NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

STUDENT SUCCESS FACTORS AT DEFENSE LANGUAGE INSTITUTE FOREIGN LANGUAGE CENTER

by

Jonathan Bermudez-Mendez

March 2020

Thesis Advisor:	Jonathan K. Alt
Co-Advisor:	Samuel E. Buttrey
Second Reader:	Colby J. Smithmeyer

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** March 2020	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE
STUDENT SUCCESS FACTORS AT DEFENSE LANGUAGE INSTITUTE FOREIGN LANGUAGE CENTER

5. FUNDING NUMBERS

6. AUTHOR(S) Jonathan Bermudez-Mendez

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release. Distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT (maximum 200 words)

The Defense Language Institute (DLI) is a DOD educational and research institution that provides language instruction in over 16 different languages and dialects to thousands of students annually. DLI implements an immersion program where select students spend time in their third semester immersed in the language and culture that they are studying in an effort to improve proficiency. At the end of a student's course of instruction, DLI administers the Defense Language Proficiency Test (DLPT). The current minimum score to pass the DLPT for all basic program students is L2/R2/S1+, and not all students meet this standard. The director of the National Security Agency (NSA) identified that the L2/R2 standard leaves too large a training gap to meet NSA's operational requirements. DLI has been directed to increase the graduation standard to L2+/R2+, which most students do not currently meet.

We developed four stepwise logistic regression models that could predict a student's probability of success at different stages in the student lifecycle. As a student progresses through the program, performance in advanced language classes was the most significant factor in predicting success. Factors such as DLAB score, prior language experience, and language category proved significant throughout the student lifecycle. We found that, after accounting for selection bias, the immersion program did not significantly contribute to improved DLPT performance.

14. SUBJECT TERMS
DLI, Defense Language Institute, machine learning, statistical model

15. NUMBER OF PAGES
63

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

STUDENT SUCCESS FACTORS AT DEFENSE LANGUAGE INSTITUTE FOREIGN LANGUAGE CENTER

Jonathan Bermudez-Mendez
Lieutenant, United States Navy
BA, Norwich University, 2013

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 2020

Approved by: Jonathan K. Alt
Advisor

Samuel E. Buttrey
Co-Advisor

Colby J. Smithmeyer
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The Defense Language Institute (DLI) is a DOD educational and research institution that provides language instruction in over 16 different languages and dialects to thousands of students annually. DLI implements an immersion program where select students spend time in their third semester immersed in the language and culture that they are studying in an effort to improve proficiency. At the end of a student's course of instruction, DLI administers the Defense Language Proficiency Test (DLPT). The current minimum score to pass the DLPT for all basic program students is L2/R2/S1+, and not all students meet this standard. The director of the National Security Agency (NSA) identified that the L2/R2 standard leaves too large a training gap to meet NSA's operational requirements. DLI has been directed to increase the graduation standard to L2+/R2+, which most students do not currently meet.

We developed four stepwise logistic regression models that could predict a student's probability of success at different stages in the student lifecycle. As a student progresses through the program, performance in advanced language classes was the most significant factor in predicting success. Factors such as DLAB score, prior language experience, and language category proved significant throughout the student lifecycle. We found that, after accounting for selection bias, the immersion program did not significantly contribute to improved DLPT performance.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

CAT	Category
CLA	Cryptologic Language Analyst
CONUS	Continental United States
DLAB	Defense Language Aptitude Battery
DLI	Defense Language Institute
DLIFLC	Defense Language Institute Foreign Language Center
DLPT	Defense Language Proficiency Test
DoD	Department of Defense
FY	Fiscal Year
ILR	Interagency Language Roundtable
k-NN	k-Nearest Neighbors
MOE	Measure of Effectiveness
NN	Neural Network
NPS	Naval Postgraduate School
NSA	National Security Agency
OCONUS	Outside Continental United States
OPI	Oral Proficiency Interview
ROC	Receiver Operating Characteristic
SDB	Student Database
SVM	Support Vector Machine
USCG	United States Coast Guard

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The Defense Language Institute Foreign Language Center (DLIFLC) is the Department of Defense's premier foreign language education and training center. DLIFLC trains the majority of the cryptologic language analysts (CLAs) that do the nation's work of translating and analyzing data and information from foreign language sources. Students must take the Defense Language Proficiency Test (DLPT), which measures foreign language proficiency across two modalities, listening and reading, to graduate. Based on student performance on the DLPT, they are assigned a foreign language proficiency score, on a scale of 0 to 5, based on the Interagency Language Roundtable scale. Currently, the standard for graduation from DLIFLC is attaining at least a 2 in both reading and listening modalities on the DLPT. The Director of the National Security Agency (NSA), through operational experience, has identified that CLAs need to perform a proficiency level of 3 or above in reading and listening. As the primary producer of CLAs, DLIFLC has been directed to increase its graduation standards to at least a 2+ on the DLPT in listening and reading.

This research identified the student factors that contribute to achieving at least a 2+ on the DLPT at four separate milestones in the student lifecycle: the first day of classes, the end of the first of three semesters, the end of the second semester, and the end of the third semester. We examined the effects of the immersion program on student success. The immersion program at DLIFLC selects students, based on their grades, to be immersed in the language and culture they are studying during the third semester in an effort to improve their foreign language proficiency. We analyzed student data taken from fiscal year (FY) 2008 to FY 2018 to fit four logistic regression models that predicted the probability of students achieving at least a 2+ on the listening and reading portions of the DLPT.

The day one model performed poorly at predicting success, but did well at identifying those who would not achieve the new graduation standard. Each subsequent model improved in its accuracy with the addition of end of semester grades. The third semester model proved most accurate at predicting student success on the DLPT.

The models identified the following factors as significant indicators of success throughout the student lifecycle:

- Prior Language Experience: Student with prior language experience as translators, transcribers, or instructors had increased odds of success in every model we fitted. Recruiting students with prior language experience could result in improved graduation rates at the 2+ level.

- Defense Language Aptitude Battery (DLAB): Student DLAB scores remained a significant predictor of success throughout the student lifecycle. The higher the DLAB score, the high the odds of succeeding on the DLPT. This indicates that higher DLAB scores translate into higher odds of success.

- Recycled Students: Recycled students proved to have significantly reduced odds of success. This may indicate that recycling students may not be enough to help them achieve a 2+.

- Category (CAT): Students in CAT 1 languages tended to have lower odds of success compared to students in categories 2, 3 and 4. This could be a result of the lower DLAB requirement to qualify for a CAT 1 language and the shorter length of their curriculum.

- Grades: Grades were the largest predictors of success, especially grades in the second and third semester language classes.

The immersion program was not a significant factor in predicting student success. At first glance, it appeared that students selected for the immersion program benefited from it, achieving a 2+ on the DLPT at rate of 56%, compared to only 34% for those not selected for immersion. Students selected for immersion, however, were already high performing students with high GPAs. When controlled for GPA, students selected for immersion performed no better than students not selected for immersion on the DLPT.

This research provides DLIFLC leadership insights into the factors that influence a student's ability to achieve the new graduation standard. The models allow DLIFLC staff to identify at-risk students early on in the student lifecycle and provide them with the appropriate intervention to get them back on track to succeed on the DLPT. By identifying that the immersion program does not significantly improve student performance on the DLPT, leadership can take a closer look at the program to decide how to improve or change it to better meet its goal of improving performance on the DLPT.

ACKNOWLEDGMENTS

This thesis could not have been accomplished without the help and support of my friends, family, and advisors. Their consistent support truly buoyed my spirits when things became difficult and pushed me through to the finish line. My advisors, Jon Alt and Professor Sam Buttrey, literally made this thesis possible. Thank you both for guiding me onto the right path and giving me the tools and nudges needed to succeed.

To my cohort, thank you all for making me look forward to coming to class. Kelly, Ross, Karl, Lindsey, Clarese, April, Pete, Alex, and everyone else, each of you helped me make it through this program and I am glad to be able to call each of you a close friend.

To my close friend, Jon Harrell, thank you for keeping me grounded, allowing me to bounce ideas off you, always being up for a friendly debate, and for just being a fantastic friend.

To my wife, Sarah Perkins, words cannot begin to express how you have supported me during my time here. Thank you for supporting me, encouraging me, giving me advice and everything else you've done for me. I love you.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

This research seeks to identify the factors that impact a military language student's likelihood of achieving a proficient reading and writing score on the Defense Language Proficiency Test (DLPT) following the completion of a course of study at the Defense Language Institute Foreign Language Center (DLIFLC). DLIFLC requires an understanding of these factors so that it can implement data-driven changes to their language curricula and administrative policies.

A. MISSION OF DLIFLC

DLIFLC's mission is "to provide the highest quality culturally based foreign language education, training, and evaluation to enhance to enhance the national security of the United States" (Defense Language Institute Foreign Language Center 2017, p. 68).

There are currently 16 languages taught at DLIFLC, divided into four categories (CAT) based on difficulty. The length of a course of instruction is tied with the language's category. CAT I and CAT II languages are considered easier languages to learn and have the shortest course of instruction at 36 weeks. CAT III languages are taught in 48 weeks. CAT IV languages are considered the hardest languages to learn and have the longest course of instruction, at 64 weeks. Similar languages are grouped together into schools. There are currently eight schools within the Department of Undergraduate Education (Defense Language Institute Foreign Language Center 2017).

Regardless of language, every curriculum at DLIFLC consists of three semesters, each containing five classes. Three of these classes are focused primarily on language learning and speaking. The other two classes focus on history/geography/culture of the area where the language is spoken, and specific military topics/skills related to that language.

Throughout the course of instruction, students must maintain a "C" average. Students who fail to maintain that standard may either be recycled, restarted, relanguaged (restarting the course of instruction in an easier language), or academically attrited. When students are recycled, they are "returned to an earlier point within the same language program… to a point no later than the onset of problem that caused the recycle" (Defense

Language Institute Foreign Language Center 2017, p. 310). If students get restarted, they start their same language over from the beginning. Students who get relanguaged start a new language from the beginning. This option is reserved for students who have shown an "aptitude for language learning but is unable to learn the most difficult languages… or when the services determine an unanticipated need for a new language" (Defense Language Institute Foreign Language Center 2017, p. 310). If DLIFLC deems that none of these options is appropriate, students are removed from the program and considered an academic attrition.

Once students make it to the end of their course of instruction, they must take the DLPT and Oral Proficiency Interview (OPI). The DLPT is the Department of Defense (DoD) standardized testing system for measuring foreign language proficiency (Department of Defense 2009). The DLPT consists of two portions: reading comprehension and listening comprehension. Based on the student's performance, each portion is assigned an Interagency Language Roundtable (ILR) Language Skill Level. The ILR is a Federal interagency organization that coordinates language-related activities at the Federal level (Interagency Language Roundtable 2019). There are six basic skill levels, ranging from 0 to 5, with each level having a possible "+" modifier, for a total of 12 possible skill levels. DLIFLC's current graduate requirement is attaining a listening and reading skill level of 2 (L2/R2).

B. THE PROBLEM

The Department of Defense requires DLIFLC to increase the graduation standards from L2/R2 to L2+/R2+ by 2023. In an effort to facilitate this change, DLIFLC is interested in identifying factors that contribute to student success, specifically, students successfully achieving at least L2+/R2. We have identified three main issues for analysis:

- What factors contribute to a student's ability to achieve a 2+/2+?
- Does the immersion program improve a student's probability of success?
- Can we construct models that allow DLIFLC staff to understand a student's probability of success at key points of the student life cycle?

II. BACKGROUND AND LITERATURE REVIEW

This chapter discusses the mission of DLIFLC, the immersion program, and the reason for increasing the graduation requirements on the DLPT in greater detail. This chapter will then review various studies pertinent to this study and this subject area in general. Finally, this chapter will discuss the methodology used to guide this analysis.

A. BACKGROUND

1. Mission and Goals of DLIFLC

The DLIFLC must provide quality foreign language instruction to linguists from the military service branches and to select civilian federal employees. DLIFLC designs language instruction in such a way that the students will be able to “understand and interpret meaning and intent within foreign language and culture including value systems, behavioral patterns, institutions, geography, and political, economic and social systems of the areas where the target language is spoken” (Defense Language Institute Foreign Language Center 2019, p. 9).

DLIFLC’s primary mission is to provide trained linguists ready to conduct Cryptologic Language Analyst (CLA) work at the standards and requirements established by the organizations that sent the student to DLIFLC in the first place (Defense Language Institute Foreign Language Center 2019).

Following the events of 9/11, the intelligence community identified a need for better-qualified CLAs. In April of 2002, Lieutenant General Michael Hayden, then Director of the National Security Agency (NSA), issued a memo requiring that all CLAs score at least a L3/R3 on the DLPT. As an organization involved directly in the training of CLAs, DLIFLC’S graduation standard of L2/R2 left too large a training gap to satisfy the new L3/R3 requirement (Department of the Army 2015).

In the years since Lt. Gen. Hayden’s memo, DLIFLC has taken numerous actions to increase improve the quality of instruction in order to produce better qualified linguists. Some of these actions include the recruitment of more instructors, introduction of new instructional materials, a focus on proficiency-oriented teaching methodologies, introduction of team teaching and reduction of faculty-student ratio from 2:10 to 2:6 for difficult languages and to 2:8 for easier languages, and increased support for the immersion program (Defense Language Institute Foreign Language Center 2019).

In May 2016, the DoD Senior Language Authority, in an effort to meet the NSA's L3/R3 proficiency requirement, directed DLIFLC to increase its graduation standards from L2/R2 to L2+/R2+ by the beginning of fiscal year (FY) 2023 (Department of the Army 2015).

2. Immersion Program

In an effort to improve the foreign language proficiency of students, DLIFLC has established an immersion program. This program takes students to sites in countries where their language of study is spoken. In the cases where this is not possible, enclaves within the U.S. where the language predominates are used (Defense Language Institute Foreign Language Center 2017). The immersion typically takes place in the third semester. Students are typically selected based on their classroom performance.

3. DLIFLC Measure of Effectiveness

DLIFLC uses several different measures of effectiveness (MOE) to quantify success. DLIFLC chose academic production as its main MOE to track its progress towards improving DLPT scores. DLIFLC defines academic production as

$$\frac{\text{Total \# Achieved Standard}}{\text{Total Enrollment - Administative Attrition}}\,. \tag{0.1}$$

By this metric, the current L2+/R2+ production rate is about 36%. DLIFLC wants to increase the L2+/R2+ production rate to 64% by FY 2023.

B. LITERATURE REVIEW

1. Previous Work

DLI and the Naval Postgraduate School (NPS) have, over the years, conducted many studies into academic performance and attrition at DLI. These studies developed predictive models of academic success for students in specific languages and explored the effects of gender and other factors on attrition.

2. Korean Academic Attrition at DLIFLC

In the thesis *Analysis of Korean Attrition at the Defense Language Institute Foreign Language Center*, Haupt (2014) explored the factors affecting attrition of students enrolled in the Korean language program. Haupt used data from students ranging from FY 2006–2013 to build eight logistic regression models. These models predicted student success at four academic milestones, the beginning of semesters 1–3 and after graduation but before the DLPT. Through his analysis, he identified several factors that proved significant indicators of at-risk students such as pay grade, service branch, DLAB scores, in status, prior language proficiency and semester GPAs. Haupt states that his models show that a faculty member can predict whether a student will succeed or fail after just the first semester with reasonable accuracy. This accuracy improved with second semester data, but did not improve appreciably with third semester data. Therefore, a reasonable guess can be made after the first semester and a pretty clear picture develops by the end of the second semester.

3. NPS Theses Investigating Graduation Factors

Many students from NPS have assisted DLIFLC in getting a better understanding of how it can utilize their data to improve their course of instructions. Two such theses were *An Analysis of Factors Predicting Graduation of Students at Defense Language Institute Foreign Language Center* (Wong 2004), and *Study of Initial Student Attrition from Defense Language Institute Foreign Language Center* (Anderson 1997).

Both authors were interested in discovering factors that affected attrition and graduation. Wong used student data from FY 1998 through 2003, separating the data into four groups (all students, CAT I students, CAT III students, and CAT IV students). Wong found many significant and interesting interactions among factors for each group. Anderson utilized student data from FY 1994 through 1996 from four languages, one from each category (Spanish for CAT I, German for CAT II, Russian for CAT III, and Arabic for CAT IV). Anderson utilized Binary Classification Trees on two different data sets, one filled with all the students who graduated or academically attrited, another with all the students who graduated or administratively attrited. Both theses found that, among other things, students' DLAB scores and whether they were relanguaged or recycled proved to be significant factors in predicting graduation/attrition. There were also interesting correlations between motivation, gender and service with graduation in some of the models.

C. ANALYSIS METHODOLOGY

We utilize logistic regression to model our data. Logistic regression is a method of regression analysis that is appropriate for use when the response variable is binary (Hilbe 2009). In his guide to machine learning, Kassambara (2018) states the following assumptions when fitting Logistic regression models:

- The response variable is binary.
- The relationship between the logit of the outcome and each predictor is linear.
- There are no influential values among the continuous variables.
- The predictors do not exhibit high multi-collinearity.

We use a training set to train the models and a test set to verify the performance of the models. The training set contains 80% of observation in the data set and the test set contains the remaining 20% of observations. The training and test sets were constructed in a way that the proportion of success and failures were equal between the two sets.

Using the statistical environment R (R Core Team 2019), a binomial logistic regression model was fit to the response variable using only the main effects of the predictor variables in training set. Stepwise logistic regression, an algorithmic method of determining variable importance, was used to identify which variables to include in the final model.

Stepwise regression starts with a model containing all variables of interest. Predictor variables are added or removed sequentially based on which action improves AIC most. Every time a variable is added or removed, the variables currently in the model are reassessed. This process continues until either adding or removing a variable only increases the AIC. We are left with a model containing only the most significant variables that results in the smallest AIC.

After verifying that our model satisfied the assumptions for a logistic regression, we evaluated the resultant model via a classification table. A classification table is an intuitive way to summarize the results of a fitted logistic regression model (Hosmer et al. 2013). Using our model, we predict the probability of success for students in our test set. We classify any student with a probability greater than a specified cut point (such as 0.5) as a success, otherwise, a failure. We then create a classification table which compares our predicted success and failures to the observed successes and failures (see Figure 1).

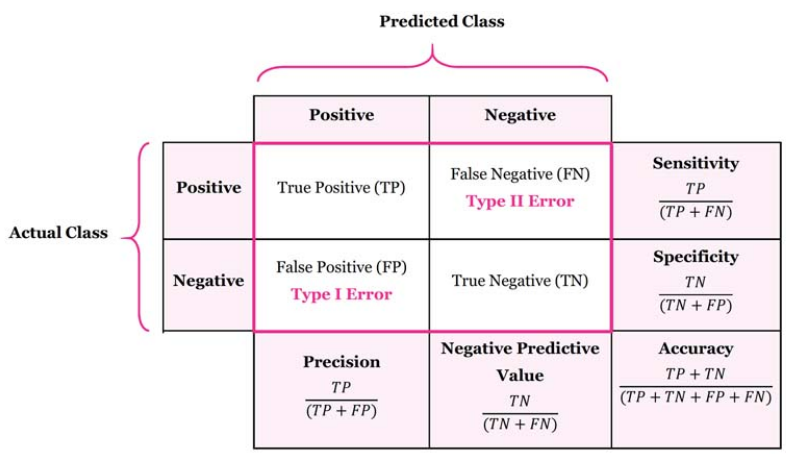

Figure 1. Sample Classification Table. Source: Sirsat M (2019).

A number of important metrics can be derived from the classification table such as overall accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). Overall accuracy tells us how many time we accurately predicted successes and failures. Overall accuracy is defined as

$$\frac{\text{True Negative} + \text{True Postive}}{\text{Total Population}} . \tag{0.2}$$

Sensitivity tells us how often we correctly predicted success for students who scored above L2+/R2+. Sensitivity is defined as

$$\frac{\text{True Positive}}{\text{True Positive + False Negative}}\ . \tag{0.3}$$

Specificity tells us how often we correctly predicted failure for students who scored below L2+/R2+. Specificity is defined as

$$\frac{\text{True Negative}}{\text{True Negative + False Positive}}\ . \tag{0.4}$$

PPV tells us the probability of student actually succeeding if they are predicted to succeed. PPV is defined as

$$\frac{\text{True Positive}}{\text{True Positive + False Positive}}\ . \tag{0.5}$$

NPV tells us the probability of a student actually failing if they are predicted to fail.
NPV is defined as

$$\frac{\text{True Negative}}{\text{True Negative + False Negative}}\ . \tag{0.6}$$

We also examine the Receiver Operating Characteristic (ROC) curve and the area under the curve (AUC). The ROC plots the sensitivity, known as the true positive rate (TPR) and 1-specificity, known as the false positive rate (FPR), for an entire range of threshold cutoff points (Hosmer et al. 2013). The AUC, which ranges from 0.5 to 1, provides a measure of the model's ability to discriminate between students who achieve and L2+/R2+ or greater and those who do not. A model with no ability to discriminate between successes and failures will have an AUC of near 0.5, indicating that the classification ability of the model is no better than a random guess or a coin flip. A model with an AUC of 1 indicates a model that can perfectly classify successes and failures.

Next, we determine the relative importance of the factors in the final model. To do this, we examine the odds ratio associated with each factor included in the final model. The odds ratio tells us the increase or decrease in odds of success for a student exposed to a factor, compared to the odds of success of a student not exposed to that same factor (Hosmer et al. 2013).

We then proceed to utilize a similar process to construct three more models. These models will be used to predict success at three different milestones at DLIFLC. The first milestone is day one. This model will initially contain only the factors that DLIFLC would have available on the first day of classes. The second milestone is the end of the first semester. This model will contain the GPA in the beginner language and culture classes in addition to all factors from the day one model. The third milestone is the end of the second semester. This model adds the intermediate language and culture classes GPAs. Finally, the last milestone is the end of the third semester. This model incorporates every factor available.

THIS PAGE INTENTIONALLY LEFT BLANK

III. DATA DESCRIPTION

A. DATA PREPARATION

DLIFLC'S Directorate of Academic Administration maintains the Student Database (SDB). The data used in this study was prepared and delivered by Mr. Bryan Emerson of the Directorate of Academic Administration. The SDB dataset contains historical data on 26,714 DLIFLC students ranging from FY 2010 to FY 2018. The dataset contains 53 variables describing enrollment and demographic data for each student. The students are given a questionnaire at the beginning of their course of instruction in which they are asked about their prior language and educational experiences. This data, if provided, is also recorded in the SDB.

The population of students of interest are those who will go on to serve in roles as CLAs. This represents the majority of students at DLI, about 82.5%. We removed all students who underwent either academic or administrative attrition because these students never took the DLPT. We removed any Coast Guard members and civilians, as they are not in our population of interest. Next, we filter our data set to keep only the 16 languages currently taught at DLIFLC; see Appendix C for a list of those languages.

Some students have multiple entries in the dataset. This arises in a couple of ways. First, if a student fails out of a language program, it is possible that he or she will either restart the course of instruction for the same language or have to restart at an earlier portion of the course. It is also possible such a student will be relanguaged (restart a new course of instruction for a different, usually easier, language). Secondly, students who make it through the course of instruction but fail to pass the DLPT can become Post-DLPT students, attending an eight-week course to better prepare them for the DLPT. Lastly, it is possible that the same student has attended DLIFLC for different languages over the course of ten years. For students who are recycled, their grades are split between multiple records. We merged these separate records into one that captures their time at DLIFLC.

B. VARIABLE TRANSFORMATIONS

There were two groups of variables that required some extra attention in our dataset: rank and the FL series. The rank variable contains the students' military rank. This variable has 18 levels, representing ranks from E-1 up to O-6. We grouped these 18 rank levels into three rank groups: Junior Enlisted, Enlisted, and Senior Enlisted and Officer. Almost 75% of students in our dataset are Junior Enlisted (E-1, E-2, E-3), with Enlisted (E-4, E-5, E-6) and Senior Enlisted and Officer comprising about 21% and 4%, respectively.

The FL series of variables represent a student's grade in one of the 15 classes that students take at DLIFLC. Grades range from an A down to an F, with P's counting as a Pass. These grades make up 13 levels across 15 variables. We decided to convert those letter grades into their GPA equivalent. See Appendix B for the letter-grade-to-GPA conversion we used.

With these conversions made, we then grouped certain classes together based on their content. FLX01, FLX02 and FLX10 are focused specifically on language skills. FLX20 and FLX40 focus on teaching job skills and the culture and history of the regions in which the language is spoken. Therefore, we grouped the language-focused classes into one group and the culture and history classes into another group. The student's score for each group was the average performance in the classes that make up the group.

C. INDEPENDENT VARIABLES

There were many independent variables in the dataset that were not included in the modeling process. These variables were mostly the questionnaire responses and dates. For a list of all variables contained in the full dataset, refer to Appendix A. Table 1 describes each independent variable that was considered for inclusion in our model, including their levels (if categorical) and range (if numeric).

Table 1. Independent Variables

Name	Symbol	Classification	Description
Service Branch	Svc	Categorical	USA (Army) USN (Navy) USMC (Marine Corps) USAF (Air Force)
Category	Lang.Cat	Categorical	Difficulty of Language: 1 (CAT I) 2 (CAT II) 3 (CAT III) 4 (CAT IV)
DLAB	DLAB	Continuous	Scores from 0 to 159
DLAB Waiver	DLAB_Waiver	Categorical	Y (Yes) N (No)
Rank Group	Rank_Group	Categorical	Junior Enlisted (E-1, E-2, E-3) Enlisted (E-4, E-5. E-6) Sr. Enlisted & Officers (E-7 and above)
Input Status	In_Status	Categorical	I (New Input) J (Relanguaged) P (Post-DLPT) Q (Recycle – Same Course)
Elementary Language Group	FL1XX_Lang_Classes	Categorical	Average grade in FL101, FL102 and FL110
Intermediate Language Group	FL2XX_Lang_Clsses	Categorical	Average grade in FL201, FL202 and FL210
Advanced Language Group.	FL3XX_Lang_Classes	Categorical	Average grade in FL301, FL302 and FL310
Elementary Culture Group	FL1XX_Culture_Classes	Categorical	Average grade in FL120 and FL140
Intermediate Culture Group	FL2XX_Culture_Classes	Categorical	Average grade in FL220 and FL240
Advanced Culture Group	FL3XX_Culture_Classes	Categorical	Average grade in FL320 and FL3240

D. DEPENDENT VARIABLE

The dependent variable of interest in our study is whether or not a student achieved a L2+/R2+ or greater on the DLPT. A column was added indicating if a student reached that threshold. Table 2 describes the dependent variable.

Table 2. Dependent Variable

Name	Symbol	Classification	Description
L2+/R2+ or Greater	L2+/R2+_greater	Categorical	0 (Failed to achieve L2+/R2+) 1 (Achieved L2+/R2+ or greater)

E. DESCRIPTIVE STATISTICS

Considering the size of the dataset, it would be beneficial to conduct some initial descriptive statistics on the data to get a better understanding of the data.

1. Students with Multiple Observations in the Dataset

There are 134 student IDs that appear twice in our dataset, comprising 268 observations. These 134 students attended DLIFLC on two separate occasions over the course of their careers, several years apart and in different languages. The remaining 14,844 observations are of individual students.

2. Distribution of Students Achieving L2+/R2+ or Greater

Of the 14,896 observations in our dataset, 9,096 (61.1%) did not achieve a score of at least L2+/R2+ on the DLPT, and 5,800 (38.9%) did.

3. Students Achieving L2+/R2+ by Service Branch

Figure 2 displays the number of students in the dataset by Service. We can see that, given our exclusion criteria, Air Force students make up the largest group, with 39% of the observations in the data set. The Army comes in second with about 27% of the student population, followed by the Marine Corps and the Navy with 13% and 21%, respectively.

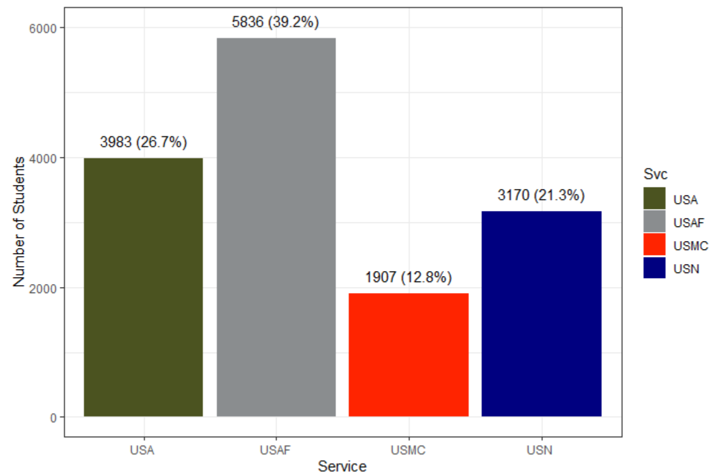

Figure 2. Number of Students by Service

4. Student Distribution by Language Category

Figure 3 displays the distribution of students by language category. We see that the majority of students attending DLIFLC are in CAT 4 languages, with nearly 58% of the observations. CAT 2 languages comprise the smallest number of students with only 166 observations over 10 years. The students in CAT 2 languages have attained an L2+/R2+ or greater on the DLPT at a rate of just over 74%, whereas students in the CAT 1,2, and 3 languages achieved at a rate of 29%, 38% and 41%, respectively. Table 3 summarizes these rates.

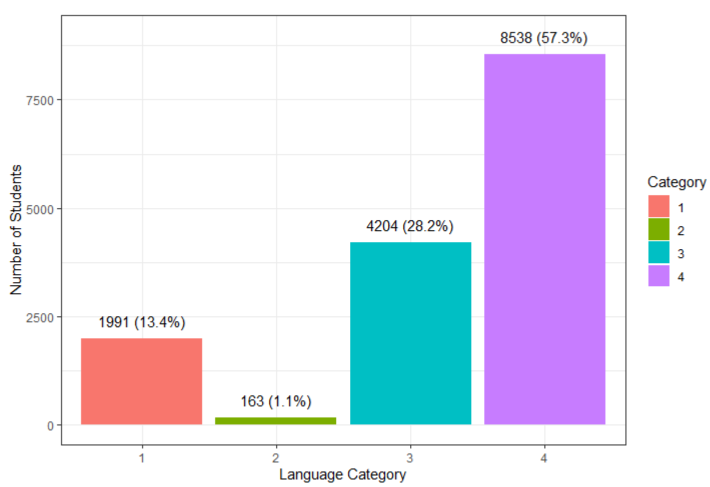

Figure 3. Student Distribution by Language Category

Table 3. Student Success by Language Category

Language Category	L2+/R2+ or Greater	Less than L2+/R2+
1	582 (29.2%)	1409 (70.8%)
2	122 (74.8%)	41 (25.2%)
3	1593 (37.9%)	2611 (62.1%)
4	3503 (41.0%)	5035 (59.0%)

IV. ANALYSIS AND RESULTS

This chapter goes over the diagnostics and results of each model developed using data available to decision makers at four milestones in the student's life cycle: first day of classes, end of the first semester, end of the second semester and end of the third semester. We start with the end of the third semester model.

A. THIRD-SEMESTER MODEL

1. Goodness of Fit

First, we examine the model for goodness of fit. According to Hosmer, Lemeshow, and Sturdivant, a model with an AUC between 0.8 and 0.9 is considered to be excellent at discriminating between successes and failures (2013, p. 177). Our model has an AUC of 0.838, leading us to conclude that the model as specified does an excellent job of discriminating between successes and failures. Figure 4 contains diagnostic plots for the model.

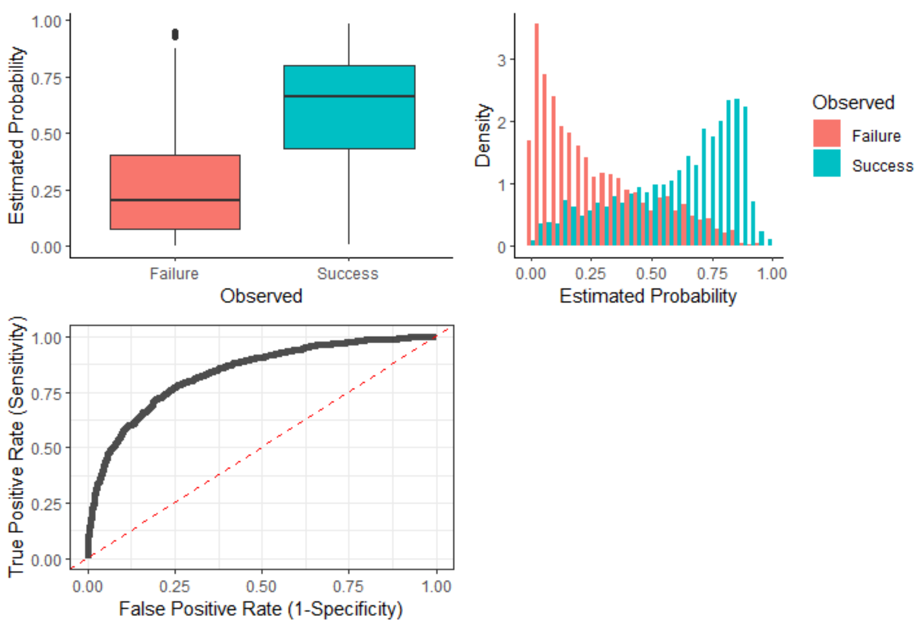

Figure 4. Third Semester Model Diagnostic Plots

The boxplot displays estimated probabilities produced by the model for students who achieved the new standard and those that did not. We can see that the model tends to produce higher estimated probabilities for students who were actually successful and produced lower probabilities for students that were actually unsuccessful. This outcome is exactly what we expect to see from our model, and the clear separation of the boxes is a good sign of model's ability to discriminate between successes and failures.

Plot B is a histogram of the estimated probabilities for successes and failures. The better a model is at discriminating, the further apart these histograms should be. We can see clear differences in the peeks of the two histograms, with the successes having a much higher density at higher estimated probabilities.

Finally, the last plot is of the ROC curve. As stated earlier, with an AUC 0.837, our model can be considered excellent at discriminating between successes and failures.

2. Classification Table

After creating our model, we utilize the test set to evaluate its performance and tabulate the results as a classification table. Table 4 summarizes the results classification table.

Table 4. Classification Table

Predicted/Observed	Failure	Success
Failure	1481	366
Success	333	800

The important metrics derived from the classification are as follows:

- Overall Accuracy: 0.77
- Sensitivity: 0.69
- Specificity: 0.82
- PPV: 0.71
- NPV: 0.80

An overall accuracy of 0.77 means that we correctly predicted a success or failure at a rate of 77%. Considering that a “naïve” person would guess “failure” every time with about 61% accuracy, this is a good result. A sensitivity of 0.69 means that we correctly predict success for a successful student about 70% of the time. A Specificity of 0.82 mean that we correctly predict failure for a student who fails to attain a L2+/R2+ 82% of the time. The PPV tells us that if we predict a student is going to be successful, there is a 71% chance that they actually will be successful. The NPV tells us if we predict a student will fail, there is an 80% chance that they will actually fail.

3. Variable Interpretation

Table 5 summarizes the coefficient estimates from the third semester model.

Table 5. Estimates of Predictor Variables for Third Semester Model

Factor	Estimate	Std. Error
(Intercept)	-17.059	0.404
Lang.Cat2	1.962	0.257
Lang.Cat3	0.308	0.084
Lang.Cat4	0.376	0.081
DLAB	0.011	0.002
In.StatusQ	-0.629	0.087
Prior Lang Exp: Instructor	0.252	0.122
Prior Lang Exp: Transcriber	0.423	0.209
Prior Lang Exp: Translator	0.890	0.201
Rank: Enlisted	-0.104	0.059
FL2XX_Lang	1.306	0.096
FL3XX_Lang	2.382	0.097
FL1XX_Culture	0.175	0.075
FL3XX_Culture	0.443	0.075

We first notice that, through the process of stepwise logistic regression, several variables were considered insignificant based on their contribution to the model, and were dropped. These dropped variables were service, elementary language classes, intermediate culture classes and DLAB waivers.

A summary of each variable's odds ratio, confidence interval and interpretation is provided in Table 6. The odds ratios for categorical factors represent the increase in the odds of success when exposed to particular factor, compared to a student not exposed to that particular factor. For continuous factors, the odds represent the increase/decrease in odds of success for every unit increase for that particular factor.

Table 6. Model Interpretation Summary

Variable	Odds Ratio	Lower 95%	Upper 95%	Summary
Lang.Cat2	7.261	4.416	12.195	CAT 2 students have substantially higher odds of being successful than CAT 1 students.
Lang.Cat3	1.306	1.102	1.548	CAT 3 students have better odds of success than CAT 1 students.
Lang.Cat4	1.410	1.196	1.665	CAT 4 students have better odds of success than CAT 1 and CAT 2 students.
DLAB	1.012	1.007	1.016	Students with higher DLAB scores have better odds of success.
In.StatusQ	0.554	0.463	0.660	Recycled students have worse odds of success than relanguaged and initial entry students.
Prior Language Exp: Instructor	1.287	1.012	1.636	Prior experience as a language instructor improves odds of success.
Prior Language Exp: Transcriber	1.527	1.017	2.306	Prior experience as a transcriber improves odds of success.
Prior Language Exp: Translator	2.435	1.648	3.621	Prior experience as a translator improves odds of success.
FL2XX_Lang	3.617	3.000	4.367	Better grades in intermediate language classes increase odds of success.
FL3XX_Lang	11.121	9.199	13.471	Better grades in advanced language classes increase odds of success substantially more than any other class group.
FL1XX_Culture	1.203	1.038	1.395	Better grades in beginner culture classes increase odds of success.
FL3XX_Culture	1.522	1.314	1.765	Better grades in advanced culture classes increase odds of success more than beginner culture classes.

We see that students of CAT 1 languages, all other things being equal, have lower odds of success on the DLPT than students of the other language categories. This is an interesting finding, considering CAT 1 languages are considered the easiest languages for native English speakers to learn. This result could be related to the shorter course of instruction or lower DLAB standards. CAT 3 and CAT 4 languages have similar odds of success, with CAT 4 edging CAT 3 out just slightly. Although CAT 4 languages are considered more difficult than CAT 3 languages, the improved odds of success for CAT 4 students could be explained by the longer duration of course of instruction and higher DLAB minimum requirements.

B. DAY ONE, FIRST SEMESTER, AND SECOND SEMESTER MODELS

The day one model had an AUC of 0.63, which according to Hosmer et al. suggests poor discrimination. That is to say that the performance of the model essentially performs no better than a coin flip (2013). This is strong indication that whether a student will achieve a L2+/R2+ on the DLPT is not able to be accurately predicted based solely on incoming student data. An interesting finding, however, is that although the day one model has an accuracy of 0.63 and a sensitivity of 0.25, it has a specificity of 0.87. This means that of the students who failed to meet the new DLPT standards, the model correctly identified 87% of those students. Of those students who actually met the new DLPT standard, the model was only able to correctly identify 25%. Figure 5 contains the diagnostic plots for the day one model and Table 7 summarizes the pertinent statistic derived from the classification table.

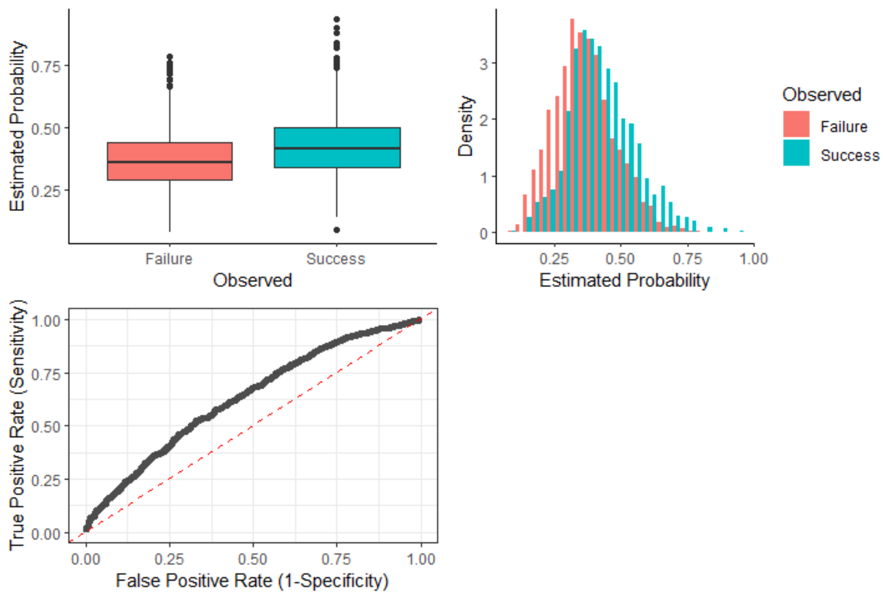

Figure 5. Day One Model Diagnostic Plots

The first semester and second semester models improve their accuracy and sensitivity with the addition of the first and second semester grades. The AUC's for these models are 0.77 and 0.82, respectively. This means that the first semester model can be considered to have acceptable discrimination and the second semester model to have excellent discrimination. These models sequentially improve in accuracy as student performance in the program gets factored in. We can see this increasing ability to discriminate by comparing the diagnostic plots. The main thing to note is the increasing separation between the boxes in the boxplots and the histograms. Figure 6 and 7 contain the diagnostic plots for the first and second semester models, respectively.

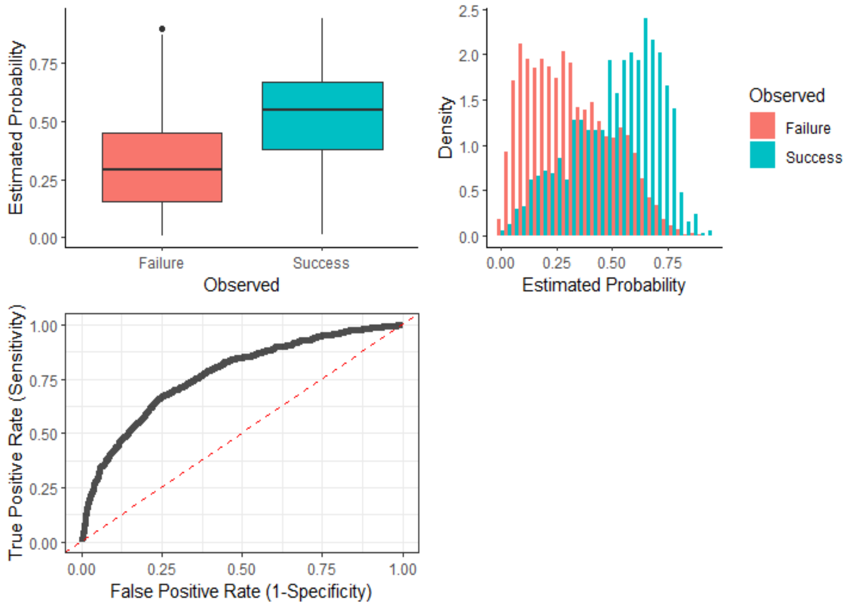

Figure 6. First Semester Model Diagnostic Plots

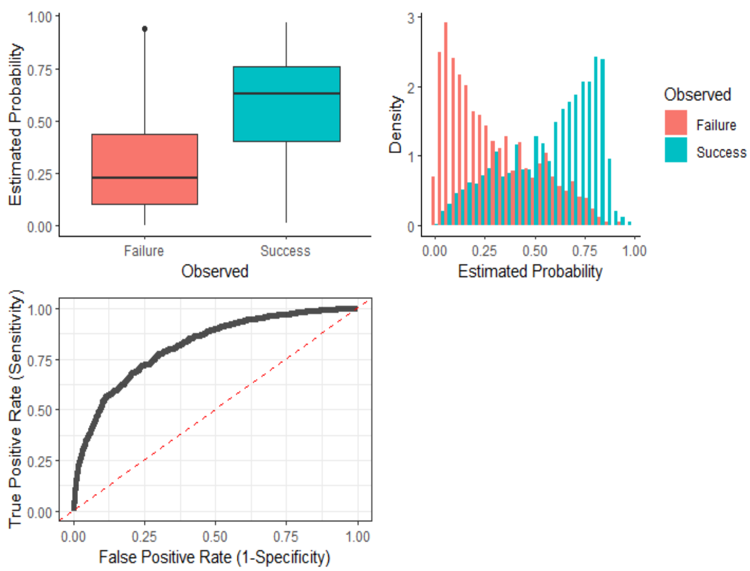

Figure 7. Second Semester Diagnostic Plots

Table 7 lists pertinent metrics for each model.

Table 7. Classification Table Metrics for All Models

	Day One Model	First Semester	Second Semester	Third Semester
Accuracy	63%	72%	75%	77%
Sensitivity	25%	58%	66%	69%
Specificity	87%	80%	81%	82%
PPV	55%	65%	69%	71%
NPV	64%	75%	79%	80%

C. IMMERSION PROGRAM

We first started by looking at the proportion of students who went on immersion and achieved at least an L2+/R2+ compared to the proportion of students who did not go on immersion and achieved at least an L2+R2+. Figure 8 displays the results.

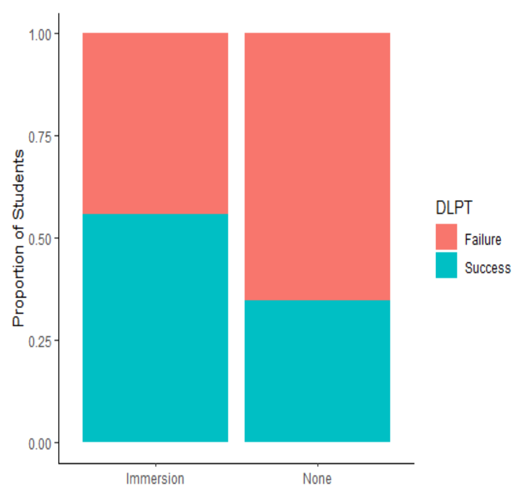

Figure 8. Proportion of Students who Did and Did Not Go on Immersion

We found that 56% of who went on immersion achieved at least L2+/R2+ on the DLPT, compared to only 34% of students who did not go on immersion. A Chi-square test confirmed that these proportions were significantly different. This alone could be considered evidence of the effectiveness of the immersion program; however, we accounted for the fact that students selected for the immersion program are already high performers. Figure 9 displays the distribution of overall GPAs for students who went on immersion and students who did not.

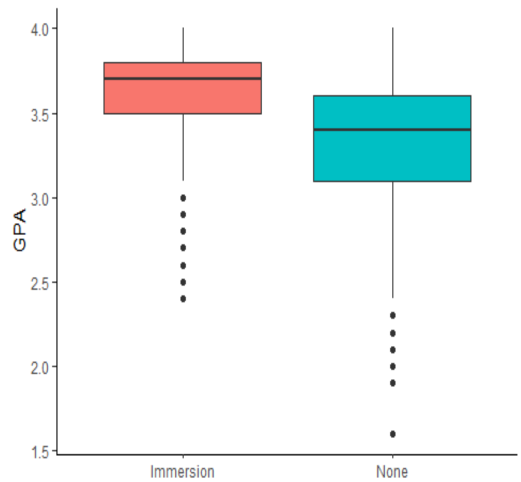

Figure 9. Distribution of GPAs of Students who Did and Did Not Go on Immersion

We can see that students who went on immersion had a much tighter spread and higher median GPA when compared to students who did not go on immersion. It is important to note that the immersion group consists of about 3,000 observations, compared to nearly 12,000 observations in the no immersion group.

To account for the disparity in GPA distribution between the groups, we sampled the non-immersion group to match the distribution of the immersion group's GPA. When we compare the two groups after controlling for GPA disparity, we found that the non-immersion group achieved L2+/R2+ at about the same rate as the immersion group. In fact, a Chi-Square test showed that, now, there was no significant difference in the proportions between the two groups. Figure 10 displays the proportions of success for each group after accounting for GPA disparity.

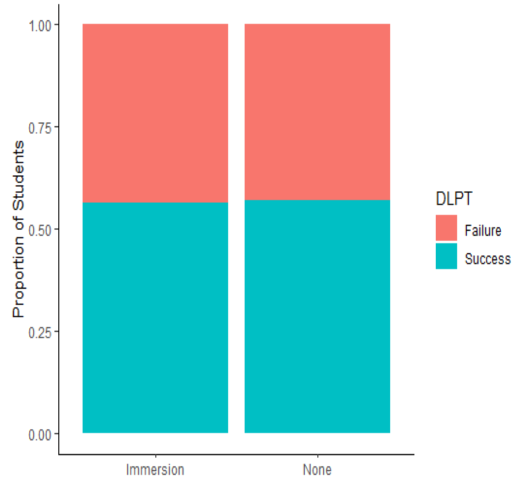

Figure 10. Proportion of Success after Controlling for GPA

V. CONCLUSIONS

A. SUMMARY

The purpose of this thesis was to identify factors that can be used for predicting student success on the DLPT at the L2+/R2+ and above level. Identifying these factors can provide senior leadership at DLIFLC with a better understanding of how to improve policies and procedures to increase the academic production rate of L2+/R2+ to the stated goal of 64% by 2023. We identified three key issues for analysis: which factors contribute to student success, whether the immersion program improves probability of success, and whether we can fit models to help DLIFLC staff understand a student's probability of success at key points of the student life-cycle.

We found that several factors associated with selection remained important throughout the student life cycle. We also found that the further into the student-life cycle, the more important grades become as indicators of success, to the point that by the end of the third semester, grades in the first semester were not significant predictors of success.

Analysis of the immersion program showed that it did not significantly improve a student's probability of success on the DLPT. Although students who went on immersion saw a 56% success rate on the DLPT compared to only 34% for those who did not go on immersion, the students selected for immersion tended to already be high-performing ones. When controlled for GPA, there was no difference in DLPT outcomes between students who went on immersion and students who did not.

We fit four models to the data available at various milestones in the student life cycle. The Day One model had poor ability to discriminate between failures and success, but with the addition of grades in the other models, we were able to fit a model with excellent discrimination and 77% accuracy.

B. FUTURE WORK

This study was only an initial look into success factors at DLIFLC. The immersion program could benefit from further study. Specifically, it might be revealing to look into the effects of immersion on DLPT performance by individual languages to determine if some language programs do benefit from immersion. Another area of future study is the effects of Post-DLPT training. If a student fails to meet DLPT standards they may be given extra time at DLIFLC to help improve the areas in which the student was lacking and retake the DLPT. If Post-DLPT training is effective in increasing a student’s probability of achieving the new DLPT standard, it could be expanded to allow for even more students to benefit. DLIFC administers an end-of-program questionnaire to all students asking them about the perceived difficulty of their instruction, how often they studied, etc. This data can also be incorporated into future studies.

APPENDIX A. DESCRIPTION OF VARIABLES

Name	Symbol	Classification	Description
Service Branch	Svc	Categorical	USA (Army) USN (Navy) USMC (Marine Corps) USAF (Air Force)
Language	Lang	Categorical	XX different language digraphs
Category	Lang.Cat	Categorical	Difficulty of Language: 1 (CAT I) 2 (CAT II) 3 (CAT III) 4 (CAT IV)
DLAB	DLAB	Continuous	Scores from XX to XXX
DLAB Waiver	DLAB_Waiver	Categorical	Y (Yes) N (No)
Gender	Gender	Categorical	M (Male) F (Female)
Rank	Rank	Categorical	Student Rank, E-1 to O-7
Input Status	In_Status	Categorical	I (New Input) J (Relanguaged) P (Post-DLPT) Q (Recycle – Same Course)
Output Status	Out_Status	Categorical	G (Graduate) H (Hold) L (Recycle Out Same Course) Z (Attrition)
Reason for Attrition	Reason	Categorical	* (NA) A or X (Academic Attrition)
GPA	GPA	Continuous	Scale 0.0 to 4.0
DLPT Listening	DLPT.L	Categorical	00 (L0) 06 (L0+) 10 (L1) 16 (L1+) 20 (L2) 26 (L2+) 30 (L3)
DLPT Reading	DLPT.R	Categorical	00 (R0) 06 (R0+) 10 (R1)

Name	Symbol	Classification	Description
			16 (R1+) 20 (R2) 26 (R2+) 30 (R3)
DLPT OPI	OPI.S	Categorical	06 (L0+) 10 (L1) 16 (L1+) 20 (L2) 26 (L2+) 30 (L3)
Years of Service	Yrs.Svc	Numerical	Range 1 - 41
Marital Status	Marital.St	Categorical	*Blank* (No input from Student) S (Single) M (Married)
Education Level	Edu	Categorical	0-1 (Non-High School) 2 (High School) 3 (1 Year College) 4 (2 Years College) 5 (3 Years College) 6 (4 Years College) 7 (Bachelor's Degree) 8 (Master's Degree) 9 (Doctorate) NA (No input from Student)
Motivation	Motive	Categorical	1 (Not Motivated, does not want to study any language) 2 (Not Motivated, prefers another language) 3 (Not My preferred language, but motivated to learn) 4 (Motivated, language is second or third choice) 5 (Motivated, language is first choice)
Prior Language	Prior.Lang	Categorical	130 various languages
Native English Speaker	Native.Eng	Categorical	*Blank* (No Student Response) Y (Yes) N (No)

Name	Symbol	Classification	Description
Native Other Speaker	Native.Oth	Categorical	*Blank* (No Student Response) Y (Yes) N (No)
Birthdate	Birthdate	Numeric	Age Range xx – xx
Prior Lang. Proficiency	Prior.Lang.Prof	Categorical	A (Poor) B (Fair) C (Good) D (Excellent) X (None)
Prior Lang. Source	Prior.Lang.Src	Categorical	A (Civilian School) B (DLI) C (Foreign Residence) D (Home Environment) E (Military – Other than DLI) F (Self Study) X (NA)
Prior Lang. Experience	Prior.Lang.Exp	Categorical	A (Conversation) B (Instructor) C (Interpreter) D (Interrogator) E (Reader) F (Transcriber) G (Translator) X (None)
Language Immersion	Immersion	Categorical	O (OCONUS Immersion) C (CONUS Immersion) U (Immersion location not provided) *Blank* (No Immersion)
Elementary Lang I	FL101	Categorical	Student’s Letter Grade
Elementary Lang II	FL102	Categorical	Student’s Letter Grade
Elementary Convo.	FL110	Categorical	Student’s Letter Grade
Intro to Job Related Skills in Lang.	FL120	Categorical	Student’s Letter Grade
Intro to Lang Culture	FL140	Categorical	Student’s Letter Grade
Intermediate Lang I	FL201	Categorical	Student’s Letter Grade
Intermediate Lang II	FL202	Categorical	Student’s Letter Grade
Intermediate Convo.	FL210	Categorical	Student’s Letter Grade
Intro to Military Topics in Lang.	FL220	Categorical	Student’s Letter Grade

Name	Symbol	Classification	Description
History and Geography of Lang Region	FL240	Categorical	Student's Letter Grade
Advanced Lang I	FL301	Categorical	Student's Letter Grade
Advanced Lang II	FL302	Categorical	Student's Letter Grade
Advanced Convo.	FL310	Categorical	Student's Letter Grade
Comprehensive Military Topics in Lang	FL320	Categorical	Student's Letter Grade
Lang Area/Cultural Studies	FL340	Categorical	Student's Letter Grade

APPENDIX B. LETTER-GRADE-TO-GPA CONVERSION TABLE

Letter Grade	GPA
A	4.0
A-	3.7
B+	3.3
B	3.0
B-	2.7
C+	2.3
C	2.0
C-	1.7
D+	1.3
D	1.0
F	0
P	2.0

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX C. LANGUAGES TAUGHT AT DLIFLC

Language	Digraph	Category
Pushtu – Afghan	PV	4
Arabic – Modern Standard	AD	4
Russian	RU	3
Chinese – Mandarin	CM	4
Persian – Farsi	PF	3
Arabic – Iraqi	DG	4
Spanish	QB	1
Arabic – Egyptian	AE	4
French	FR	1
Arabic – Levantine Syrian	AP	4
Urdu	UR	3
Korean	KP	4
Hebrew – Modern	HE	3
Indonesian	JN	2
Tagolog	TA	3
Japanese	JA	4

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX D. CLASSIFICATION TABLES FOR DAY ONE, FIRST SEMESTER, AND SECOND SEMESTER MODELS

Table 8. Day One Model Classification Table

Predicted/Observed	Failure	Success
Failure	1572	874
Success	242	292

Table 9. End of the First Semester Classification Table

Predicted/Observed	Failure	Success
Failure	1455	487
Success	359	679

Table 10. End of the Second Semester Classification Table

Predicted/Observed	Failure	Success
Failure	1461	398
Success	353	a768

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX E. VARIABLE IMPORTANCE FOR DAY ONE, FIRST SEMESTER, AND SECOND SEMESTER MODELS

Table 11. Day One Model Variable Importance

Variable	Odds Ratio	Lower 95%	Upper 95%
Service: USAF	1.21	1.10	1.34
Service: USMC	1.16	1.01	1.32
Service: USN	1.18	1.04	1.33
Category: 2	7.75	5.12	12.00
Category: 3	1.49	1.28	1.74
Category: 4	1.32	1.14	1.54
DLAB	1.03	1.03	1.04
In Status: Relanguaged	0.62	0.42	0.89
In Status: Recycled	0.47	0.40	0.54
Prior Lang Prof: Excellent	1.42	1.12	1.79
Prior Lang Exp: Conversation	1.18	1.02	1.36
Prior Lang Exp: Instructor	1.29	1.03	1.62
Prior Lang Exp: Transcriber	1.81	1.26	2.60
Prior Lang Exp: Translator	2.11	1.50	2.99
Rank : Senior Enlisted & Officer	1.39	1.12	1.71
Same Language: Yes	1.24	1.03	1.48

Table 12. End of First Semester Model Variable Importance

Variable	Odds Ratio	Lower 95%	Upper 95%
Category: 2	5.19	3.33	8.27
Category: 3	1.19	1.01	1.40
Category: 4	1.39	1.19	1.63
DLAB	1.02	1.01	1.02
In Status: Relanguaged	0.56	0.38	0.82
In Status: Recycled	0.58	0.49	0.68
Prior Lang Exp: Translator	1.96	1.39	2.77
First Semester Lang Classes	8.82	7.70	10.13
First Semester Culture Classes	1.68	1.46	1.92
Same Language: Yes	0.76	0.63	0.92

Table 13. End of Second Semester Model Variable Importance

Variable	Odds Ratio	Lower 95%	Upper 95%
Category: 2	8.09	5.06	13.23
Category: 3	1.34	1.14	1.58
Category: 4	1.47	1.25	1.72
DLAB	1.01	1.01	1.02
In Status: Relanguaged	0.80	0.52	1.20
In Status: Recycled	0.47	0.39	0.55
Prior Lang Exp: Translator	2.16	1.49	3.15
Rank : Senior Enlisted & Officer	0.77	0.61	0.98
First Semester Lang Classes	1.33	1.09	1.55
Second Semester Lang Classes	16.10	13.36	19.43
First Semester Culture Classes	1.17	1.01	1.36
Second Semester Culture Classes	1.61	1.40	1.86

LIST OF REFERENCES

Anderson RE (1997) Study of initial entry student attrition from Defense Language Institute Foreign Language Center. Master's thesis, Naval Postgraduate School, Monterey, CA, http://hdl.handle.net/10945/9027.

Defense Language Institute Foreign Language Center (2019) General Catalog 2019–2020, v10c (Monterey, CA). https://www.dliflc.edu/wp-content/uploads/2018/11/DLIFLC_catalog_2019-20_v10c.pdf.

Defense Language Institute Foreign Language Center (2017) Institutional self-evaluation. Report, DLIFLC, Monterey, CA. https://www.dliflc.edu/wp-content/uploads/2018/01/DLIFLC-Self-Study-December-2017_small.pdf.

Department of the Army (2015) Defense Language Institute Foreign Language Center (DLIFLC) plan to achieve 2+/2+ executive summary. Memorandum, Washington, DC.

Department of Defense (2009) DoD language testing program. DoD Instruction 5160.71, Washington, DC, https://dlnseo.org/sites/default/files/DoDI_5160.71.pdf.

Haupt AC (2014) Analysis of Korean academic attrition at the Defense Language Institute Foreign Language Center. Master's thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA.

Hilbe JM (2009) Logistic Regression Models (Chapman & Hall, Boca Raton, FL).

Hosmer DW, Lemeshow S, Sturdivant RX (2013) Applied Logistic Regression Third Edition (John Wiley & Sons, Hoboken, NJ).

Interagency Language Roundtable (2017) About the ILR. Accessed December 20, 2019, https://www.govtilr.org.

Kassambara A (2018) Machine Learning Essentials: Practical Guide in R, https://stdha.com.

R Core Team (2019) R: A language and environment for statistical computing. R Foundation for Statistical Computing, http://www.R-project.org.

Sirsat M (2019) What is confusion matrix and advanced classification metrics. Accessed January 10 2020, https://manisha-sirsat.blogspot.com/2019/04/confusion-matrix.html.

Wong CH (2004) An analysis of factors predicting graduation of students at Defense Language Institute Foreign Language Center. Master's thesis, Operations Research Department, Naval Postgraduate School, Monterey, CA, http://hdl.handle.net/10945/1296.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California