, pp. 99–113.

