



Language Aptitude as Predictor of Performance

John Pendergast

United States Military Academy, West Point

Thomas Dean

United States Military Academy, West Point

Robert Jannusch

United States Military Academy, West Point

West Point attempts to enroll cadets in their required core language courses based on their top three choices/preferences. Because seats are limited in the most popular languages, it was necessary to devise an objective method for reassigning cadets to courses with available seats, most of which are in Cat III and IV languages, like Arabic, Chinese, Russian, and Persian. West Point has used the Modern Language Aptitude Test (MLAT) and Pre-Defense Language Aptitude Battery (Pre-DLAB) as tools for making this decision. Analyzing the records of 17,000 cadets, this study uses logistic regression to assess the effectiveness of these aptitude tests in predicting expected performance. The analysis finds that aptitude as measured by the MLAT and Pre-DLAB modestly predicts performance.

Keywords: *Language Aptitude, Modern Language Aptitude Test, Defense Language Aptitude Battery, Language Placement, Logistic Regression, L2 Performance*

INTRODUCTION

With each entering class, the Department of English and World Languages (DEWL) at West Point finds more cadets wanting to take certain languages than there are seats available. When this occurs, we try to assign as many of these cadets as possible to one of our Category III or IV “strategic languages”: Arabic, Chinese, Persian, and Russian (Savko et al., 2022, p. 96). From 2007 to 2019, DEWL used the Modern Language Aptitude Test (MLAT) to assess overall language aptitude among incoming cadet candidates who needed to be moved into one of their second-

APPLIED LANGUAGE LEARNING (ISSN 2164-0912) is the journal of the Defense Language Institute Foreign Language Center, in the U.S. Department of Defense. The views expressed herein are those of the author(s), not the Department of Defense or its elements. Further reproduction is not advisable. Whenever copyrighted materials are reproduced in this publication, copyright release has ordinarily been obtained only for use in this specific issue. Requests for reprints should be directed to the individual authors.



or third-choice strategic languages. An MLAT score of 110 was established as a good predictor that a cadet would achieve a passing grade in the strategic languages (per the Department of Defense's language placement threshold scores on the Defense Language Aptitude Battery, DLAB). In 2020, the MLAT could no longer be used for placement, however, due to restrictions created by the COVID-19 pandemic. As a result, the department¹ switched to using a device DEWL refers to as the DLAB Screening Tool (DST)². The DST was created by the Defense Language Institute Foreign Language Center and was based on the DLAB. The department's research and analysis cell determined that a score of 22 on the DST was roughly equivalent to a score of 110 on the MLAT, and for the last three cycles of placement testing, DEWL has used that score for assignment and reassignment purposes.

This paper reports on research comparing the possibility of DST scores and MLAT scores to predict student performance in their language classes.

LITERATURE REVIEW

Language Aptitude Tests

Prior to World War II, the Army used various instruments for recruitment screening, which served as precursors of psychological evaluations, IQ tests, and the Armed Services Vocational Aptitude Battery (ASVAB), but they were not oriented toward linguistic ability (Carroll, 1965, p. 90-91). Interestingly, one of the earliest language aptitude prototypes was administered at West Point in 1954, called the West Point Aptitude Test or WPAT (p. 117-19).³

The main shortcoming of the early tests is that they were firmly grounded only in the test taker's mastery of English. Carroll and Sapon revolutionized the testing paradigm by studying decades of previous research on language aptitude and concluding that aptitude depended on factors that indicate an individual's linguistic behavior and observations (Carroll, 1965, p. 95). In the mid 1950s, Carroll and Sapon developed these factors by trial and experiment and released the Modern Language Aptitude Battery in 1959 (later re-designating it as the Modern Language Aptitude Test). It soon became the standard for language screening in the DoD and the Foreign Service, therefore being used at West Point from 2011-2020.

While many studies have determined that language aptitude is a key factor in students' language learning success (Andringa & Dąbrowska, 2019; Carroll, 1959; Curcic et al., 2019; Dahlen & Caldwell-Harris, 2013; Petersen & Al-Haik, 1976), some scholars have noted limitations inherent

¹ In 2024, the Department of Foreign Languages merged with the English division to become the Department of English and World Languages.

² The developers call it the Pre-DLAB. Because of the possibility for cadets to confuse this test with the actual DLAB (which has often happened with the MLAT), we chose to call it the DST. ("CASL Research Fact Sheet")

³ The results on that test were compared to performance in the language classes taught at that time: French, German, Portuguese, Russian, and Spanish. Additionally, these results were compared to the cadets' performance in English and Mathematics. The similarities to our present study are striking.

in these exams. Ehrmann and Oxford (1995), for example, point out that “Some of these measures have been questioned because of their focus on analytic, structural skills, their potential incompatibility with less structured, highly communicative language teaching approaches, and their strong relationship with a general intelligence factor” (p. 68). Nevertheless, aptitude tests continue to provide useful information in many institutions.

Additionally, since student motivation has been identified as an equally important predictor of success in L2 coursework (Ehrman & Oxford, 1995, pp. 68, 72, 76; Gardner & Lambert, 1972; Hardison et al., 2012), DEWL uses both a measure of aptitude as well as a short language survey which allows cadets to rank their top three world language preferences for study at West Point and to indicate whether they have had previous language study experience. Both instruments allow the department to consider multiple datapoints when placing cadets into language classes.

Research Question

This study was guided by the following question: To what degree do the MLAT and DST aptitude tests serve as reliable predictors of student performance in West Point’s two-semester core language sequence?

METHOD

Setting

In June 2023, the research team submitted a proposal to West Point’s Human Research Protections Program to study the degree to which the MLAT and DST aptitude tests serve as reliable predictors of performance in the two-semester core language sequence. Our proposal appeared before the Academy’s Institutional Research Committee in July 2023 and received nearly immediate approval. Cadet records from the class of 2011 to the present, were requested and compiled comprising over 17,000 individual records.

Logistic Regression

Logistic regression was used as a statistical technique to model the probability of success based on predictor variables. These data were used to analyze the relationship between MLAT scores and DST scores on the probability that students will succeed in their world language courses. In this context, success means whether they met/exceeded or failed to achieve expected language course grades. These regression models were standardized across all graduating classes. This reduces the impact of any general inconsistencies across the range of language aptitude test averages.

Data Sources

This study utilized five data sources for analysis: (1) students' scores on the Modern Language Aptitude Test (MLAT) from 2011-2020, (2) students' scores on the DLAB Screening Tool (DST) from 2020-2025, (3) students' grade point averages (called the APSC at West Point), (4) students' language course grades over two semesters, and (5) students' responses on West Point's Language Preference Survey (used to identify students' language study experience and preferences, see Appendix A). Other predictors often used by West Point's registrar's office, such as SAT/ACT and College Entrance Examination Rank (CEER) scores, were also considered, but preliminary analyses indicated that they offered no additional predictive information for this study (See Appendix B for a sample of all placement data considered for this study).

For this study, the operative focus is on performance and not on language proficiency. As important as language proficiency is in second language acquisition, it was not analyzed beyond the grade cadets earned in their language class. The grade in a language class, however, is also not to be construed as an indicator of proficiency, as many graded events were partially assessed based solely on whether they were completed/not completed.

Calculating Student Performance

As mentioned above, student performance was calculated as a binary variable of whether a student either "met/exceeded" a certain performance expectation or not. This was accomplished by comparing cadet grades in their core language courses to their overall GPA. For example, a student with a cumulative GPA of 4.0 and an assigned language course grade of B was coded as "not meeting" expected performance because their GPA suggests their performance should be higher. Comparatively, a student with a cumulative GPA of 2.7 and a language course grade of B was coded as "meeting/exceeding" expected performance because their language course grade aligns more closely with their GPA. This data point was then compared with the cadets' score on the DST and MLAT to determine whether a score of 22 and 110 (respectively) predicts this performance.

Participant Data

The segment of the cadet population of greatest interest (hereafter called "treated group") were those reassigned from their first-choice language to a strategic language due to enrollment limitations (class size, available instructors, etc.). The remaining cadets' records were used as control data. When comparing DST scores to GPA and language core course grades for the classes of 2025 and 2026 across all eight languages, data for roughly 1,250 cadets were collated and analyzed. Because most cadets complete their two-semester core language sequence by the end of the sophomore year, data were available for all 2025 cadets, except those who deferred core language to their junior year (mostly life science majors, often called "pre-med" at other colleges). For the class of 2026, these data were only available for those who completed their core language sequence in their freshman year (approximately 175 cadets).

Because considerably more MLAT data are available for classes 2011-2024, the same comparison was conducted for those class years. This has two primary benefits. It permitted us to validate the initial assessment of the MLAT's reliability as a performance predictor and made possible a consistent comparison between the MLAT and DST using the same criteria across all sets of data. There are important implications to the reassignment variable above. First, reassigned students are those who were not placed in their first-choice language. As a result, in this segment of the study, these students were placed in various strategic languages (Arabic, Chinese, Russian and Persian). Thus, for the following logistic regressions, the impact of reassignment can be seen when the model highlights data for those strategic languages.

Empirical Methodology

Often called Grade Point Average (GPA) at other colleges, at West Point, cadets' GPA is called Academic Program Score Cumulative (APSC). As an aggregate measure of a cadet's academic record, APSC provides a holistic indicator of academic achievement at West Point. For this reason, we incorporated APSC as a benchmark to evaluate whether a student's performance in language classes was consistent with their overall academic performance. Although we compared cadet APSC scores with their DST scores for the target population (2025-2026), we determined APSC to be ill-suited for our control group (2011-2024). Generally speaking, in our model, a good control variable occurs before the analyzed result. In this case, APSC was calculated after an individual's performance in a language class was assessed. Because of this, APSC was not included as a control variable because it interfered with our ability to isolate causal effects.

To create a dynamic and reasonable performance expectation for a cadet with a given APSC, we established a threshold at 1.5 percentage points below the APSC. For example, an individual with a 3.00 GPA equates to a B average. The percentage range for a B in DEWL is 83% to 86.99%. Using the lower bound of the range and subtracting 1.5 percentage points, for example, 83% - 1.5% equates to a performance threshold of 81.5%. We also established this threshold to account for language category variables and minor performance fluctuations. Thereby, academic success is measured against a student's established potential rather than an inflexible standard while also providing a realistic buffer that recognizes individual variation without demanding statistical perfection.

Empirical Analysis

The following logistic regression models were coded using R. For each model, performance was analyzed based on averages of cadets' grades in their two core language courses at West Point. The variable is binary, meaning student performance either met or exceeded the expectation for the model, or not. See Figure 1.

Figure 1
Summary of Balance for Matched Data

Summary of Balance for Matched Data			
	Means Treated	Means Controlled SD	Mean Difference
Distance	0.222	0.222	0.0000
CEER	624.581	623.651	0.9300
MLAT	114.875	114.799	0.0760
SAT_Math	657.684	656.226	1.4580
SAT_Verbal	635.282	631.789	3.4930
Persian_Enrolled	0.205	0.203	0.0020
Russian_Enrolled	0.348	0.341	0.0070
Chinese Enrolled	0.158	0.164	-0.0060
Arabic_Enrolled	0.289	0.292	-0.0030

Figure 1 summarizes the difference in the means of each variable showing minimal difference in means (“matched data”⁴) between the control group and the treated group thereby minimizing any confounding variables. As a result of this balancing procedure, variables such as SAT and CEER scores were not included in the final models (Figure 4b). For additional diagnostic data, see the Absolute Standardized Mean Difference figure in Appendix C.

RESULTS

The results have confirmed that there is a modest correlation between aptitude scores and academic performance in language classes. This supports the conclusions of research cited above that scores on language aptitude tests have a direct statistical relationship with academic performance in language courses.

A series of logistic regression models was estimated including aptitude scores (MLAT/DST), language indicators, and reassignment status. See Appendix D for an overview of the underlying formulas used in these models.

⁴ “Matched” means that each individual in the treated group has been paired (“matched”) with a similar individual in the control group based on key characteristics (e.g., SAT scores, MLAT, CEER). This creates two groups that are as similar as possible, except for the variable being studied.

MLAT Analysis

Figure 2 reflects the impact of MLAT scores on performance. The regression coefficient for MLAT (0.234, $p < .001$), indicates that higher MLAT scores are associated with a higher likelihood of meeting or exceeding the expected performance threshold. In this regression, the intercept is the predicted value when all other independent variables are set to 0. The intercept of 2.285 indicates that a student in the baseline category (with predictors set to zero) has an approximately 91% chance of achieving the expected grade (i.e., a grade at or above their specified threshold).

Figure 2
Effect of MLAT on Performance

MLAT	0.234*** (0.000)
Intercept	2.285*** (0.000)

Note. *** represents statistical significance at the 95% confidence level

Figure 3 presents the same variables as Figure 2, with the addition of the “Reassigned” variable, which is negatively associated with performance. This suggests that involuntary reassignment to strategic languages is associated with lower performance compared to those in non-strategic languages. Given the intercept of 1.106, a student in the baseline category (with all predictors set to zero) has a roughly 75% chance of getting a grade at or above their specified threshold.

Figure 3
Effect of MLAT on Performance, including Reassignment

MLAT	0.248*** (0.000)
Reassigned	-0.539*** (0.000)
Intercept	1.106*** (0.000)

Note. *** represents statistical significance at the 95% confidence level

Figure 4a illustrates the correlation of MLAT scores to cadet success in the strategic languages when those strategic languages were chosen by the cadets as their first-choice language. The estimates are in comparison to the average student in a non-strategic language course. This indicates that a cadet choosing to take a strategic language was less likely to meet the expected performance threshold compared to the average cadet not taking a strategic language. Moreover, the effect of the reassignment variable has decreased from the model in Figure 3 and has become statistically insignificant. The reassignment variable captures differences between

students who were involuntarily assigned to a strategic language and those who requested it, holding language enrollment constant.

Figure 4 (a and b)

Effect of MLAT on Performance (Standard vs. Matched)

Effect of MLAT on Performance		Effect of MLAT on Performance (Matched)	
MLAT	0.029 *** (0.000)	MLAT	0.043 *** (0.000)
Russian Enrolled	-0.055 *** (0.000)	Russian Enrolled	0.039 (0.059)
Persian Enrolled	-0.042 *** (0.001)	Persian Enrolled	0.076 ** (0.002)
Arabic Enrolled	-0.084 *** (0.000)	Chinese Enrolled	-0.052 * (0.047)
Chinese Enrolled	-0.137 *** (0.000)	Reassigned	0.0004 (0.983)
Reassigned	-0.008 *** (0.447)	Intercept	0.831 *** (0.000)
Intercept	0.936 *** (0.000)		

Note. *** represents statistical significance at the 95% confidence level

Figure 4b depicts a smaller dataset (approximately 1250 students) of students who were involuntarily reassigned into a strategic language. These students had virtually the same ACT, SAT, CEER score and any other pertinent academic credentials with the only difference between the two groups being that one group requested enrollment in the strategic language rather than being involuntarily reassigned into it.⁵ In this model, Arabic was used as the baseline variable for each strategic language comparison. The “Reassigned” variable here had a value of 0.0004 and a p-value of 0.983 and the “MLAT” variable had a value of 0.043 with $p < .001$. Further, Russian and Persian language variables were positively associated with cadet performance while Chinese is negatively associated.

DST Analysis

In 2021 after COVID, DEWL switched from the MLAT to the DST. Because of this, only 3,000 cadets have taken the DST and completed their respective core language courses as shown in Figures 5a and 5b.

⁵ This model's key component is “ceteris paribus” - a Latin term that translates in English to “all else equal” (Persky, 1990, p. 191-92). Unlike the first three models which use all the students in the dataset, this model finds a set of students who were reassigned with specific academic credentials held constant and compares them to a subset of students with similar academic credentials. The only difference between the subsets of students is that one subset was reassigned to a strategic language, and the other was not.

Figure 5 (a and b)

Effect of DST on Performance (Standard vs. Matched)

Effect of DST on Performance		Effect of DST on Performance (Matched)	
DST	0.054 *** (0.000)	DST	0.065 *** (0.000)
Russian Enrolled	0.122 *** (0.000)	Russian Enrolled	0.020 (0.650)
Persian Enrolled	.137 *** (0.001)	Persian Enrolled	-0.032 (0.557)
Arabic Enrolled	.127 *** (0.000)	Chinese Enrolled	-0.103 ** (0.025)
Chinese Enrolled	0.03 (0.309)	Reassigned	-0.413 *** (0.000)
Reassigned	-0.417 *** (0.000)	Intercept	0.877 *** (0.000)
Intercept	0.754 *** (0.000)		

Note. *** represents statistical significance at the 95% confidence level

On average students with higher DST scores performed better. However, in both the standard (Fig 5a) and matched model (5b) the “Reassigned” variable was negatively associated with performance and statistically significant ($p < .001$). In other words, students who were involuntarily reassigned into strategic languages had significantly lower performance than those who were not. We found that in the class of 2026, 285 cadets were reassigned to strategic languages compared to an average of 80/year in previous years.⁶ Of these 285 cadets, there was a much higher rate of lower-than-expected performance, as shown by the negative reassignment coefficient. This appears to be an outlier.⁷ The “Reassigned” variable here had a value of -0.413 with $p < .001$ while the “DST” variable was 0.065 also with $p < .001$. Further, the Russian language variable was positively associated with cadet performance while Persian and Chinese were negatively associated.

⁶ For the other class years, an average of 80 cadets were reassigned (with a range of from 32-141). Thus, the class of 2026 had three to four times as many reassigned cadets than other years.

⁷ There are at least four factors that played a role in making that class year an outlier. When the class of 2026 were sophomores, a new course called Advanced Beginner Spanish was introduced. The administrative requirements for implementing that course demanded that the planners be a bit more draconian in processing cadets’ paperwork. As an example, the first step for cadets who wanted to be considered for Advanced Beginner Spanish was to submit a survey outlining their prior experience with the Spanish language. Cadets who did not submit that survey on time were not given a reminder or second try because the Spanish office already knew they needed to trim down the number of cadets by at least 80 to avoid exceeding their enrollment cap. As a result, more than 120 cadets failed to submit their survey on time. Additionally, two professors were on sabbatical in that academic year, which eliminated another 50-60 seats from European languages that could only be filled in Arabic and Chinese. Finally, 40 life science majors pushed back their language requirement from sophomore to junior year and were almost entirely placed in strategic languages. Therefore, it is plausible that the negative relationship between DST score and its effect on performance is not an accurate result.

DISCUSSION

Recall that we defined performance not as an absolute score or grade but rather how closely the grades in the core language courses match those of the cadets' GPA. Overall, scores on both the MLAT and DST were directly associated with performance, meaning that both tests are good predictors of performance. Scores associated most consistently with expected or better than expected performance were as follows:

MLAT 117 (7 points higher than current cut-score)

DST 24 (2 points higher than current cut-score)

Most notably, for cadets who took the MLAT, the effect of reassignment on performance was statistically insignificant, meaning reassignment did not adversely affect performance. For cadets who took the DST, a negative impact of forced reassignment can be observed, but we believe that this result is an outlier and is attributable more to structural changes in DEWL placement priorities than to the DST itself.

The findings associated with the Class of 2026 should be interpreted as a longitudinal data limitation rather than a reflection of cadet aptitude. Because the data set coincides with significant administrative restructuring within DEWL, the results are likely confounded by institutional transition. In this context, the DST scores do not serve as a clean metric of performance, but rather as a data point captured during a period of high systemic variance.

Additionally, the findings of this study are constrained by several unquantifiable variables that introduced statistical noise into the performance data. A primary limitation lies in the instructor variation across the language programs and within individual languages, as the difference in instructional delivery method means that a cadet's success can be a reflection of an individual instructor's teaching style rather than their own language aptitude. This aligns with Ehrman, who used personality tests to study the compatibility of certain teaching styles with student learning styles but later concluded that such tests are weak predictors of success (Ehrman, 1990; Ehrman & Oxford, 1995). Furthermore, our data do not appear to account for so-called "affective factors," such as cadet motivation, which can cause performance to deviate from predicted benchmarks (Hardison et al., 2012, p. 11). The research is also impacted by environmental and extracurricular stressors, including absences due to intercollegiate sports, club commitments, or unexpected illness. In other words, researchers broadly agree that these factors affect language learning but do not agree about how and to what extent. Because these factors remain unattributed in the current model, they function as confounding variables that obscure the relationship between the MLAT/DST and academic performance.

Implementing the recommended adjustments to enrollment thresholds by increasing the cut-scores, specifically raising the MLAT to 117 and the DST to 24, provides a more rigorous baseline for ensuring cadets possess the necessary aptitude for linguistic success. By establishing these higher benchmarks, the Academy can likely reduce the incidence of training failures and might

lower attrition rates, as cadets will be better equipped to handle the intensive demands of the language curriculum from the outset. Moving forward, policy refinement should focus on an iterative evaluation of these thresholds to ensure they remain calibrated with evolving academic standards.

Future research should prioritize longitudinal tracking of cadet performance to validate these benchmarks, ensuring that West Point continues to optimize its selection process while fostering a linguistically agile officer corps capable of leading in an increasingly complex global environment. Looking beyond the specific goals of this study toward a broader research context, the large corpus available in the study could be used to explore evidence that, in contrast to students with lower attainment in their native (L1) language, those with higher attainment get higher scores on L2 aptitude tests (MLAT) and stronger L2 achievement (similar to the APSC numbers in this study) than students with weaker L1 skills. (Sparks et al., 2011, p. 255).

CONCLUSION

The study establishes a clear association between scores on the MLAT and the DST and performance in core language classes. Those with higher aptitude scores consistently perform at least as well in their language courses as they do in their other academic courses, including those involuntarily reassigned to a strategic language based on their aptitude scores. One change that the results suggest is that the threshold score used to enroll cadets in Cat III and IV languages, especially Chinese, may need to be adjusted several points higher.

Acknowledgement

There are no financial or personal relationships that could influence the manuscript.

Authors

John Pendergast, Associate Professor and Program Director of Russian at West Point, writes about second language acquisition and Russian and German music and letters. Recently Chair of the MLA Opera and Musical Performance Executive Committee, this year he co-authored a chapter on decolonizing Russian study abroad in *Inclusive Strategies and Critical Pedagogies for East European and Eurasian Languages* (Academic Studies Press) and was Editor-in-chief of *Arms, Letters, and the Humanities* (West Point Press).

Thomas Dean, Logistics Officer serving at the 101st Division Sustainment Troops Battalion at Fort Campbell, Kentucky, graduated from the United States Military Academy at West Point in May of 2025. He graduated with a bachelor's degree in science, majoring in Economics with a minor in Applied Statistics with a concentration in systems engineering. He graduated from Logistics Basic Officer Leader Course at Army Sustainment University in Fort Lee, Virginia in November of 2025.

Robert Jannusch, Infantry Officer serving at the 3rd Infantry Division at Fort Stewart, Georgia, graduated from the United States Military Academy at West Point in May of 2025. He graduated with a bachelor's degree in science, majoring in Economics with a concentration in robotics engineering. He graduated from Infantry Basic Officer Leader Course at Fort Benning, Georgia in January of 2025.

REFERENCES

- Andringa, S. & Dąbrowska, E. (2019). Individual differences in first and second language ultimate attainment and their causes. *Language Learning*, 69(S1), 5-12.
- Carroll, J. B. (1964). *Language and thought*. Prentice-Hall.
- Carroll, J. B. (1959). *Modern language aptitude test: MLAT manual*. The Psychological Corporation.
- Carroll, J. B. (1965). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research and education* (pp. 87-136). John Wiley & Sons.
- CASL research fact sheet: Predicting DLAB performance (2009). Center for Advanced Study of Language [CASL], University of Maryland.
- Curcic, M., Andringa, S., & Kuiken, F. (2019). The role of awareness and cognitive aptitudes in L2 predictive language processing. *Language Learning*, 69(S1), 42-71.
- Dahlen, K., & Caldwell-Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal* 97(4), 902-16.
- Ehrman, M. E. (1990) The role of personality type in adult language learning: An ongoing investigation. In T. Parry & C. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 126-178). Prentice-Hall.
- Ehrman, M. E. & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *The Modern Language Journal* 79(1), 67-89.
- Gardner, R. C. & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning*. Newberry House.
- Hardison, C. M., Miller, L. W., Li, J. J., Schroeder, A. N., Burkhauser, S., Robson, S., and Lai, D. (2012). *Second-language skills for all?* RAND Corporation.
- LaValley, M. P. (2008). Logistic regression. *Circulation* 117(18), 2395-2399.
- Persky, J. (1990). Retrospectives: Ceteris paribus. *Journal of Economic Perspectives* 4(2), 187-93.
- Petersen, C. R. & Al-Haik, A. A. (1976). The development of the Defense Language Aptitude Battery (DLAB). *Educational and Psychological Measurement* 36(2), 369-380.
- Savko, P., Savukinas, R., Cutter, N., O'Reilly, E., & Sharif, S. (2022). *DLIFLC general catalog 2021-2022*. Defense Language Institute Foreign Language Training Center. <https://www.dliflc.edu>.
- Sparks, L., Patton, J., Ganschow, L. & Humbach, N. (2011). Subcomponents of second-language aptitude and second-language proficiency. *The Modern Language Journal*, 95(2), 253-273.

APPENDIX A

Language Preference Survey (administered during Cadet Basic Training)



CBT LANGUAGE SURVEY FORM

CO ___ PLT ___ NAME _____

1. **Language Experience:** Fill in the buttons beside each language indicating the number of years' study or experience you have in that language.

Language	Years					
	None	1	2	3	4	5
ARABIC	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
CHINESE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
FRENCH	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
GERMAN	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PERSIAN	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
PORTUGUESE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
RUSSIAN	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
SPANISH	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2. **Other Language Experience.**

Other Language:

3. **Plebe Year Study:** Select "yes" or "no" to indicate your desire to study a Foreign Language during your Plebe Year.

Plebe Year Study: Yes No

4. **Order of Preference:** Indicate, in order of preference, your three choices for language study by selecting a language for each choice.

Note:

At least one selection must be Arabic, Chinese, Persian, or Russian.

1st CHOICE	2nd CHOICE	3rd CHOICE
<input type="text"/>	<input type="text"/>	<input type="text"/>

ARABIC
CHINESE
FRENCH
GERMAN
PERSIAN
PORTUGUESE
RUSSIAN
SPANISH

5. **Concentration:** Select "yes" or "no" to indicate your desire to concentrate in Foreign Languages.

Concentration: Yes No



APPENDIX B

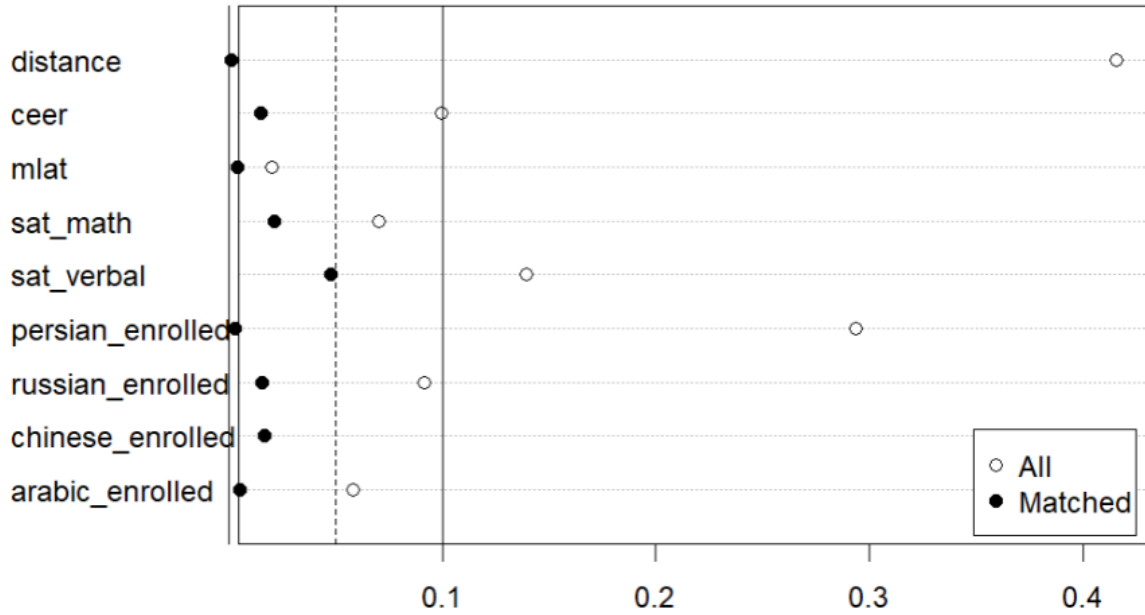
Sample Language Survey Data + Registrar Data (CEER, SAT/ACT, and APSC scores)

Class Year	A	C	F	G	Pe	Po	R	S	Other Exp	1st Choice	2nd Choice	3rd Choice	Placement	Plebe_Lang	Major	Minor	MLAT	DST	first sem	first sem	second	second	second	CEER	SAT math	SAT verbal	ACT math	ACT reading comp	APSC
2017	0	0	0	0	0	0	0	0	2	German	Arabic	Chinese	German	Y	Y	145		LG203	A	94.05	LG204	A	94.55	729	760	750	34	35	3.807
2017	0	0	0	3	0	0	0	0		German	Russian	French	German	N	Y	145		LG371	A+	97.5	LG475	A	94.45	628	650	590	29	22	3.339
2017	0	0	0	3	0	0	0	0		German	Spanish	Arabic	German	N	N	145		LG371	A+	100.25	LG475	A+	97.4	769	720	800	35	34	4.132
2017	0	0	0	1	0	0	0	0		German	French	Persian	German	N	N	143		LG203	A+	97.1	LG204	A	95.15	645	650	720	30	34	3.379
2017	0	0	0	2	0	0	0	0		German	French	Arabic	German	Y	Y	142		LG371	A+	97.75	LG475	A	93.55	738	690	760	35	35	3.743
2017	0	0	0	2	0	0	0	0		German	Russian	Arabic	German	N	N	142		LA203	B+	88.13	LA204	A-	90.35	653	710	620	34	30	3.196
2017	0	0	0	1	0	0	0	0	NONE	German	Russian	Persian	German	N	N	141		LG203	A	93	LG204	B+	87.64	579	550	730	23	36	2.712
2017	0	0	0	0	0	0	0	3		German	French	Russian	German	N	N	139		LG203	A-	91.8	LG204	A	94.45	682	740	700	30	34	3.547
2017	0	0	0	0	0	0	0	0	KOREAN	German	French	Chinese	German	N	N	138		LG203	A	94.5	LG204	A-	90	622	710	540			2.565
2017	0	0	0	0	0	0	0	1		German	Russian	Chinese	German	N	N	137		LG203	A-	90.48	LG204	A-	91.6	724	750	720	30	34	3.507
2017	0	0	0	0	0	0	0	0	LATIN- 1	German	French	Arabic	German	N	Y	137		LG203	B	84.28	LG204	A	93	636	550	670	28	36	2.609
2017	0	0	0	0	0	0	0	1	LATIN 2	German	Russian	French	German	N	N	136		LG203	A-	91.25	LG204	A	93.3	648	630	630			3.4
2017	0	0	1	0	0	0	0	0	NONE	German	French	Russian	German	N	N	136		LG203	A	95.53	LG204	A	95.3	663	680	600	34	29	3.439
2017	0	0	0	2	0	0	0	0	N/A	German	Spanish	Russian	German	N	N	136		LR203	A	94.58	LR204	A	95.58	657			31	24	3.814
2017	0	0	0	0	0	0	0	2		German	Spanish	Arabic	German	N	N	135		LG203	A	94.9	LG204	A-	90.45	704	700	570	35	29	3.742
2017	0	0	0	3	0	0	0	0		German	Persian	Portuguese	German	Y	Y	135		LG371	A	94.33	LG475	B+	88.58	691	750	730	31	30	3.583

APPENDIX C

Absolute Standardized Mean Difference for Sample Subsets

The sample subsets used in this study were very similar. The figure below shows the means for each statistic after matching reassigned students to students that have not been reassigned. This provides a visual demonstration of how similar each subset was as each black mark indicates the mean standardized difference in each category for both subsets.



APPENDIX D

Logistic Regression Formulas for MLAT and DST Analyses

For this study, probability that an outcome $y=1$, given a set of predictors x was modeled using this general formula:

$$P(y=1|x) = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\beta_5X_5)}}$$

Here, $P(y=1|x)$ represents the probability of a student receiving a grade equal to or above their given threshold. As y approaches 1, the event is more likely to occur; as y approaches 0, the event is less likely to happen (LaValley, 2008). For this study, the closer y is to 1, the more likely the student is to achieve a grade above the given threshold. As y approaches 0, the student is less likely to achieve this grade. β_0 is the intercept, and the coefficients β_i measure the influence of each predictor x_i on the log-odds of the outcome (where i represents a given number associated with both the coefficient and predictor).

For the MLAT analysis, the predictors were first replaced with meaningful variables such as MLAT (Modern Language Aptitude Test score) and dummy variables representing different languages ($x=0$ if the student did not enroll in the language and $x=1$ if they did). MLAT is a continuous variable, while Russian, Persian, Arabic, and Chinese are binary indicators (equal to 1 if the language applies, and 0 otherwise). Each coefficient shows how its predictor changes the log-odds of the outcome, holding everything else constant. After fitting the model with data, the final estimated equation becomes:

$$P(y=1|x) = \frac{1}{1+e^{-(0.936 + 0.029MLAT - 0.0546russian - 0.042persian - 0.084arabic - 0.137chinese - 0.008restrat)}}$$

In this fitted equation, the intercept $-.936$ represents the baseline log-odds when all predictors are zero. The MLAT coefficient, 0.029 , means that each one-point increase in MLAT increases the log-odds of success by 0.029 . Conversely, the language variables have negative coefficients, such as -0.546 for Russian and -1.37 for Chinese, indicating a negative association as these languages decrease the log-odds compared to the baseline category (not being in a strategic language). Finally, the “restrat” variable - indicating whether a student was reassigned into a strategic language - has a small negative effect of -0.008 indicating a negative association with student performance compared to students not being reassigned into a strategic language. Thus, logistic regression provides a probability between 0 and 1 for the outcome, with coefficients interpreted in terms of their effect on the log-odds (and, equivalently, the odds) of expected student performance in their foreign language class.

Similarly, the empirical equations for evaluating the DST utilized the same variables and line of reasoning. The completed equation follows:

$$P(y=1|x) = \frac{1}{1+e^{-(0.6502+ 0.0126DST + 0.2307russian + 0.1868persian + 0.2282arabic + 0.1416chinese - 0.6245restrat)}}$$