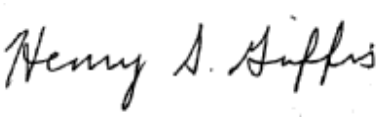
Accession Screening for Language Skills and Abilities

Edward J. Schmitz • Peter H. Stoloff • Jessica S. Wolfanger
Seema Sayala

CRM D0020788.A2/Final
September 2009

CNA
ANALYSIS & SOLUTIONS

Approved for distribution: September 2009



Henry S. Griffis, Director
Defense Workforce Analyses
Resource Analysis Division

This document represents the best opinion of CNA at the time of issue.
It does not necessarily represent the opinion of the Department of the Navy.

Contents

Executive summary

Background

Post-9/11 military operations have brought to the fore the importance of understanding language and culture for our Nation's security. As a result, the Department of Defense (DOD) has undertaken substantial efforts to improve its organic capabilities for developing language skills. One important part of DOD's strategy is to improve its ability to identify military personnel with foreign language capabilities.

Approach

The Defense Language Office asked CNA to investigate DOD's capabilities to screen for language aptitude. We collected information on the current screening practices of the Army, Air Force, Marine Corps, and Navy. We developed a model of the screening process and investigated the outcomes of each stage—from recruiting through completion of language training. We also collected information on officer language aptitude from the Air Force and the Marine Corps.

Based on our data collection, we estimated models that could be used to identify language aptitude for enlisted and officer accessions. We evaluated attributes—usually available in enlistment records—that could be used to identify the best candidates for language training. We were able to identify several measures that could be used to predict both officer and enlisted performance in training.

We used our models to evaluate a variety of screening policies. We tested a range of policies to identify how different screening strategies could be constructed and evaluated by DOD and each of the services. We compared policies of using the Defense Language Aptitude Battery (DLAB), only using recruit attributes in combination with the DLAB, and replacing the DLAB with readily available recruit attributes.

Findings

As in previous research, we found that the DLAB is a useful indicator for performance at the Defense Language Institute Foreign Language Center (DLIFLC). Other factors that are important in explaining student performance include education and motivation. In addition, we found that aspects about training, including the resources provided and the test used in evaluation, are important to control for in assessing student performance.

We examined the services' current screening practices and found that all four used the Armed Services Vocational Aptitude Battery (ASVAB) as the single most important factor in identifying candidates for DLAB testing. However, the services differed in their rates of testing; the characteristics of those screened were also different.

We applied our model for predicting DLAB scores to the FY 2008 accession population to see how different approaches to screening could affect the process of identifying recruits for language training. We found that a pre-screening model that estimates a candidate's DLAB score from other information in the enlistment record could greatly improve the efficiency of testing. Also, we considered whether it would be possible to defer DLAB testing for some high-aptitude recruits in order to reduce the burden on pre-enlistment testing.

We also created a model for officer screening. Using information common in most officers' academic records, we were able to estimate their potential for learning foreign languages. This model could be used for screening officer populations to assess which officers could be trained in a foreign language.

Implications and recommendations

The services can use the model we developed to identify candidates for language aptitude screening. Because it helps determine who is more likely to pass, it has the potential to greatly reduce the numbers of DLAB test takers who fail the test.

Many more recruits would qualify for language training than are currently required. However, we do not have measures of interest in

language training. Additional measures of both recruit interest and the relative priorities of various training slots would be needed for each service to determine if it could improve efficiency in screening with information on recruit interests. If we knew which recruits were interested in learning a foreign language, we would be better able to assess the difficulty of finding qualified candidates for training. Also, the high-aptitude recruits who are candidates for language aptitude screening often are eligible to fill other high-priority training requirements. It would be useful to identify how many of these recruits are interested in foreign languages, but are assigned to other high-priority jobs.

There is potential to improve screening efficiency through use of the predicted DLAB score by waiving the DLAB testing requirement until the candidate is in recruit training. For example, recruits who are predicted to score 110 or above on the DLAB could be accepted into language training conditional on passing the DLAB later.

The model we developed to predict DLAB scores for officers can be used by DOD and the services for general screening for language aptitude. Only those officers who are candidates for particular language training programs would need to take DLAB. Using the officer DLAB prediction model would virtually eliminate any initial screening requirements for officer accessions.

Of the many factors we had to control for, besides student characteristics, the most significant was the version of the proficiency test being used. It will be extremely important to track changes in the testing regimen, as well as in the teaching environment, to understand how to produce sufficient numbers of graduates in the future.

Motivation also appears to be a critical factor. The limited information collected at DLIFLC shows that motivation can be very important for successfully learning a foreign language. Better information on motivation may be particularly useful for language placement. Also, it may be possible to use information on motivation to identify when and how to intervene in a student's training, and to determine whether remedial actions are warranted.

We accomplished the analyses in this project only because we were able to assemble databases from a number of different sources. Important information on testing, assignment, and student characteristics is not readily available. DOD will need to develop a system that tracks information on everyone who takes the DLAB. In addition to test scores and recruit characteristics, the system should contain updates on how people are classified and what type of language training they have obtained. Also, while in this paper we are most concerned with the process from screening through completion of initial training, this database should follow linguists throughout their careers in DOD. These data would help support future analyses on job performance and retention.

Introduction

Post-9/11 military operations have reinforced the importance of language and culture in national security. In 2005, the Department of Defense developed a roadmap that laid out steps to significantly improve its capabilities for developing foreign language skills and cultural competency [1]. As part of achieving this goal, DOD seeks to identify people who have an aptitude to learn foreign languages.

This report describes our analysis of DOD's language screening capabilities. We discuss how the current system identifies language aptitude and provide statistics on recent performance in language training. We investigate how the services recruit people and how those people progress through testing and assignment to attendance at the Defense Language Institute Foreign Language Center. We identify several criteria that can be used to assess their performance at DLIFLC.

In our analysis, we develop models for predicting enlisted performance on the Defense Language Aptitude Battery screening test. Building on previous analytical work, we expand the list of predictors to include several other background characteristics that we found to be significantly related to performance on the DLAB test.

We modeled the performance of both officer and enlisted students at DLIFLC. In addition to comparing the relationship of student characteristics with performance at DLIFLC, we investigated other factors, such as teaching resources, that could also affect outcomes, including completion rates and scores on language proficiency tests. This analysis demonstrates some of the ways that screening of candidates for DLIFLC could be improved.

We also developed a screening model for officers. We looked at data from both the Marine Corps and the Air Force to identify common background factors that could be related to DLAB scores. Using data

from the Air Force Reserve Officers' Training Corps (ROTC) program, we tested how background characteristics relate to DLAB scores and developed a screening model to identify candidates for language training.

We also performed an analysis of different screening policies that could be applied to enlisted recruiting. We estimated models for each of the four services of how they have been selecting candidates for testing. We then compared a screening approach based on our model for predicting DLAB scores. We calculated the number of people that would need to be screened under different scenarios to produce an equivalent number of successful graduates from DLIFLC. This analysis can provide policy-makers across the services with an assessment of the relative impact on screening and training that could be produced through different screening policies.

In the last section of this report, we discuss how DOD and the services can use our analysis to improve the screening process in the future. While we have developed screening models that can be used today to estimate the language aptitude of all enlisted and officer personnel in DOD, several additional steps can be taken. It is important to investigate further how student abilities, language assignment policies, and training resources can be used more efficiently.

We also believe that a tracking system for language screening information would be helpful. Such a system would serve three purposes:

1. It would provide a list of immediate candidates for language training, should there be a need to surge the number of students.
2. It would serve as a database for future studies on screening and performance.
3. It can serve as a source of information for tracking and reporting on language screening activities to Congress.

Background

DOD's strategy in military operations includes the goal of improving its capabilities with respect to foreign languages and dialects that have recently emerged as important for national security. As part of this strategy, DOD issued the Defense Language Transformation Roadmap in February 2005 [1]. The roadmap addresses four major goals, one of which is to create a foundation of language and cultural expertise throughout the military.

Specifically, one of the actions being pursued by DOD calls for improving the testing system across the Defense Language Program to increase the pool of potential language personnel. Accordingly, the Defense Language Office (DLO) has the lead in establishing guidance for improving language testing within DOD.

To establish new screening and testing policies, however, one must first examine the current testing system. We do that in this paper. In addition, we model the process, predict DLAB scores for accessions, and look at the costs and benefits of possible screening alternatives. Finally, we discuss the need to improve data collection efforts to support future analysis.

This page intentionally left blank.

Current practices for identifying and developing language skills

The process of developing linguists is one of the most challenging personnel development processes for enlisted recruits in the military. In this section, we lay out a model of the linguist development process and explain how each service implements the screening process. We provide descriptive statistics on recent results from the screening process. We also describe the flow of personnel into and through the DLI, as well as a picture of the outcomes that result from DLI training.

A model of the screening and development process

The identification and development of military linguists is a long and challenging process. Figure 1 lays out several of the key steps in the screening and training process. In addition to the standard aptitude, physical, and moral qualifications, each candidate must meet the service's Armed Services Vocational Aptitude Battery composite. Table 1 shows the ASVAB composites used by each service for language aptitude screening. All services use the verbal composite, which is also part of the Armed Forces Qualification Test (AFQT) score. All services except the Navy use arithmetic reasoning, which is another component of the AFQT. The Army currently uses a composite that is based on regression-determined weights calculated against all ASVAB composites. The Navy uses math knowledge and general science, while the Marine Corps includes mechanical comprehension in its composites.

Figure 1. Linguist development process

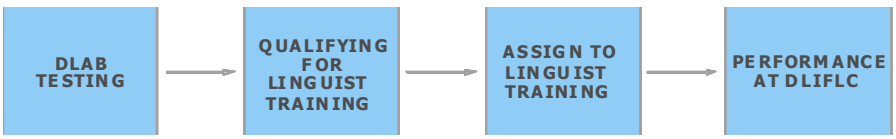


Table 1. ASVAB subtests used for language screening

	Verbal[a]	Arithmetic Reasoning	Math Knowledge	General Science	Mechanical Comprehension	Auto-Shop	Electronics Information
AFQT	X	X	X				
Army	X	X	X	X	X	X	X
Air Force	X	X					
Navy	X		X	X			
Marine Corps	X	X			X		

a. Verbal is calculated by combining the word knowledge and paragraph comprehension subtests.

While some of the qualifications for military linguist training can be ascertained as part of the general military application, the most challenging part of the process is qualifying on a separate test, the Defense Language Aptitude Battery (DLAB). The DLAB takes 1.5 hours and usually requires a separate trip to the Military Entrance Processing Station (MEPS). This involves finding recruits who are interested in becoming linguists and who have the aptitude to pass the DLAB.

The second part of the screening process is passing the DLAB. While only high-aptitude recruits are sent for DLAB testing, the historical qualification rate is under 50 percent. So the services must identify and test large numbers of high-aptitude recruits in order to find sufficient numbers who qualify on the DLAB.

After recruits have been tested and pass the DLAB, they are classified into linguist positions, which involves further processing by each service. Most candidates who qualify are likely to be assigned to linguist training, but this is not always the case. If a high-aptitude recruit decides against accepting linguist training, the service will usually accommodate them into another training program. Also, a significant number of language training candidates fail to complete initial recruit training.

The process of training linguists continues at the Defense Language Institute Foreign Language Center in Monterey, CA. Here the students enter training programs that last up to 15 months. Many students fail to complete the training. Others may complete the course, but they fail to achieve the final course grade requirement.

What is the DLAB?

DOD uses the DLAB to test a person's potential for learning a foreign language and to determine who may pursue training as a military linguist. Half the test is audio, and half is written. The test requires at least 1.5 hours to administer, must be proctored, and is given at the MEPSs on selected days only. The test does not attempt to gauge a person's ability in a language, but rather to determine his or her ability to learn a language.

As defined by DLIFLC, the languages are broken into categories, and the minimum DLAB score rises from Category I to Category IV based on difficulty level for a native English speaker [2]:

- Category I: 95 or better required on DLAB (French, Italian, Portuguese, and Spanish)
- Category II: 100 or better required on DLAB (German)
- Category III: 105 or better required on DLAB (Hebrew, Hindi, Kurmanji, Pashto, Persian-Farsi, Persian Afghan Dari, Russian, Serbian, Croatian, Sorani, Tagalog, Thai, Turkic, Uzbek, and Urdu)
- Category IV: 110 or better required on DLAB (Arabic, Chinese, Japanese, and Korean).

The DLAB is used to identify language aptitude for both enlisted and officer personnel. It is typically administered to new and prospective recruits at the MEPS sometime after the ASVAB is taken but before a final career field is determined. Recruits may take the DLAB if they score high enough on the ASVAB cut score determined by their individual service for linguist training and are interested in doing so.

In FY 2008, the average DLAB test taker had an AFQT score of 85 and scored 98 on the DLAB. In some cases, waivers are granted for accessions that do not meet the minimum scores set by the services. The DLAB is also administered to some ROTC members while they are still attending college and to some service members at the academies.

Military personnel interested in retraining into a linguist field typically also must pass the DLAB. In some cases, the DLAB requirement may be waived if proficiency in a foreign language is already demonstrated via the Defense Language Proficiency Test (DLPT).

Current approach

Recruiting and screening

The services have a requirement to train 3,000 or more new linguists annually. Most of these people are recruited specifically for this training. Because of cost and time constraints, it is impractical to test all DOD accessions with the DLAB (over 100,000 recruits annually).

To narrow the candidate pool, various composites from the ASVAB are used as a pre-screen to determine which recruits should be tested with the DLAB. Each service uses its own ASVAB composite for this purpose.

In addition to ASVAB, citizenship is also used as a screening criterion. Most linguists work in top secret environments, which eliminates non-citizens from eligibility for assignment. Many applicants who have an aptitude for another language or already speak another language are unable to enlist for the linguist military occupational specialty (MOS) because of their citizenship status. Hence, the services usually must rely on recruiting people with no prior knowledge of a language to become proficient in it.

In this section, we look at pre-screening and testing for the language field for each of the four services as it is practiced today.

Navy

For a Navy recruit to attend DLIFLC and become a Cryptologic Technicians Interpretive (CTI) (the Navy's linguists), all of the following requirements must be met:

1. **ASVAB score requirement:** Verbal (VE) + Mathematics Knowledge (MK) + General Science (GS) = 165

2. **Security clearance requirement:** Top Secret
3. **DLAB requirement:** 100.

The Navy uses the foregoing ASVAB composite as a pre-screen to determine who should be administered the DLAB and who is potentially qualified for language study at DLIFLC. When a Navy applicant enters the MEPS, the applicant's score is fed into the classifier's system, and, based on that score, a list of potential jobs for that applicant is generated. The Navy classifier then determines which of these jobs has available openings.

If a CTI position is available and the applicant has an interest in this field, classifiers may send the applicant to take the DLAB if it's being offered that day at the MEPS, or they may choose to bring the applicant back to the MEPS if it isn't being offered that day.

In addition, to qualify for the CTI position, the applicant must be able to obtain a security clearance. In most cases, if a recruit cannot receive a Top Secret clearance, there is no language career field for him or her.

Navy applicants need a score of 100 or higher on the DLAB to qualify to be a CTI. Although the Navy does not typically issue waivers for scores below 100 on the DLAB, there are exceptions. For example, an applicant who already speaks other languages and scores in the high 90s may be waived.

According to Navy Recruiting Command, the DLAB is currently only offered 1 day a week at certain MEPSs. Because of this, applicants will often be given an initial rating other than CTI but will be reclassified as a CTI if they are very interested in that job, are brought back to test, and are successful in meeting the minimum DLAB score.

In FY 2008, the Navy awarded recruits accepted into the CTI program with bonuses of up to $16,000. Only recruits in special operations and the nuclear field programs received larger bonuses.

In addition to screening candidates for DLAB testing, the Navy also screens recruits for heritage language skills. Recruits self-assess their

skill in languages other than English. These data are recorded in the enlisted master file.

Marine Corps

For a Marine recruit to attend DLIFLC and become a Cryptologic Linguist, the following requirements must all be met:

1. **ASVAB score requirement:** General Technical (GT) Score: Verbal (VE) + Arithmetic Reasoning (AR) + Mechanical Comprehension (MC) = 105
2. **Security clearance requirement:** Top Secret
3. **DLAB requirement:** 100.

During the initial screening process (before contract/enlistment), Marine recruiters determine the aptitude of potential Marines from their ASVAB scores. As in the Navy, the DLAB is administered at the respective MEPSs to those who show interest and meet the linguist program's required scores.

The Marine Corps has two intelligence programs consisting of both linguist and non-linguist intelligence positions. The intelligence positions include crypto-linguists and translators who are required to take the DLAB. The non-linguist intelligence program does not require the DLAB. Both programs have a 5-year enlistment term. The bonuses for these programs differ substantially, however. Those classified as linguist intelligence can receive a bonus of up to $25,000. This bonus was added in FY 2009 and is the largest bonus paid by the Marine Corps. Those who enter the non-linguist intelligence program can receive a bonus of $6,000.

The Marine Corps screens recruits for foreign language proficiency. If the recruit self-reports knowledge of a foreign language, it is recorded on the DD1966 form.

Air Force

To qualify to attend DLIFLC and become a Cryptologic Linguist in the Air Force, the following requirements must all be met:

1. **ASVAB score requirement:** General (G) Score: Verbal (VE) + Arithmetic Reasoning (AR) = 72
2. **Security clearance requirement:** Top Secret
3. **DLAB requirement:** 100.

The Air Force actively recruits entry-level personnel into the two Cryptologic Linguist career fields. Potential Cryptologic Linguists must score a minimum of 72 on the Word Knowledge, Paragraph Comprehension, and Arithmetic Reasoning subtests of the ASVAB to be considered for this duty. Personnel who score at least 72 and who wish to enter the Cryptologic Linguist career field are then administered the DLAB. Like the other services, the minimum DLAB score to enter the field is 100.

The Air Force has two types of linguist positions, airborne linguists and ground linguists. Airborne cryptologic linguists must meet additional physical requirements in addition to the testing requirements noted above.

The Air Force reports no problems with the number of days the DLAB is offered at the MEPS. They note that there is currently a waiting list for applicants who are qualified and waiting to become linguists. In this case, the applicant signs a contract and waits for a linguist position to open.

Recruiting bonuses are offered to those who qualify for the Cryptologic Linguist career fields. To encourage recruits to enlist for 6 years, the Air Force offers a $12,000 signing bonus, while 4-year enlistees are paid $3,000. These enlistment bonuses are not paid to heritage/native speakers or already trained linguists.

The Air Force also asks all recruits about their foreign language proficiency. Recruits may list up to two foreign languages in which they are proficient. This information is recorded on the SF86 form.

Army

To qualify to attend DLIFLC and become a linguist in the Army, all of the following requirements must be met:

1. **ASVAB score requirement:** Skilled Technical Score = 110. A combination of all subtests with regression-determined weights is denoted Skilled Technical Score.
2. **Security clearance requirement:** Top Secret
3. **DLAB requirement:** 100.

The Army currently requires a 110 Skilled Technical score for applicants to join the Delayed Entry Program (DEP) for 35W (an Army linguist rating). These applicants must return to the MEPS within 14 days and take the DLAB. If the applicant scores below a 110, he or she can still enlist for 35W but must have a passing DLAB score before enlistment. Language skills in the Army can qualify for an enlistment bonus of up to $40,000.

According to Army recruiting, the DLAB test-failure rate in the Army is very high. If the applicant fails the DLAB, he or she must be sold on another MOS—a task that can be difficult. As a result, there are significant numbers of applicants in the Army who decide not to choose another MOS and drop out of the DEP completely. Currently, the Army is getting waivers approved for DLAB scores of 95 to 99. The Army feels that, without these waivers, it would be very difficult to meet the requirements for this MOS.

The Army asks all recruits: “Do you speak a foreign language?” If they answer “yes,” the response is captured on the DD1966 form. As in the other services, the linguist MOS in the Army requires a Top Secret clearance. According to Army Recruiting, the failure rate for Top Secret clearance interviews is currently about 7 percent. Many applicants who speak another language are unable to enlist in the linguist MOS because of their citizenship status.

The Army informed us that there are many applicants who score well on the ASVAB and have qualifying Skilled Technical scores but fail the DLAB. It feels that applicants often lose interest in the test and give up.

The Army also tests special operations recruits with the DLAB. These recruits attend other schools for language study, however, and do not attend DLIFLC [3].

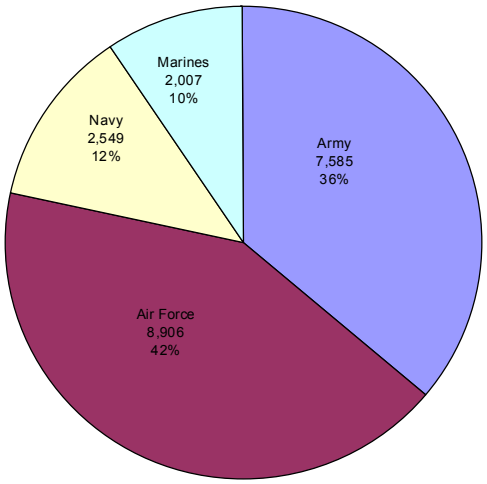
Testing

The first step in the enlistment screening process (refer back to figure 1) is to identify candidates for DLAB testing. We obtained data from the Defense Manpower Data Center (DMDC) and the Military Enlistment Processing Command (MEPCOM) on DLAB testing from 2005 to 2008. Here we provide some background on recent service testing.

By service

Between these years, 21,047 service members were tested with the DLAB. Of those tested, the Air Force accounted for 42 percent, followed by the Army at 36 percent. The Navy accounted for 12 percent of those tested and the Marines for 10 percent (see figure 2).

Figure 2. Percentage tested by service (FY 2005–2008)



Each service tested a percentage of total accessions from 2005 to 2008. Table 2 displays total accessions for each of the services by year compared with the total number of DLAB takers. The Air Force tested the largest percentage of total accessions, followed by the Army. The Navy and Marines test the smallest percentage of their total accessions a year.

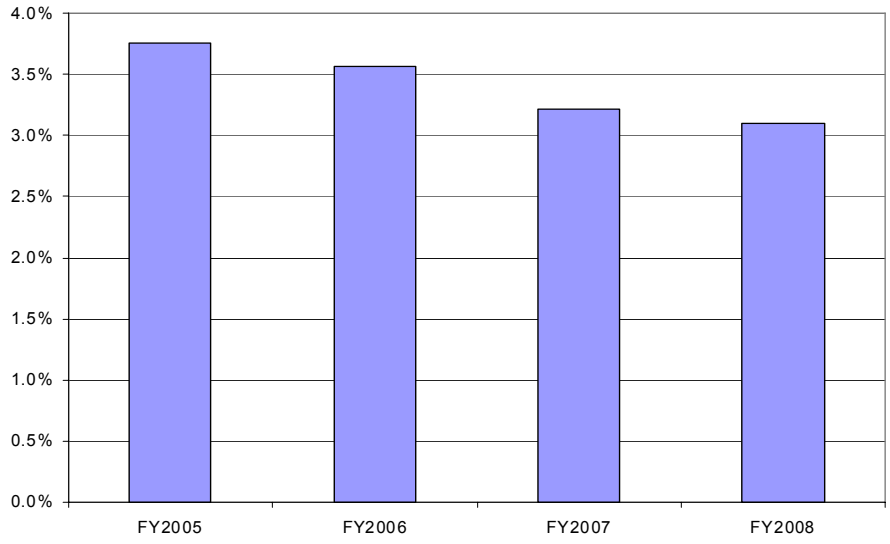
Overall, a relatively small percentage of accessions is currently given the DLAB. The overall percentage of DOD accessions taking the

DLAB decreased from approximately 3.8 percent in FY 2005 to 3.1 percent in FY 2008, despite the requirements for language skills in DOD increasing during this time period (figure 3).

Table 2. Total accessions and DLAB takers by service (FY 2005–2008)

	2005	2006	2007	2008
Army	63,000	69,758	54,085	79,889
DLAB	2,783	1,779	1,107	1,916
Percentage	4%	3%	2%	2%
Air Force	19,902	30,429	23,494	21,897
DLAB	1,953	3,196	2,191	1,566
Percentage	10%	11%	9%	7%
Navy	37,729	35,840	29,706	29,264
DLAB	507	558	634	850
Percentage	1%	2%	3%	3%
Marines	32,016	31,362	29,730	27,804
DLAB	467	448	477	615
Percentage	1%	1%	2%	2%

Figure 3. Percentage of accessions with DLAB score (FY 2005–2008)



The Armed Forces Qualification Test

The AFQT score is computed from ASVAB subtests and is used to determine eligibility for military service. AFQT is also used to identify candidates for DLAB testing.

The four areas of the ASVAB used to compute the AFQT score are Word Knowledge (WK), Paragraph Comprehension (PC), Arithmetic Reasoning (AR), and Mathematics Knowledge (MK). To determine the AFQT "raw score," the Verbal Expression (VE) score must first be computed. The VE score is determined by adding the raw scores from the PC and WK tests and using a table to get the VE score from that combined PC and WK raw score. AFQT is computed using the formula = 2VE + AR + MK. This score is then compared to another chart to determine an overall percentile score.

The recruiters we spoke with indicated that they look at the various ASVAB composites or the AFQT to determine who will likely do well on the DLAB. The data support this and indicate that accessions with AFQT scores of 80 or greater (top quintile) are tested at the highest rates in each of the services. As noted in table 3, the majority of DLAB takers in each service have an AFQT of 80 or higher, while very few DLAB takers score below 50 on AFQT. We discuss the significance of AFQT in predicting DLAB scores in greater detail in the policy analysis section of the paper.

Gender

In all services except the Army, women represent a disproportionate share of DLAB test takers. This may be caused in part by the competition for people with high AFQT scores. The services prioritize selection into many military occupations based on aptitude scores. That is, people meeting the ASVAB pre-screen and passing the DLAB will tend to have high AFQT scores. To the extent that the services assign a lower priority to selecting an individual for a linguist occupation than some other occupation, despite "passing" the DLAB, that person may not be available for linguist training.

Table 3. AFQT scores of DLAB takers (FY 2005–2008)

	<50	50-65	66-79	80+
Army				
2005	20	188	749	1,811
2006	10	115	419	1,197
2007	18	77	265	741
2008	14	136	518	1,248
Air Force				
2005	0	13	442	1,497
2006	1	28	839	2,327
2007	0	19	463	1,705
2008	0	6	331	1,229
Navy				
2005	2	29	111	365
2006	3	44	132	377
2007	7	57	184	385
2008	7	63	237	543
Marines				
2005	5	30	106	325
2006	7	32	92	317
2007	7	34	95	339
2008	1	41	129	444

For example, 41 percent of DLAB takers in the Navy in 2008 were women, compared with 19 percent of recruits. This is likely related to the competition for men with high aptitude scores in other careers, such as the nuclear field. Since the number of women permitted to enlist in the nuclear field is restricted in the Navy, those with higher scores are more likely to end up in a career such as the linguist field (table 4). We will discuss the impact of gender on the probability of becoming a linguist in greater detail later in this paper.

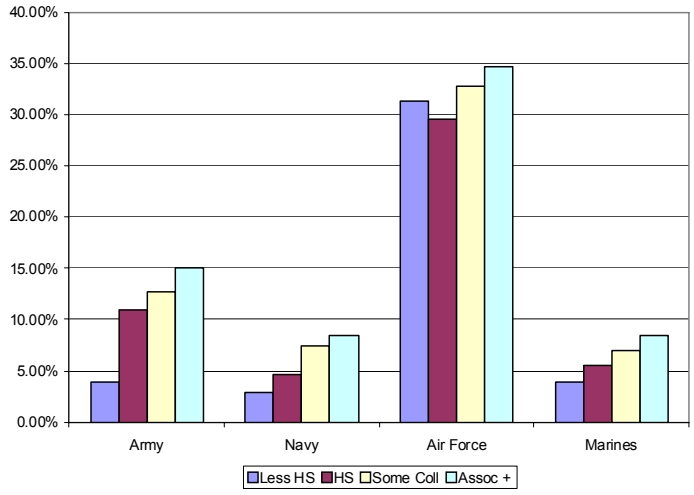
Table 4. DLAB takers by gender (2008)

	Women (%)	Men (%)
Army	15	85
Air Force	31	69
Navy	41	59
Marines	17	83

Education

Next, we looked at the education level and AFQT scores of those accessions taking the DLAB. The education categories examined were “Less than High School Degree,” “High School Graduate,” “Some College,” and “Associate Degree or Better.” The results varied significantly by service. The Air Force tests a high percentage of accessions with high AFQT scores (>80) regardless of education level. The Army, Navy, and Marine Corps, however, test accessions who score 80 and greater on the AFQT at a higher rate based on their education level (figure 4). Again, later in the paper (i.e., the analysis portion), we will look in greater detail at the impact of education level on the probability of becoming a linguist.

Figure 4. Percentage tested by education (AFQT>=80)

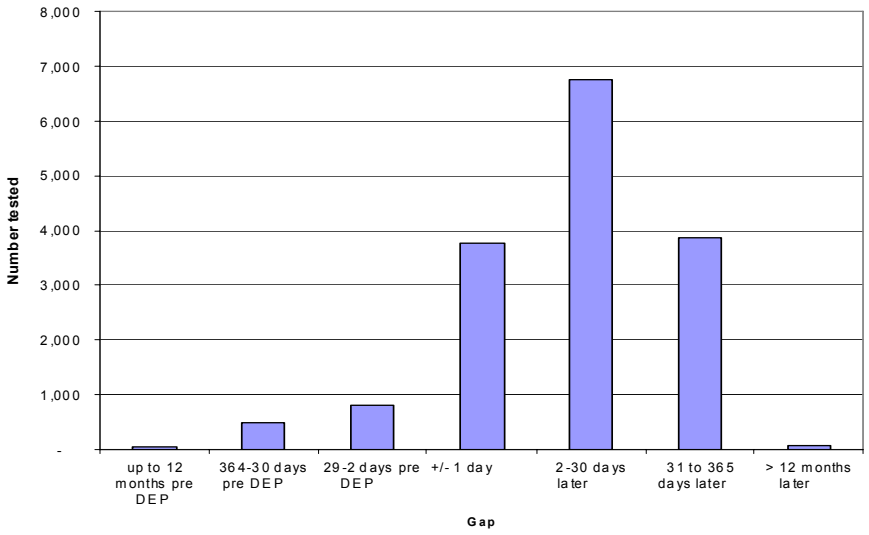


Test scheduling

An important consideration in the screening process is when the DLAB test is taken. A common concern with the services is that testing usually requires a second trip to the MEPS, which adds an expense and burden to the recruiters.

We examined when recruits were tested relative to when they entered the Delayed Entry Program. Figure 5 shows the distribution of when recruits took the DLAB, confirming the concerns of the services. The most common occurrence was that recruits took the test between 2 and 30 days after entering the DEP, while the next highest number took the DLAB 1 month or more after entering the DEP. The next highest number of recruits who took the DLAB did so within 1 day (+/-) of entering the DEP.

Figure 5. DLAB test date vs. DEP entry data (2004–2006)



Qualifying for DLIFLC

Scoring 100 or greater on the DLAB has been the general requirement for attending DLIFLC since FY 2007 when the minimum score was raised. In 2005 and 2006, 43 percent of those given the DLAB scored 100 or better. By 2008, 47 percent of those taking the DLAB scored 100 or better, indicating that the standard remained difficult to achieve.

DLAB scores for linguists

We identified the training enlistment specialty from MEPCOM data for Army and Air Force recruits and from Navy and Marine Corps accession records to determine the eventual occupation of those who had recorded DLAB scores. We looked at how many of those who scored less than 100, 100 or better, and 110 or better on the DLAB in FY 2008 ended up in training for a linguist rating in each of the services.

Our results indicate that, for each service, the majority of those who score 100 or better on the DLAB do end up in a linguist rating. Table 5 displays the results. The Army has the largest percentage of linguists with scores below 100—33.6 percent—indicating that they grant waivers to a third of their test takers. The Air Force, however, did not grant any waivers according to the data. In addition, between 59 and 65 percent of test takers who enter linguist training in the Air Force, Navy, and Marines score 110 or better, indicating that a high percentage of test takers in each of these services qualify for Category III and IV languages.

Table 5. Linguists' DLAB scores (2008)

	Linguists scoring <100 (%)	Linguists scoring 100+ (%)	Linguists scoring 110+ (%)
Army	33.6	66.4	45.7
Air Force	0	100.0	58.9
Navy	2.3	97.7	60.3
Marines	0.9	99.1	65.4

Who attends DLIFLC?

Once the services have determined who is assigned to become a linguist and they complete recruit training, these personnel are sent to DLIFLC in Monterrey, CA. DLIFLC is a DOD school established for the purpose of teaching armed forces personnel foreign languages. All required costs are paid by DLIFLC from its mission funds or by the sponsoring agency.

DLIFLC does not have a standard school year because classes are scheduled to respond to customer demands. Classes begin and end on a continual basis throughout the calendar year.

Prior to the first day of each fiscal year, DLIFLC computes the student quotas requested by the various services and agencies to be taught in each language. From these quotas, classes in each language are scheduled for the year. In general, classes are conducted 6 hours a day, 5 days a week [2].

Service attendance

According to data from 2003 to 2006 provided by DLIFLC, the services send varying numbers of enlisted personnel and officers to the school to meet service-specific requirements. Table 6 displays DLIFLC attendance numbers by service for FY 2003 through FY 2006.

Table 6. DLIFLC attendance by service (FY 2003–2006)

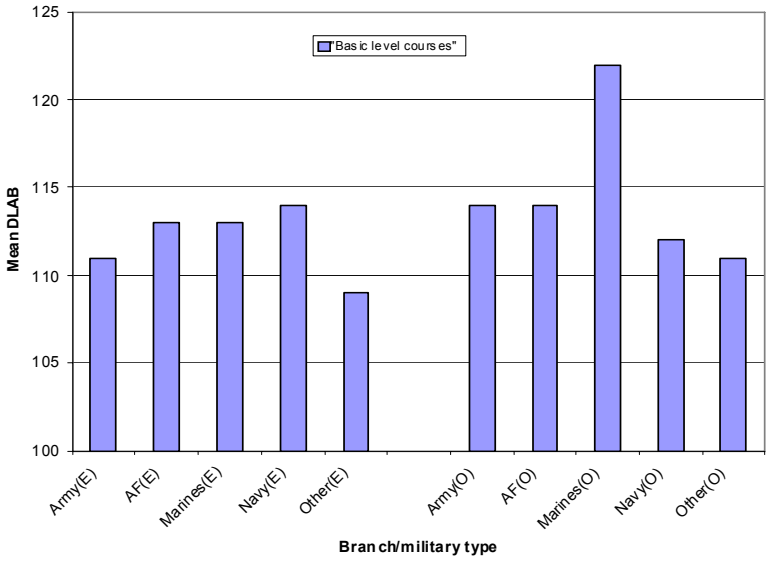
Service	Enlisted	Officer
Army	2,693	133
Air Force	3,350	119
Navy	1,040	123
Marines	705	46

While DLIFLC trains both officers and enlisted personnel, enlisted personnel make up the majority of the student body. Only 5 percent of the students during this period were officers. The Air Force and Army had the largest number of enlisted attendees enroll in basic level courses during this time period with 3,350 and 2,693 attendees, respectively. The Navy had 1,040 enlisted DLIFLC attendees from FY 2003 to FY 2006, and the Marines had 705. The Army had the highest number of officers attending DLIFLC in basic level courses during these years (133), while the Navy had 123, the Air Force had 119, and the Marines only had 46.

Average DLAB scores (enlisted and officer)

As noted earlier, a DLAB score of 100 or better is required by all services for attendance at DLIFLC (with the exception of those who are waived). Average DLAB scores for enlisted DLIFLC students over this period range from a high of 114 in the Navy to 111 in the Army (see figure 6). The Marines have the highest average score for officers at 123, while the Navy had the lowest average at 112. Average DLAB scores for officers are slightly higher than enlisted averages in the Army and Air Force, much higher in the Marines, and slightly lower in the Navy.

Figure 6. Average DLAB scores of DLIFLC students (enlisted and officer)



Languages studied

The majority of the languages taken in recent years have been the in-demand Category III and IV languages. Table 7 displays the number of students in each language as well as the category of each language. Category IV languages, such as Arabic, Korean, and Chinese-Mandarin, are the languages most frequently studied at DLIFLC

because of the demand for personnel with these skills in DOD. Over 61 percent of students were enrolled in Category IV languages, and 26 percent were taking Category III languages.

Table 7. Languages studied (FY 2003–FY 2006)

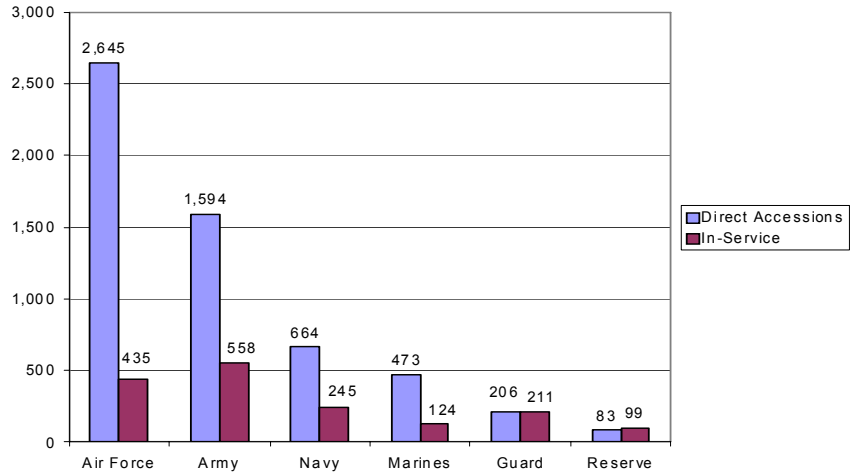
	Number of students	Language category
Arabic Modern	2,468	4
Korean	1,821	4
Chinese Mandarin	1,126	4
Persian Farsi	831	3
Spanish	655	1
Russian	536	3
French	270	1
Serbian Croatian	237	3
Pushtu Afghan	223	3
Persian Afghan Dari	160	3
Hebrew Modern	145	3
Tagalog	79	3
Japanese	66	4
German	60	2
Thai	54	3
Italian	52	1
Portuguese	34	1
Urdu	27	2
Uzbek	21	3
Turkish	20	3
Hindi	15	2
Kurmanj	14	3
Sorani	12	3
Kurdish	9	3
Vietnamese Hanoi	2	3

Direct accessions vs. in-service attendees

DLIFLC students are made up of not only new accessions but also personnel already in the service in another occupation. Direct accessions, however, make up the majority of the DLIFLC attendees as

compared with in-service students. Figure 7 represents the different numbers of direct accession students vs. in-service students by service. About 22 percent of enlisted students were in-service. Here we see that the Air Force had the highest number of direct accession enrollments with 2,645, followed by the Army with 1,594. These two services also have the highest number of in-service enrollments with 435 and 558, respectively.

Figure 7. Direct accessions vs. in-service DLIFLC students (FY 2003–FY 2006)



The Navy's direct accessions had the highest average DLAB score of 115.6, followed by the Marines, Air Force, and Army, respectively. As expected, direct accessions studying the Category IV languages had the highest average DLAB scores. DLIFLC attendees studying Arabic scored an average of 119.8 on the DLAB, followed by an average score of 116.4 for Chinese Mandarin and 114.9 for Korean (table 8).

The Marines had the highest average DLAB score for in-service attendees of DLIFLC at 112.9, followed by the Air Force, Guard, and Navy. Again, in-service attendees selected for the Category IV languages had the highest average DLAB scores.

Table 8. Average direct accession DLAB scores by language

Language	Avg DLAB	Category
Arabic Modern	119.8	4
Chinese Mandarin	116.4	4
Korean	114.9	4
Persian Afghan Dari	107.6	3
Pushtu Afghan	107.1	3
Persian Farsi	106.1	3
Russian	105.8	3
Hebrew Modern	104.2	3
Serbian Croatian	103.6	3
Tagalog	103.5	3
Spanish	102.9	1
French	97.4	1

Success at DLIFLC

To better understand the screening process, we explored student performance at DLIFLC. We focused our analysis on the performance of those enrolled in "basic" courses. The DLIFLC provided data for about 10,000 students enrolled in basic-level courses during FY 2004 through FY 2006. The performance measures available to us included course grades, summarized by grade point average (GPA), whether the student completed the course work (graduated), and the scores on the Defense Language Proficiency Test (DLPT). DLIFLC also provided background data on each student.

The DLPT is based on the Interagency Language Roundtable (ILR) scale, which separately assesses reading, listening, and speaking proficiency. The test results are expressed as levels ranging from 0 to 5 for each area evaluated, with half levels permitted, such as 0+, 1+, 2+.

For purposes of the analysis, we focused on two success criteria: graduation and proficiency test results. First, we considered those who did not complete their course work as failing to graduate. The second criterion for success was to achieve an acceptable score on the DLPT. The current course goals require "basic" students to achieve proficiency levels of 2 in listening/comprehension (L) and reading (R),

and 1+ in speaking (S). The tables in this section also show pass rates for higher levels (2/2/2 and 2+/2+/2), which were being considered at DLIFLC.

Student characteristics

The DLIFLC records include a number of student and class characteristics that could be useful for understanding outcomes at DLIFLC. Here we summarize selected student characteristics and performance measures. We categorize the characteristics examined as follows:

- *Personal characteristics.* Gender, prior education, language spoken in the home, etc.
- *Motivation*. As measured by whether students are assigned to their choice of language
- *Training enhancement*. Participation in the Proficiency Enhancement Program (PEP), lower student-to-instructor ratios, etc.
- *Recycles and relanguages.* As measured by whether a student had to start a language over again (recycled) or start over in a new language (relanguaged).

Table 9 summarizes the results of selected DLIFLC outcomes by officers, enlisted, and service status. Officers outperform enlisted personnel, although this may be because officers are more likely to take languages that fall into Categories I and II. Officers take Category I languages 45 percent of the time, compared with less than 10 percent for enlisted.

Those with more education are also more likely to succeed. Even though only a few non-graduates of high school (HS) attended DLIFLC, they were less likely to succeed. There is little difference in DLAB scores across the education continuum (table 10).

Table 11 shows the results for men versus women. The differences by gender are very small.

Non-native (English) speakers outperform native speakers. About 2 percent of DLIFLC students were not native English speakers. These students performed better at the higher levels of performance on the

Table 9. Officer and enlisted performance by service[a]

Branch	Prop. sample	Grad-uated	Met DLPT goal (2/2/1+)	(2/2/2)	(2+/2+/2)	DLAB	GPA
Enlisted							
Army	0.298	0.63	0.49	0.24	0.15	111	3.15
Air Force	0.371	0.63	0.48	0.26	0.15	113	3.15
Marines	0.078	0.69	0.55	0.29	0.16	114	3.06
Navy	0.115	0.69	0.57	0.29	0.17	114	3.22
Other	0.084	0.58	0.44	0.27	0.16	109	3.07
Officer							
Army	0.015	0.83	0.74	0.53	0.37	114	3.46
Air Force	0.013	0.91	0.82	0.68	0.46	114	3.54
Marines	0.005	0.83	0.67	0.50	0.24	122	3.39
Navy	0.014	0.83	0.69	0.57	0.39	112	3.37
Other	0.007	0.61	0.39	0.16	0.07	111	3.08

a. Note: Basic level courses.

Table 10. Performance by education[a]

Education	Prop. sample	Grad-uated	Met DLPT goal (2/2/1+)	(2/2/2)	(2+/2+/2)	DLAB	GPA
Non-HS grad	0.003	0.32	0.28	0.12	0.008	113	2.78
HS/GED	0.361	0.61	0.47	0.25	0.14	111	3.10
1 year college	0.156	0.62	0.49	0.26	0.14	112	3.14
2 years college	0.159	0.63	0.50	0.26	0.16	112	3.15
3 years college	0.065	0.64	0.46	0.22	0.13	112	3.11
4 years college	0.036	0.68	0.50	0.26	0.16	112	3.19
Bachelor's	0.139	0.72	0.57	0.33	0.21	114	3.25
Master's	0.079	0.73	0.60	0.39	0.25	115	3.33
Doctorate	0.002	0.74	0.53	0.37	0.21	115	3.34

a. Note: Basic level courses.

Table 11. Performance by gender[a]

Gender	Prop. sample	Graduated	Met DLPT goal (2/2/1+)	(2/2/2)	(2+/2+/2)	DLAB	GPA
Women	0.270	0.63	0.48	0.27	0.16	112	3.22
Men	0.730	0.65	0.51	0.28	0.17	112	3.14

a. Note: Basic level courses.

DLPT, and they had higher GPAs, despite lower DLAB scores. It is possible that DLAB may underpredict the ability of non-native speakers, which suggests a possible English language component (bias) in DLAB. The graduation rate was the same for native English and non-native English speakers (table 12).

Table 12. Effect of native language[a]

English is native language	Prop. sample	Graduated	Met DLPT goal (2/2/1+)	(2/2/2)	(2+/2+/2)	DLAB	GPA
No	0.021	0.64	0.53	0.35	0.21	108	3.25
Yes	0.979	0.64	0.51	0.28	0.17	112	3.16

a. Note: Basic level courses.

Upon entering DLIFLC, students are asked for their preference as to which language they would like to learn. DLIFLC administrators consider this a measure of motivation. As can be seen from the data shown in table 13, over 40 percent of students got their first-choice language. Those receiving their first choice did better, both in terms of graduation rates and DLPT pass rates.

Table 13. Effect of motivation[a]

Motivation	Prop. sample	Graduated	Met DLPT goal (2/2/1+)	(2/2/2)	(2+/2+/2)	DLAB	GPA
Not choice; don't want to study foreign language	0.022	0.53	0.34	0.22	0.14	109	3.01
Not choice; not motivated to study foreign language	0.012	0.50	0.29	0.11	0.07	112	2.93
Not choice; but motivated to study foreign language	0.309	0.62	0.46	0.23	0.12	111	3.10
Second or third choice	0.235	0.62	0.49	0.26	0.15	113	3.15
First choice	0.422	0.68	0.56	0.33	0.20	113	3.21

a. Note: Basic level courses.

DLIFLC introduced the PEP to improve student performance. PEP changes include reducing the student-to-instructor ratio, increasing the number of classrooms, creating improved expanded curricula, and expanding overseas training. In the Category III and IV languages, PEP decreases the student-faculty ratio from 10:2 to 6:2. In Categories I and II, PEP decreases the student-faculty ratio from 10:2 to 8:2. During the period we examined, only about 20 percent of students were in a class that used PEP.

The results in table 14 indicate that, in general, those students in PEP classes do better despite lower aptitude. The differences are particularly noticeable at the more stringent goal levels.

Table 14. Effect of PEP[a]

	Prop.		Met DLPT goal				
PEP	sample	Graduated	(2/2/1+)	(2/2/2)	(2+/2+/2)	DLAB	GPA
No	0.805	0.63	0.50	0.25	0.15	113	3.14
Yes	0.195	0.68	0.54	0.37	0.23	109	3.23

a. Note: Basic level courses.

Some students have difficulty in learning their language at DLIFLC. Many of these students are dismissed from DLIFLC, but others start the course over again (are recycled), or are switched into another language (relanguaged) (table 15). The services differ in their use of these procedures. The Air Force relanguages very few people; the Army, Marines, and Navy relanguage proportionately many more.

Table 15. Recycled vs. relanguaged by service

	Recycled (%)	Relanguaged (%)
Air Force	97	3
Army	77	23
Marines	75	25
Navy	64	36

Arabic, Korean, and Chinese Mandarin students are largely recycled instead of relanguaged, likely because of the difficulty of these languages for native English speakers, and the scarcity of students who score highest on the DLAB (table 16). The percentage recycled, in fact, decreases the most for Category I languages. Spanish is the only language in which more students are relanguaged than recycled.

Table 16. Recycled vs. relanguaged by language

	Recycled (%)	Relanguaged (%)
Arabic Modern	99	1
Korean	99	1
Chinese Mandarin	96	4
Hebrew Modern	82	18
Pashto Afghan	80	20
Russian	79	21
Persian Afghan Dari	75	25
Serbian Croatian	64	36
Persian Farsi	57	43
French	56	44
Spanish	34	66

When comparing performance of those taking the course for the first time with recycled and relanguaged enrollees (table 17), we find that recycled students are actually more likely to graduate.

Table 17. Performance and course entry status[a]

Course entry status	Prop. sample	Graduated	Met DLPT goal (2/2/1+)	(2/2/2)	(2+/2+/2)	DLAB	GPA
Initial	0.840	0.63	0.51	0.29	0.18	112	3.17
Recycled	0.126	0.72	0.51	0.21	0.08	111	3.14
Relanguaged	0.033	0.56	0.43	0.26	0.18	110	2.96

a. Note: Basic level courses.

Another way to examine language students is to compare those who enter training directly after recruit training (new accession) with those who have a year or more of military service. New accessions perform better at DLIFLC. Most students (84 percent) in our sample enter DLIFLC as new accessions (table 18).

Table 18. Performance and time in service[a]

New accession	Prop. sample	Graduated	Met DLPT goal (2/2/1+)	(2/2/2)	(2+/2+/2)	DLAB	GPA
Yes	0.838	0.65	0.51	0.28	0.17	112	3.17
No	0.162	0.62	0.45	0.25	0.14	112	3.07

a. Note: Basic level courses.

Finally, we looked at performance by language category (table 19). Most DLIFLC students during this time were registered in Category III and IV courses. Few were in Category II languages, such as German and Hindi. As expected, those taking Category I and II languages have higher success rates.

Table 19. Performance and language difficulty[a]

Language category	Prop. sample	Graduated	Met DLPT goal (2/2/1+)	(2/2/2)	(2+/2+/2)	DLAB	GPA
1	0.112	0.80	0.63	0.49	0.36	104	3.18
2	0.011	0.82	0.53	0.39	0.12	107	3.31
3	0.261	0.66	0.47	0.24	0.13	106	3.09
4	0.615	0.60	0.49	0.25	0.14	116	3.18

a. Note: Basic level courses.

Modeling enlisted screening practices

In the previous section, we presented an overview of the screening process from recruiting through completion of language training at DLIFLC. In this section, we develop models of two of the key points in the process for enlisted recruits: a model for predicting DLAB scores and a model to explain performance at DLIFLC.

The services and DOD are interested in identifying people who can do well at learning a language at DLIFLC. As part of an analysis of how AFQT and DLAB could be used to screen candidates for DLIFLC, Segall [4] estimated a model of the relationship of AFQT to DLAB score. As noted earlier, the services currently use a DLAB score of 100 to qualify for DLAB training.

Factors related to DLAB score

For estimating DLAB score, we developed a single equation based on data from all the services. We investigated the relationship between DLAB score and the following variables:

- AFQT
- Age
- Gender
- Education—some college.

We considered a number of different ASVAB-based predictors, including each of the service composites. However, after analyzing the different subtests and service composites, we selected AFQT because it produced the highest overall validity scores and is easily understandable across all services and within DOD.

We estimated the model using a linear regression model for the 4,512 accessions during FY 2008 for which we had DLAB scores. Table 20

shows the regression coefficients for the four variables. All were significant at the 1-percent level. AFQT was the dominant factor for predicting DLAB score. Each additional AFQT point beyond 50 increased the recruit's expected DLAB score by just over 1 point. This confirms the kind of relationship that Segall [4] had found previously. Gender was also very important; all else equal, women scored 5 points higher than men. Some college was also significantly positive; those with college credits scored about 1.6 points higher on the DLAB.

Table 20. Model coefficients for estimating DLAB scores for enlisted recruits

	Coef.	Std. Err.	t	P>|t|
AFQT	1.032	.02	48.70	0.000
Male	-5.24	.52	-10.12	0.000
Some college	1.59	.61	2.59	0.010
Age	-.369	.07	-4.92	0.000

Older recruits scored slightly lower, with recruits losing about 1 DLAB point if they were 3 years older than an otherwise comparable recruit.

Performance at DLIFLC

After linguists complete recruit training, they proceed to language training in Monterey, CA, at DLIFLC.

We analyzed personnel who entered DLIFLC during FY 2004 through FY 2007. We consider three separate indicators of success:

- Course grades
- Course graduation
- Performance on the DLPT.

We used both graduation and DLPT test score as our success criteria. We did not use course grade, which was only available for students who had completed training.

Completion of the course at DLIFLC is the minimum expected for a student to be considered successful. The DLPT score is also required to receive language incentive pay. So, we estimated two separate models for outcomes at DLIFLC. First we analyzed the factors associated with completing the language course. Then we estimated the effect of various student and class attributes on whether the student achieved the desired DLPT score at the completion of the course.

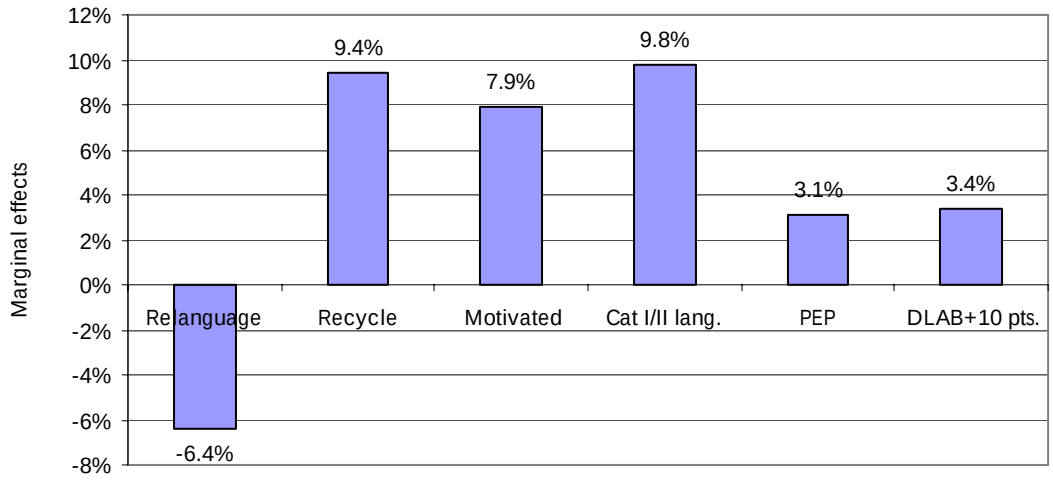
Analysis of course completion rates

We estimated the probability that a student would complete a basic language course as a logistic regression, where:

- P(C) = 1 if the student completed the course, 0 otherwise
- DLAB = the student's DLAB score
- EDLVL = 1 if the student had completed a Bachelor's degree, 0 otherwise
- MOTIVE = 1 if the student received their first choice of language, 0 otherwise
- EASY-LANG = 1 if the student was enrolled in a Category I or II language, 0 otherwise
- RELANG = 1 if the student had been reassigned to a second language, 0 otherwise
- RECYCLE = 1 if the student had been recycled (restarted) within the same language, 0 otherwise
- PEP = 1 if the student was enrolled in a PEP class, 0 otherwise.

Just over 78 percent of the students completed their course at DLIFLC. Table 24 in appendix A provides the results of the logistic regression on course completion probabilities. All factors except EDLVL were statistically significant at the 95-percent level or higher. Figure 8 shows the effects of changing each of the factors on the expected probability of completing the course. For example, all else equal, a recruit who was relanguaged had a 6-percentage point lower probability of completing the course than one who was not relanguaged.

Figure 8. Impact of selected factors on course completion (78 percent overall average)



Analysis of DLPT success rates

The DLPT consists of three parts: listening/comprehension, reading, and speaking. During the period covered by our data, scores of 2/2/1+ were considered passing. So, for our analysis, we considered a student to be successful if he or she got those scores or better and completed the course.

We evaluated a set of potential predictors of success against each of these two measures of success at DLIFLC: completing and passing the DLPT, and passing the DLPT contingent on completing the course. We evaluated the following explanatory factors:

- DLAB score
- New accession (compared with someone with more than 1 year of service)
- Enlisted (compared with officer)
- Service (compared with Army)
- Other service (Coast Guard)
- Male (compared with female)
- Language Category I or II

- PEP—participation in the PEP program
- DLPT 5 (compared with DLPT version IV)
- Category I or II language (compared with Category III or IV)
- RECYCLE—a student who is recycled
- RELANG—a student who is assigned to a new language
- MOTIVE—a student who received his or her first choice of language
- NATIVE—English speaker
- Education—three different categories were evaluated: whether the student had some college (EDLVL2), a Bachelor's degree (EDLVL3), or more than a Bachelor's degree (EDLVL4).

The dependent variable, passing the DLPT, is a dichotomous variable; the student is either successful or not. We used a logistic regression model to estimate the coefficients against a population of 7,000 students at DLIFLC who began their studies during 2003 through 2006. The complete equation can be found in table 25 in appendix A. Our estimation found that being a new accession, male, being a native English speaker, and being enlisted did not significantly affect a person's likelihood of successfully completing the course. Also, Air Force and Marine students did not differ significantly from Army students, and education levels other than a Bachelor's degree were not significant. Since our model was empirically based rather than theoretical, we dropped those variables that did not add any explanatory power from the model and reestimated the equation. Table 26 in appendix A shows the results from this equation.

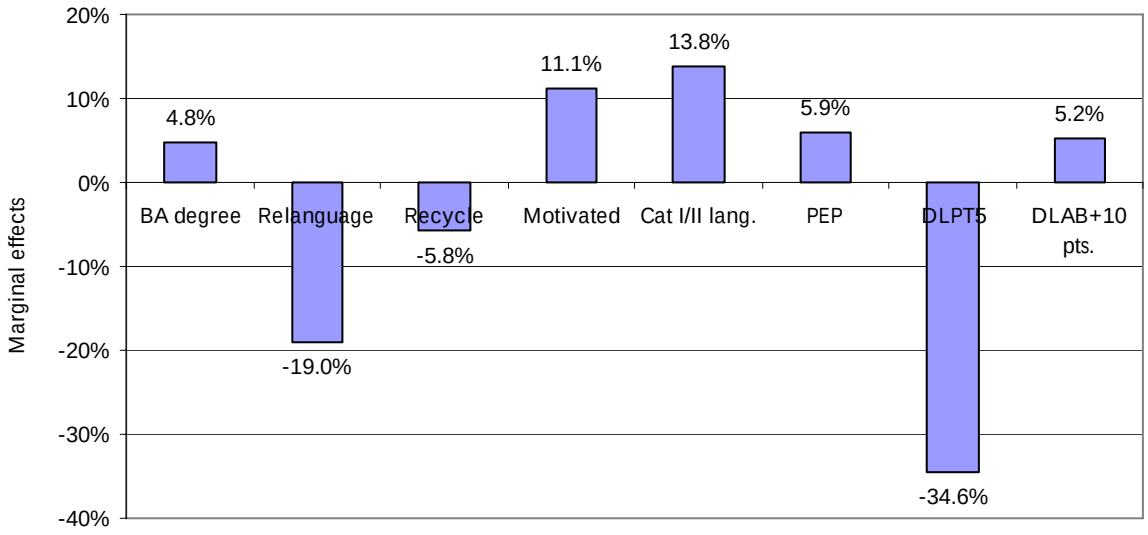
A number of student characteristics were significant for predicting successful goal achievement. DLAB score, as expected, was significant and positive. Students with Bachelor's degrees (including both officers and enlisted personnel) were also more likely to be successful at DLIFLC. Motivation was also an important factor in achieving success at DLIFLC.

Students from the Navy performed significantly better than those from other services during the period we examined. It could be that

the Navy was sending relatively small cohorts of students during this period and was somehow more selective in finding candidates, or it could be that the Navy was somehow matching students with languages more successfully than other services. Since our model controlled for DLAB score as well as many of the other factors that could be hypothesized to be related to performance, this finding may warrant additional investigation, particularly if it is sustained.

The overall success rate on the DLPT during the period we examined was 64 percent. Figure 9 shows the effects of changing selected factors on DLPT success rates. For example, those who had received a BA degree were nearly 5 percentage points more likely to achieve scores of 2/2/1+ or better on the DLPT than those without a degree.

Figure 9. Impact of selected factors on DLPT success (64 percent overall average)



Students who were relanguaged or recycled were less successful than those who were not. Relanguaging students had a much more negative outcome than recycling students, however.

Three characteristics of the learning environment were significant in their impact on student success. Students taking Category I and II languages were much more likely to achieve their learning goal, even not controlling for DLAB score. A student in one of these languages

was more likely to be successful than a student in a Category III or IV language, even if his or her DLAB score was 20 points lower.

The PEP program was significantly positive in its impact on students achieving their goals; in fact, it was as significant as having a Bachelor's degree. It increased the success rate more than selecting students with DLAB scores that were 10 points higher.

The single factor that had the strongest relationship to whether a student passed the DLPT was whether the student had taken DLPT 5. DLPT 5 is more difficult than the earlier version and includes more text synthesis and interpretation items. The DLPT 5 uses a new score scale and is therefore delinked from the previous version. Students evaluated against the DLPT 5 test were much less likely to pass the standard. Appendix B contains a separate analysis that we performed of the DLPT IV and DLPT 5 results for three languages, comparing results on specific parts of the DLPT. We found that it was important to compare results for reading and listening separately and to differentiate results by language. In any analysis of DLIFLC outcomes, it is important to separate the results by the different DLPTs used.

Quantifying the impact of selected factors on DLIFLC success

The DLIFLC performance models can be used to estimate how changing particular factors will affect success rates. We can do this by taking the logistic regression models and estimating how changing a particular factor will affect the probability of successfully completing a course and achieving 2/2/1+ on the DLPT test.

For our analysis, we investigate changing the following factors:

- Increasing average DLAB scores by 10 points
- Increasing the number of students receiving their first choice of language
- Increasing the number of students enrolled in PEP
- Recycling a student
- Relanguaging a student.

We found that in our sample of over 9,000 students, only slightly more than half both completed the course and passed the DLPT IV examination at the 2/2/1+ level.

One way to increase success rates at DLIFLC is to increase the DLAB scores of students. If the average student's DLAB score was increased by 10 points, from 113 to 123, we estimate that the success rate would increase by 6.4 percentage points to 56.8 percent.

We found that motivation was a significant factor in DLIFLC student performance. Replacing a student with one who received his or her language of choice would increase the success probability by 14.6 percentage points.

While the PEP program was relatively new during the period we studied, it was already showing significant improvements in success. A student placed in PEP would have a 6.8-percentage-point increase in his or her success rate at DLIFLC, increasing the overall success rate to 57 percent.

Recycling and relanguaging students resulted in markedly different impacts on success rates. Recycles show a slight increase in overall success rates, while relanguaged students were 18 percentage points less likely to both complete the course and achieve the DLPT standard.

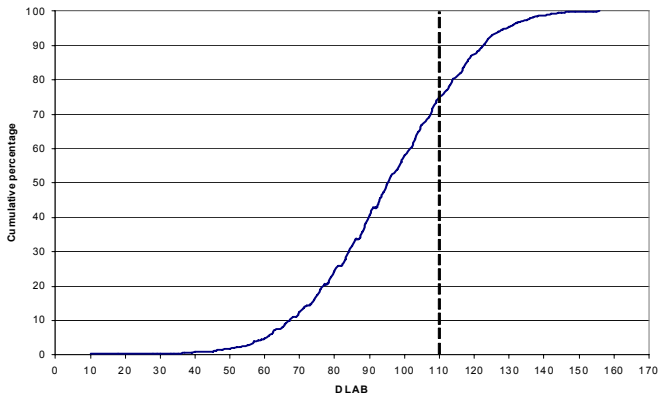
Modeling officer screening practices

In addition to new recruits, officers are also screened for language aptitude to attend DLIFLC and other language education programs. We obtained data from the Air Force ROTC program to gain insights about linguist selection and language training of officers. The Air Force ROTC program provided data on DLAB scores, academics, and demographics. The data included 6,200 Air Force ROTC students with DLAB scores with projected graduation dates of FY 2008 through FY 2013.

Of the 6,200, 4,781 had information on either SAT or ACT scores. We converted SAT scores to equivalent ACT scores. We then estimated the probability of scoring greater than or equal to 110 on DLAB given ACT / SAT, cumulative grade point average, gender, and college major (technical vs. non-technical), as categorized by the Air Force. Technical majors were those studying mathematics, engineering, or physical sciences.

Figure 10 displays the actual DLAB score distribution of the students.

Figure 10. Air Force ROTC DLAB score distribution



As the graph shows, only about 25 percent of the Air Force ROTC students score 110 or better on the DLAB, illustrating the difficulty of the test.

Next, we used logistic regression to predict which students would score 110 or greater on the DLAB given their characteristics (table 21). We found that ACT/SAT were significant positive predictors of DLAB score. All else equal, male students had significantly lower odds of scoring 110 or more. GPA was positively related to scoring 110 or above. Being a technical major had no effect on meeting the DLAB standard.

Table 21. Logistic regression results—predicting DLAB >=110

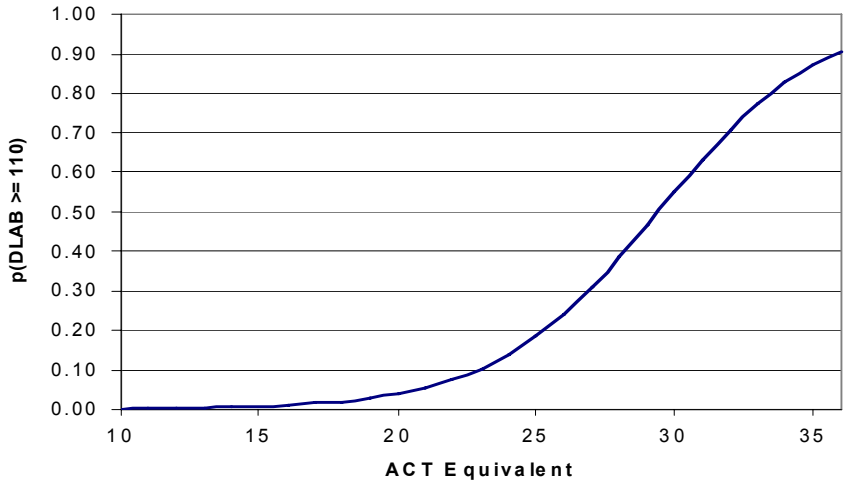
Var.	Coef.	z	P >
cum_gpa	0.370	4.260	0.000
male	-0.652	-6.860	0.000
ACT	0.347	24.290	0.000
tech major	0.080	0.980	0.329
_cons	-10.862	-24.890	0.000

R-sq = 0.20

Looking more closely at the relationship between DLAB and the single strongest predictor, ACT/SAT scores (figure 11), we see that the probability of scoring well on the DLAB increases as ACT/SAT scores increase. Students scoring 30 or above on the ACT would be predicted to have a 50-percent or greater chance of scoring above 110 on the DLAB. This score would be roughly equivalent to 1,400 on the SAT.

This analysis of Air Force ROTC students can be used to help select officers for language training. Since we know the characteristics of students that are likely to do well on the DLAB, it may not be necessary to test all officers with the DLAB. If it is desirable to test only officers with a probability of at least a 50-percent chance of getting a DLAB of at least 110, only those with an ACT of 30+ (or equivalent SAT) should be tested.

Figure 11. Probability of DLAB >=110 and ACT/SAT



We looked at officer screening in the Marine Corps and discuss these results in appendix C. We also found that, all else equal, SAT score correlated positively with DLAB scores, and that women scored significantly higher than men.

This page intentionally left blank.

Developing best practices for screening for language aptitude

We have described the current process for developing linguists in DOD, following the steps from recruiting through completion of language training at DLIFLC. We developed models to predict performance on the DLAB test, as well as performance at DLIFLC.

In this section, we apply our models to the screening process to identify what sort of tradeoffs are possible. We also test models that can be used to evaluate changes that could be made to the current screening process.

Current service screening procedures

Each service has different procedures for screening recruits to take the DLAB. We found in our discussions with each service that the typical linguist is identified during the recruiting process based on screening against the service's composite, the candidate's citizenship, his or her interest in the position, and the needs of the service.

Figure 1 showed a model of the screening process. The potential population of military linguist candidates is defined by the enlisted recruited population. The initial screening decision made by each service is whom to test with the DLAB. Hence, the first empirical model we estimate is an analysis of the factors that relate to whom to test for each service.

DOD wants to identify those people who will do well at DLIFLC, so the second model we develop predicts DLAB performance.

In addition to identifying candidates for language screening, several other steps need to occur in producing a linguist. After taking and passing the DLAB, will the person actually accept a linguist job, or will

he or she enter another training program? Also, the linguist will need to complete boot camp in order to progress to language training.

We analyzed the following decision points to model the factors involved in producing linguists:

- Testing probability
- DLAB passing probability
- Linguist placement probability
- DLIFLC success probability.

Factors related to the probability of taking the DLAB test

First, we investigated the factors that are related to each service's current testing program. We estimated separate probability models for each service. These models included the following factors:

- AFQT
- Gender—male
- Education—some college
- Age
- Marital status
- Paygrade
- Distance to MEPS
- Per diem rates
- AD—the time between DLAB testing and accession.

We used AFQT as our aptitude predictor in this analysis since we had found it to be overall more highly correlated with DLAB scores than the service-specific composites. We further limited our sample to recruits with AFQT scores of 50 or above since virtually all recruits (over 99.5 percent) currently taking the DLAB score 50 or above on the AFQT. Education was defined as having at least some college experience because we found previously that some services test more

people with additional education. Age, marital status, gender, and paygrade were also used as controlling factors because we believe that they could be used to screen recruits.

We examined three operational factors developed from the MEPCOM data that we thought could be related to testing rates: (1) distance from the recruit's home to the MEPS, (2) per diem rates for the MEP location, and (3) time between entering the DEP and accession (AD). The first two factors are associated with the potential costs of testing a recruit. We wanted to see if the services were restricting testing on the basis of potential costs.

The other factor, AD, is a measure of how much time is available to screen a recruit before he or she leaves for boot camp. It may be that recruits with longer times in DEP are more likely to be tested since they would have time to be brought back to the MEPS and tested with the DLAB.

We estimated multivariate regression models in which the dependent variable was whether a recruit would be given the DLAB test and the explanatory factors were the ones discussed above. We used a logistic formulation to estimate the probability of taking the DLAB. Appendix D contains the results of the regression models. Table 22 summarizes the results for each of the four services. Factors that are positively related to testing rates are marked with a plus (+) sign, and negative relationships are indicated by a minus (-) sign. Two asterisks indicate that the factor was significantly different from zero at the 1-percent level; one asterisk indicates significance at the 5-percent level.

While all services use ASVAB (either AFQT or a composite correlated with AFQT) to identify candidates for the DLAB test, the significance of the other factors differs considerably across the services. For example, men were significantly less likely than women to be tested in all the services except the Army. Having at least some college credit increased the likelihood of testing in all the services, but significantly for only the Air Force and the Navy. Single recruits were more likely to take the DLAB, but the increase was significant for only the Army and the Navy.

Table 22. Factors related to testing rates

	Army	Air Force	Navy	Marines
Number	39,683	21,657	28,909	23,704
Variable				
AFQT	+ **	+**	+**	+**
Male	+	-**	-**	-**
Some college	+	+**	+**	+
Single	+ **	+	+**	+
Age	-	+*	+	+
E1	-**	+	+**	-
Distance to MEPS	-	-*	-	-**
Per diem	-	+**	-	+
Accession gap (AD)	-	+**	-*	+

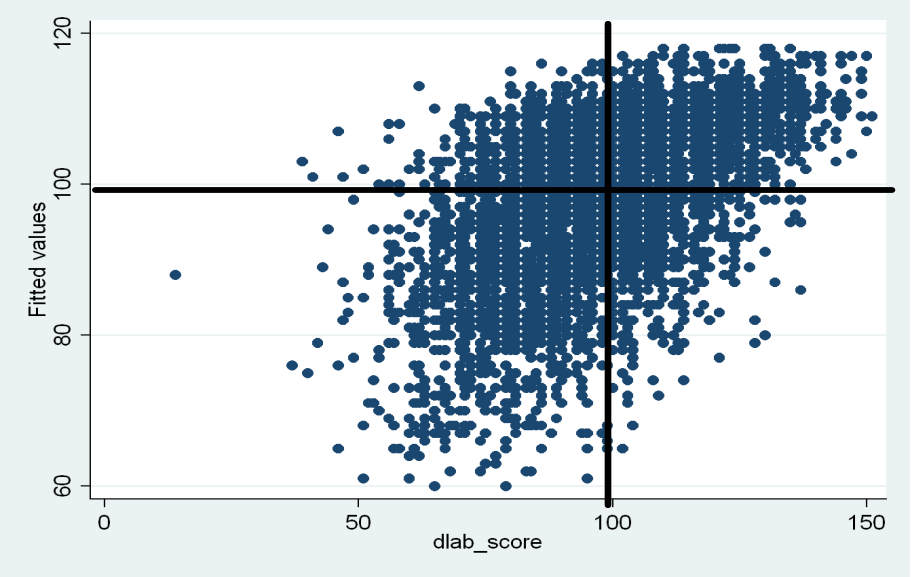
The factors relating to recruiting costs—distance to MEPS and per diem costs—were only marginally useful. Distance was negatively related to testing in all the services, which was the expected direction of effect, but only significantly related for the Air Force and the Marine Corps. Per diem costs were significant only for the Air Force, and there the sign was the opposite of what we expected. The AD variable produced different results across the services. It proved to be significant for both the Air Force and the Navy, but with opposite signs.

Current screening practices

In an earlier section (see table 20), we developed a model that can be used to predict DLAB scores. Our model used AFQT, education, age, and gender as factors associated with explaining differences in DLAB scores. We then applied our model to the FY 2008 recruit population that had already taken the DLAB to determine how well our model's results compared with the actual testing results.

The DLAB was given to 4,925 recruits in FY 2008. For each recruit we produced two numbers: the recruit's actual DLAB score, and his or her predicted score, based on our screening model. Figure 12 shows the distribution of test scores for all services, with the actual score plotted on the X axis, and the predicted score plotted on the Y axis.

Figure 12. Distribution of actual DLAB scores and fitted DLAB scores for 2008 accessions



We have marked the score of 100 on each axis. Those scoring 100 or above on the X axis would be those recruits who would qualify for language training at DLIFLC; those scoring below 100 would not qualify (although, as we discussed earlier, some services provide waivers to some of these individuals).

The percentage of those tested who fall to the right of this line defines the selection ratio. Figure 13 provides the overall selection ratio, as well as the ratio for each service. For all services, we found that 47 percent of those tested passed the DLAB standard. By service, the success rate ranged from 39 percent for the Army up to 52 percent for the Air Force and the Navy. This means that more than two people must be tested using the current screening policy for every one that is selected.

Figure 12 also illustrates how our screening model could be used to identify candidates for DLAB testing. Those we predict to score 100 or above are identified by the scores above 100 on the Y axis. These are the recruits that our model suggests should be tested because they would be predicted to score 100 or above on the DLAB test. Those in

the upper right quadrant are the recruits who scored 100 or above and were predicted to score 100 or more as well. The proportion of those in the upper right quadrant compared with all those in the two quadrants above 100 on the Y axis is the percentage of successful predictions we would make using our model.

Figure 13. FY 2008 selection ratio

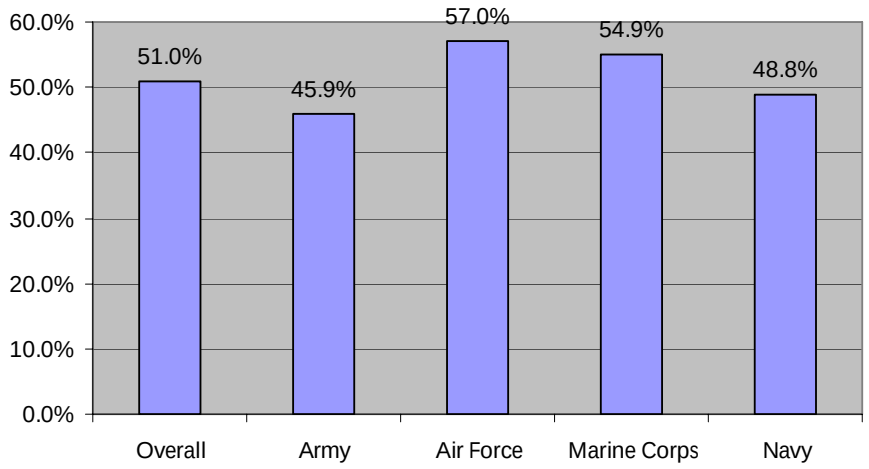


Figure 14 shows the successful predictions our model would have made for each service, as well as overall. Screening based on our predicted DLAB score proved to be correct in 68 percent of the population, ranging from 63 percent for the Air Force and up to 76 percent for the Navy.

If the screening model could be applied to identify DLAB test takers, it is theoretically possible to reduce the number of test takers. The screening model would only test slightly more than half of the current DLAB test population. Figure 15 shows that only 46 percent of the Army test takers would have been selected for testing under the screening model. The highest percentage was for the Air Force; 57 percent of its test population would have been tested.

Would it be feasible for the services to find sufficient candidates qualified under our screening model to take the DLAB? We took the FY 2008 accession population that scored 50 or above on the AFQT

and applied our DLAB estimation model to project the number of recruits who would qualify against a screen of 100 or above. Figure 16 shows the percentage of each service that would be predicted to score 100 or more. Overall, 16 percent of the high-AFQT population would be expected to score 100 or above, ranging from 15 percent for the Army up to 19 percent for the Air Force.

Figure 14. Percentage of successful predictions, by service

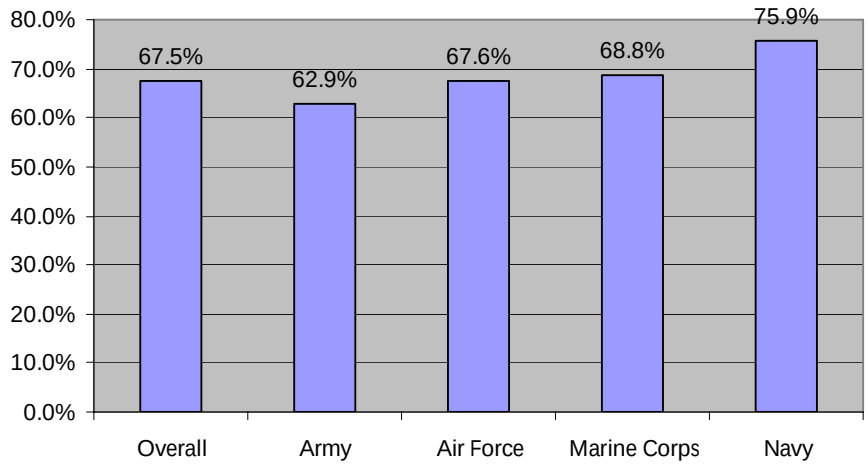


Figure 15. Percentage of DLAB test takers predicted to score 100 or above

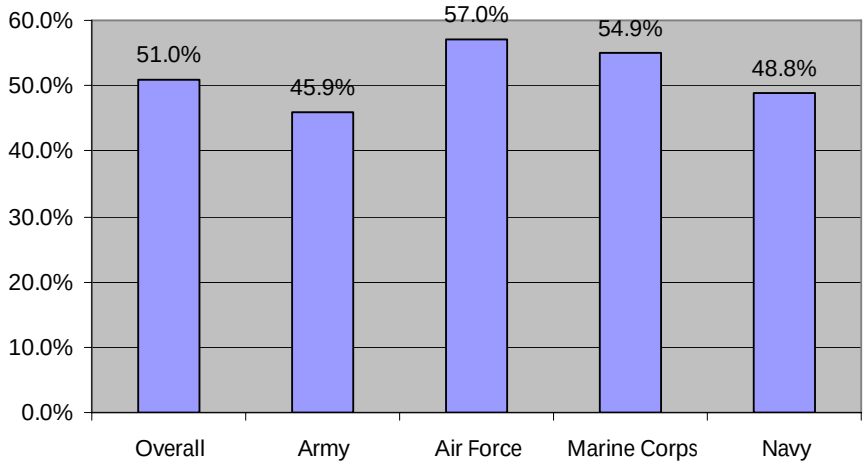
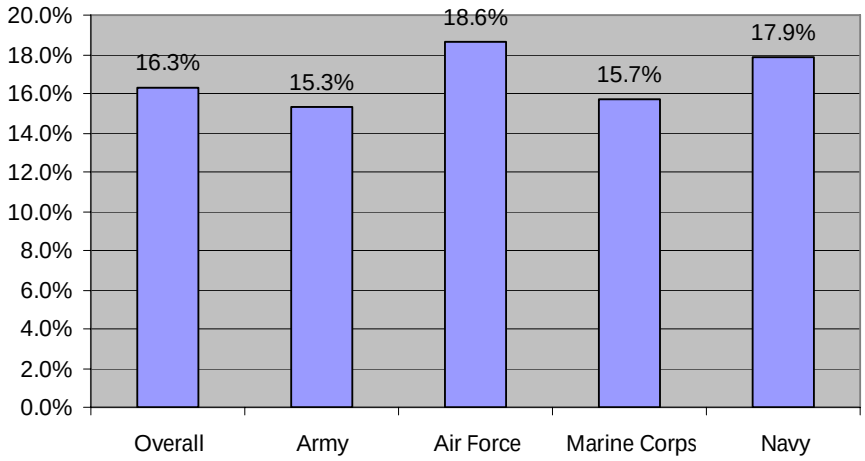


Figure 16. Percentage of recruits predicted to score 100 or above



One measure of the feasibility of applying this screen is to compare the selection ratio for testing needed to achieve recruiting goals. We calculated the number of recruits predicted to score 100 or more and compared it with the actual number recruited by each service and overall. The testing ratio was above 10 overall, ranging from about 5 for the Air Force and up to 14 for the Marines. Thus, overall there is at least a potential supply of DLAB testing candidates that exceeds the current requirement.

Modeling selection strategies

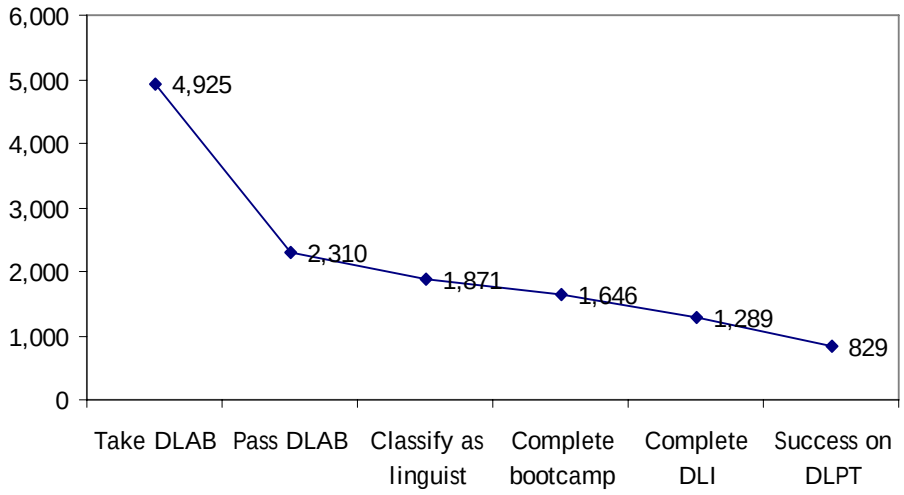
Each service performs the following steps to produce a graduate of DLIFLC who passes the DLPT:

1. Select recruits for testing.
2. Find recruits passing the DLAB score.
3. Place recruits into linguist position.
4. Ensure that recruits complete boot camp and start at DLIFLC.
5. Ensure that students complete DLIFLC curriculum and pass DLPT.

The services currently accomplish each of these steps in order to end up with a linguist. Figure 17 shows an example of the number of

recruits who would progress to the next stage out of FY 2008 candidates who were tested on the DLAB.

Figure 17. Actual FY 2008 test takes and projections for completing recruit training, language training, and passing DLPT



In FY 2008, less than half of the recruits tested scored 100 or above on the DLAB. Thus, out of 4,925 screened with the DLAB, 2,310 have qualified for linguist training. We observed that for the Army, Navy, and Air Force about 81 percent were assigned to linguist training, resulting in 1,871 entering military service on track to become a linguist. About 12 percent would be expected to attrite during boot camp, with 1,646 progressing to DLIFLC. In the period we examined, 78 percent completed language training and 64 percent passed the DLPT IV, resulting in 829 successful graduates, or 17 percent out of the original screened population.

Comparing different screening strategies

The current screening process uses service-specific screens based on ASVAB and other information on recruits, such as interest, motivation, and availability. This first screen is combined with the DLAB screen to produce the results shown above, with roughly six individuals tested for every one successful student at DLIFLC.

We identified several strategies that could be used for screening:

1. Using the current two-step process of selecting individuals for testing, followed by the use of the DLAB screen with a cut score of 100
2. Using a two-step screening method based on predicted DLAB score of 100, as well as passing a cut score of 100
3. Using a single screen based on a predicted DLAB score of 100
4. Using the current two-step screening method with a higher DLAB cut score (DLAB >=110)
5. Using a two-step screening method based on predicted DLAB score and a higher standard (DLAB >=110)
6. Using a single screen based on predicted DLAB with a higher score (DLAB >=110).

Referring back to figure 12, the current process selects all those scoring to the right of the score of 100 on the X axis. These would be the recruits who would pass the DLAB cut score of 100.

Figure 12 also illustrates the application of the two-step selection process with predicted DLAB scores being used to identify candidates for testing with the DLAB on the second step. In this scenario, only those scoring above 100 on the Y axis would be tested, with the upper right quadrant being those passing.

The third scenario shows the results of a one-step process using predicted DLAB score. Here all recruits falling above 100 on the Y axis would pass the screen.

The fourth scenario would use a cut score of 110 and apply it to the above three examples: current screening policies, screening based on a predicted DLAB score of 110 or above, and no screening other than a predicted DLAB score of 110 or more.

To evaluate these alternatives, we apply the following steps to each screening policy:

1. We use the overall 81-percent factor to identify the percent of those who qualify under the screening who end up being assigned to linguist training.
2. We factor in the probability that they would complete recruit training and begin training at DLIFLC at 88 percent, the recent historical average. We calculate a success probability for each policy based on predicted DLAB score.

Based on this procedure, we compute the number of FY 2008 recruits who would be expected to progress through the testing, assignment, and training pipelines. To estimate the success rate from training, we apply the logistic regression equations developed in the previous section to the people starting DLIFLC. We use either DLAB score or predicted DLAB score as the factor we vary according to the scenario.

This approach makes a number of assumptions about how to adjust for changes in screening and performance. We assume that the percentage of people who pass, but who are not assigned as linguists, is the same under all scenarios. For those who end up not entering the service as a linguist, we implicitly assume that the outcome is due largely to effects that are independent of testing and that are unlikely to be affected by screening. People may take the DLAB, and do well, but either they are more interested in learning other skills or they have other circumstances change concerning their enlistment that preclude linguist training.

The proportion of people who progress fromrecruit training to DLIFLC is assumed to be invariant as well. Since we do not have specific data on boot camp completion rates for linguists, this could be inaccurate. However, we have no reason to believe that linguists differ in their completion of boot camp in any significant fashion.

More troublesome is the use of DLAB and projected DLAB scores to estimate success at DLIFLC. Unlike [4], we do not have any information on AFQT scores for DLIFLC students. Because all students are screened against DLAB, it is possible that the effects of using an ASVAB-based screen could be biased. At present, no candidate who scores high on the ASVAB but low on the DLAB is sent to DLIFLC. It is possible, however, that such candidates could do better than we

currently project. At a minimum, it would be useful to determine how AFQT and DLAB scores separately affect performance at DLIFLC.

Figure 18 shows the screening pass rates for each of the first three policies—based on a minimum standard equivalent to a DLAB score of 100. Selection rates range from 46.9 percent for the current system up to 100 percent where the only screening is done based on predicted DLAB score.

Figure 18. Screening pass rates for selected policies (DLAB = 100 standard)

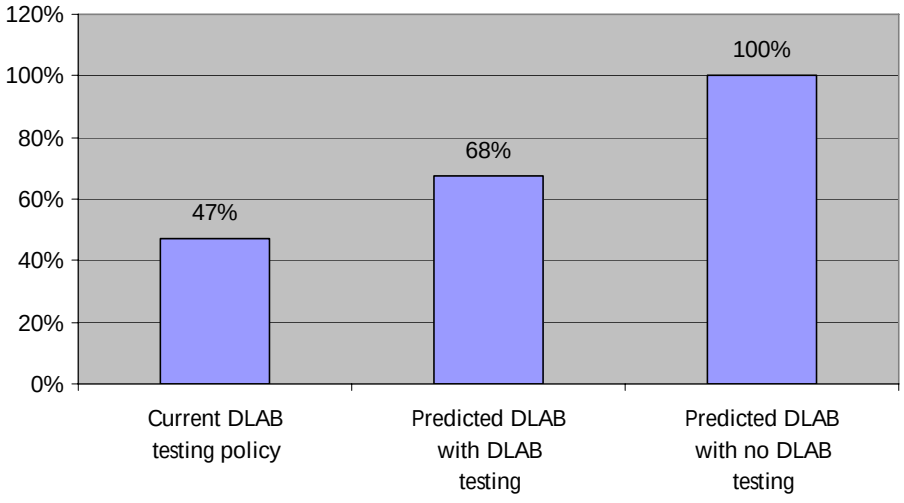


Figure 19 shows the percentage of those who would be expected to complete the DLIFLC curriculum and pass the DLPT IV standard. The combination of predicted DLAB score above 100 with an actual score above 100 would yield a success rate of 52.8 percent. The current system produces a pass rate of 51.6 percent. Using solely predicted DLAB scores would result in only 47.1 percent passing.

Figure 20 shows the screening success rate when a cut score of 110 is used. In this case, only about 28 percent of those tested would pass the screen under current testing practices. The selection percentage using the predicted DLAB screen would remain high at about 70 percent. Of course, without any DLAB screen, all of those predicted to score 110 or more would be accepted.

Figure 19. Percentage of successful students (DLAB =100 standard)

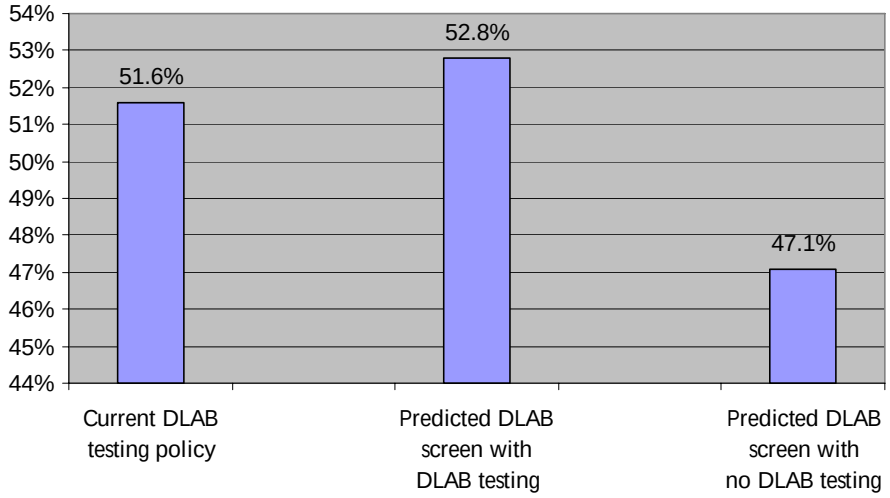


Figure 20. Screening pass rates for selected policies (DLAB = 110 standard)

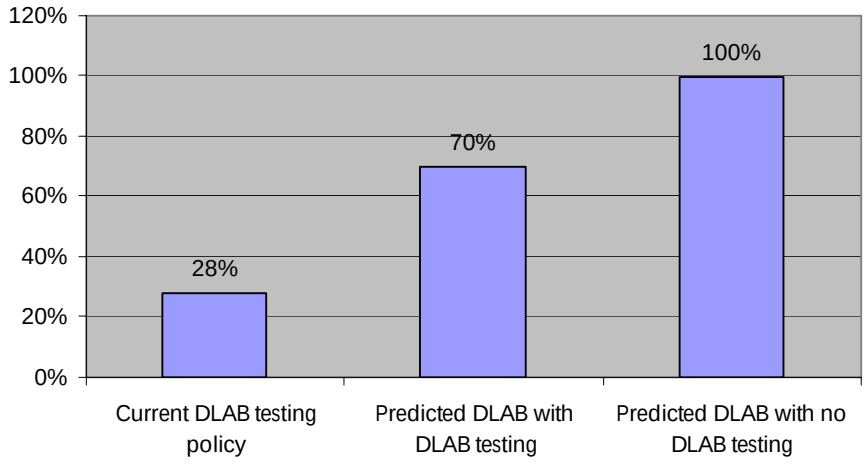
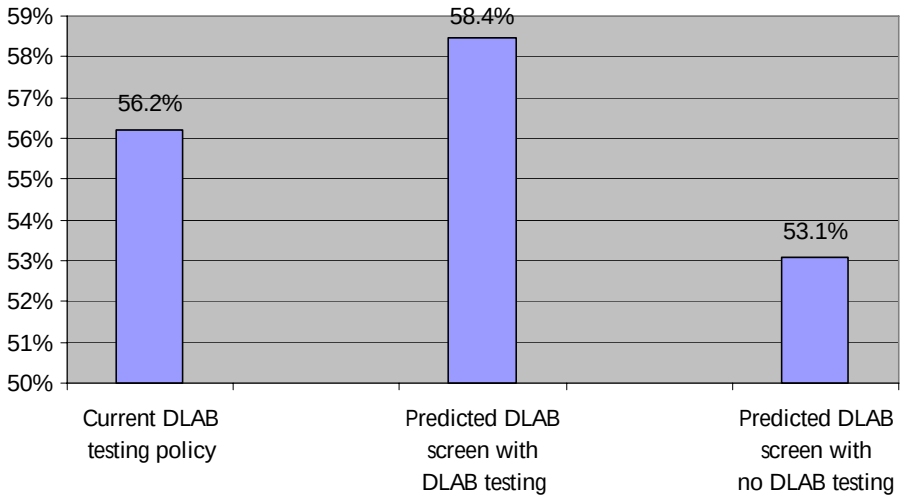


Figure 21 shows the percentage of students at DLIFLC who would be expected to pass the course and DLPT IV standard if a minimum qualifying score of 110 was used. Again, the combination of predicted DLAB score with actual DLAB score would produce the highest success rate—about 58 percent. Current screening would result in about

56 percent being successful, and no DLAB screening would lead to about 53 percent completing DLIFLC successfully.

Figure 21. Percentage of successful students (DLAB =110 standard)



To judge the efficiency of the different standards, we estimated the numbers required to be screened at the MEPS and starting DLIFLC to produce the same number of graduates. These are the two key steps in the screening process that can be connected to resources. Screening candidates using the DLAB usually requires a separate trip to the MEPS, incurring, at a minimum, transportation costs. To the degree that recruiter effort is required to identify candidates for screening, actual costs may be substantially higher. Students sent forward to DLIFLC will incur training resources. Furthermore, any policy that could increase the number of students required to achieve graduation requirements will bump up against capacity constraints at the school.

We calculated that the linguists who were recruited in FY 2008 would result in 850 graduates at DLIFLC who would also pass the proficiency test with scores of 2/2/1+. Thus, we evaluated each scenario against the number of people who would need to be screened to produce the number of students starting at DLIFLC that would be needed to achieve this number of successful completions.

Table 23 shows the numbers screened and arriving at DLIFLC that would be required to produce 850 graduates under each scenario. The smallest number to be screened would be under the scenarios in which ASVAB only screening are is for selection. If the DLAB standard of 110 was used, only 2,246 individuals would be selected for classification. However, these policies would increase the number of students at DLIFLC.

Table 23. Number of recruits screened and beginning training at DLIFLC under different screening regimes

Scenario	Number screened	Change from current policy	DLI students	Change from current policy	DLAB score	Average AFQT score
100 DLAB standard						
DOD current screen	4,925	--	1,646	--	114.6	90.1
ASVAB +DLAB screen	3,344	-32%	1,609	-2%	116.5	94.0
ASVAB only screen	2,531	-49%	1,804	10%	107.3	93.2
110 DLAB standard						
DOD current screen	7,604	54%	1,512	-8%	121.7	92.1
ASVAB +DLAB screen	2,916	-41%	1,455	-12%	125.1	97.2
ASVAB only screen	2,246	-54%	1,601	-3%	116.9	97.0

Conversely, the highest numbers requiring screening will occur when the current two-stage process is used at the 110 standard. Nearly 8,000 would need to be tested under the restrictive 110 minimum DLAB score standard. At present, we estimate that there are over 25,000 potential candidates for screening, so this policy would require finding and testing 30 percent of the eligible candidates.

Three of the alternatives produced the required number of graduates with both smaller numbers of test takers and fewer students at DLIFLC than the current policy:

- Using a predicted DLAB score of 100 as a pre-screen with DLAB (strategy 2)
- Using a predicted DLAB score of 110 as a pre-screen with DLAB (strategy 5)
- Using a predicted DLAB score of 110 as the only screen (strategy 6).

Use of the predicted DLAB score could potentially reduce the number of recruits that the services would need to send to the MEPS for DLAB testing. While it is unlikely that the 32-percent reduction in screening over current policies could be achieved operationally, the pre-screening model could substantially reduce the number of marginal candidates sent forward to the MEPS.

The strategy of eliminating DLAB testing and simply using a pre-screen of a predicted score of 110 or above would also produce the required number of graduates with no screening costs, yet slightly higher success rates at DLIFLC. This example illustrates the tradeoffs between screening and other aspects of the process. Only 13 percent of current recruits who are tested would be waived; these language candidates would constitute over a quarter of the FY 2008 quota of DLIFLC candidates. Table 23 showed that the average AFQT score of recruits meeting this screen would be 97. A high percentage of very high aptitude recruits would need to be directed into language training, yet performance would not change appreciably over the current system.

Conclusions and further discussion

We have reviewed how the services currently screen recruits to identify language aptitude. While ASVAB, notably AFQT, plays a major role in the screening process, there are differences across the services in how screening costs as well as recruit attributes are related to who is tested with the DLAB.

We have developed models that can be used to predict DLAB scores from enlisted accession characteristics, as well as a model of factors that affect success at DLIFLC. We have also developed a model that can be used to screen officers based on information from their college records, such as grades and standardized test scores.

These models enable DOD and the services to assess the probability of any military population—officer and enlisted—to qualify for language training using a few key individual factors. We have given some examples of how these models can be applied to the current screening process, including a range of the possible benefits that could be produced through more efficient screening.

We have identified four areas that are candidates for improvement, either by the services or DOD:

- Quantifying the benefits and costs of alternative screening approaches
- Developing additional predictors
- Analyzing programs, resources, and outcomes from training
- Developing databases that can be used to support both operational needs and analysis.

Quantifying the benefits and costs of alternative screening approaches

While there are many steps involved in the screening process, two steps generate costs and can produce different benefits, depending on the approach: the cost of screening a candidate with DLAB and the cost of training a student at DLIFLC.

While the costs of training are undoubtedly much higher than the costs of testing, there is considerable variability in the numbers of recruits who would be tested under different approaches. We obtained an estimate of $600 for transporting a recruit to the MEPS for testing, but this figure did not include recruiter effort in identifying candidates. Also, it is assumed that screening costs are only the costs of placing people who have already been recruited into linguist positions. Since the typical candidate is a very high quality recruit, the services have many competing positions into which the person could be placed. In the Navy, the nuclear power field would be one competitive option. In the Army, Special Forces would be a competitive field, especially since foreign language training would be part of the skill training for this MOS as well. Part of the process of finding candidates for skilled linguists involves understanding the costs of classifying potential candidates as linguists versus other positions demanding similar skills.

Developing additional predictors

ASVAB and DLAB have been well developed as predictors of language aptitude [4, 5]. We have expanded the list somewhat by including education, gender, and age as additional predictive factors. However, motivation showed up in the analysis of DLIFLC performance as a significant factor affecting success.

Motivation could be used two ways in the screening process: in recruiting and in training. In recruiting, motivation could be used to identify those candidates who exhibit either persistence or interest in the career field. This could reduce recruiter effort in identifying candidates for screening and improve the likelihood of completing training, thus reducing future requirements.

The motivation test could also be applied at DLIFLC, where it could serve primarily prescriptive purposes, including identifying students who should be recycled or relanguaged. It would be possible to develop strategies that can counteract low initial motivation, resulting in higher graduation rates.

Analyzing programs, resources, and outcomes from training

We found that the PEP program produced significant improvements in the success rate of students at DLIFLC. The gains from PEP were in fact much larger than could be obtained from increasing entrance DLAB scores. Conversely, we found that changing to the DLPT 5 standard reduced the success rates of students by 50 percent, an amount that could not be overcome through improved student selection.

Collection of data at DLIFLC on important aspects of training will become increasingly important to DOD. The new performance standards, particularly the speaking portion of the test, will require considerable improvement. Additional data on such aspects of training as resources, instructional approaches, and time use should be collected. Researchers at DLIFLC can provide feedback as to whether changes are producing the desired effects. Also, as performance standards change, it would be important to revisit how DLAB, ASVAB, education, and other student characteristics predict success.

Developing databases that can be used to support both operational needs and analysis

The analyses that we have performed of the linguist screening and development process were accomplished through the development of databases from multiple sources. We collected information from DMDC, MEPCOM, DLIFLC, as well as specific services. The lack of compatibility across databases, as well as incomplete records in many places limited our ability to research particular aspects of screening. DOD would find itself unable to respond to a query on the current status of screening should it need to respond to changing requirements, or need to report to Congress on how well it is achieving particular objectives. The unified databases that can track individuals

throughout the entire screening process from recruiting through training do not currently exist.

Based on our experience with the entire process from recruiting through completion of training, we have developed an outline of the requirements for a tracking database.

First, the database should begin with the recruiting records of all individuals who take the DLAB or DLPT. Some of the elements that the database should contain, in addition to the standard education, ASVAB, and other enlistment information, follow:

- Date and location of DLAB testing
- Date of initial MEPS processing
- Date of enlistment contract
- Date of accession or DEP attrition
- Accession MOS / rating / Air Force Specialty Code.

Currently, the services often record only the DLAB scores of people scoring 100 or above. This eliminates the possibility of evaluating how changes in screening procedures can affect the ability to identify candidates. It is important to routinely collect data on all DLAB test takers.

It is also important to be able to reconstruct when the DLAB test was taken in the enlistment process. Screening costs can be reduced if the testing is done at the same time the candidate is entering the MEPS for his or her physical and other processing. Keeping track of when DLAB testing occurs can enable the services to reduce the number of trips made to the MEPS.

Statistically, it is possible to identify good candidates for testing as soon as they have completed the enlistment application and taken the ASVAB. The kind of information contained in the enlistment record, such as education and AFQT can help identify good candidates for DLAB testing at that time. By reducing the testing of people unlikely to pass the DLAB, the services can reduce their screening

costs. Also, these reductions in the numbers of people who need to be tested with the DLAB can reduce the MEPCOM testing workload.

It is critical to identify which recruits access in linguist positions. The DMDC and MEPCOM records in this area are incomplete. While we were able to supplement these files with service-specific information, it would greatly improve future tracking to be able to identify who was actually classified as a linguist or into other positions that may necessitate language training.

Enlistment records should be merged with currently available DLIFLC student records. This combined accession-training database would facilitate additional analyses of both screening and education programs. In addition, it would help separate out the effects of ASVAB, DLAB, and education, and their contributions toward student performance.

Such a database would support potential operational needs, such as identifying how many people in the services have been screened against DLAB, as well as the numbers meeting different cut scores. Furthermore, as additional data are collected on DLAB test takers, such as information on motivation or details on curriculum, they could be combined with data from the DLAB database. The combined database could be used to evaluate program effects, as well as selection effects in identifying important factors in developing linguists.

This page intentionally left blank.

Appendix A: Complete DLIFLC performance equation

We estimated logistic regression models that tested all hypothesized variables against our two standards of success: completing the course and scoring 2/2/1+ or better on the Defense Language Proficiency Test (DLPT).

The first performance equation we estimated was the probability that a student would complete the basic language course. Table 24 shows the results of a logistic regression for completing the course at DLILFLC. While DLAB score was strongly and positively related to course completion, students taking Category I and II languages (EASY LANG), and who were recycled, were also important factors in explaining completion rates. Students in the Proficiency Enhancement Program (PEP), and who received their desired language (MOTIVE), were also significant positive factors.

Table 24. Logistic regression model coefficients for course completion

Complete = 1	Coef.	Std. err.	z	P>|z|
DLAB	0.021	0.002	8.92	0.000
RELANG	-0.354	0.140	-2.53	0.012
MOTIVE	0.438	0.128	3.41	0.001
RECYCLE	0.646	0.087	7.44	0.000
EASY LANG	0.689	0.096	7.15	0.000
PEP	0.195	0.071	2.77	0.006
Cons	-1.638	0.289	-5.66	0.000

Number of obs = 8,937
LR chi2(6) = 192.18
Prob > chi2 = 0.0000
Log likelihood = 04575.7255
Pseudo R2 = 0.0206

The second part of our model of student success was an equation that estimated the significance of various factors to performance on the DLPT. Table 25 provides the results for all factors hypothesized to be important. Some of the factors that were not significant and were dropped from the final model included direct accessions (YOS0), enlisted personnel, male, and those who were not native English speakers. The final equation that we used to analyze policy changes is provided in table 26.

Table 25. Logistic regression for meeting DLIFLC goal with full set of variables

met_goal=1	Coef.	Std. err.	z	P>|z|
q_dlab	0.024	0.002	10.25	0.000
YOS0	0.006	0.076	0.08	0.939
enl	-0.129	0.141	-0.91	0.362
Other Service	-0.172	0.099	-1.74	0.082
AF	0.056	0.067	0.83	0.404
Navy	0.333	0.091	3.67	0.000
USMC	0.201	0.107	1.88	0.060
Male	0.101	0.059	1.70	0.089
PEP	0.282	0.073	3.87	0.000
dlpt 5	-1.515	0.091	-16.61	0.000
easy_lang	0.695	0.095	7.29	0.000
Recycled	-0.263	0.072	-3.67	0.000
Relang	-0.832	0.148	-5.64	0.000
Motive	0.578	0.149	3.89	0.000
native_eng	-0.307	0.190	-1.61	0.107
edlvl2	0.011	0.060	0.19	0.849
edlvl3	0.228	0.093	2.46	0.014
edlvl4	-0.040	0.105	-0.38	0.703
_cons	-2.325	0.383	-6.07	0.000

Number of obs = 7,000
LR chi2(18) = 630.35
Prob > chi2 = 0.0000
Log likelihood = -4242.9918
Pseudo R2 = 0.0691

Table 26. Logistic regression for meeting DLIFLC goal

met_goal=1 |	Coef.	Std. err.	z	P>|z|
DLAB	0.024	0.002331	10.44	0
EDLVL3	0.232	0.079817	2.91	0.004
RELANG	-0.853	0.146031	-5.83	0
OTHER	-0.211	0.091756	-2.3	0.021
MOTIVE	0.579	0.148201	3.91	0
NAVY	0.283	0.082035	3.45	0.001
RECYCLE	-0.271	0.071294	-3.8	0
EASY_LANG	0.716	0.092556	7.74	0
PEP	0.290	0.072532	4	0
DLPT 5	-1.510	0.088416	-17.08	0
_cons	-2.649	0.302105	-8.77	0

We also did not include service-specific factors, although Navy students were significantly more successful than Army students with the same characteristics. It would be interesting to investigate inter-service differences in the future since the Navy may have approaches to identifying and classifying recruits into languages that could be applicable to the other services. For example, the Navy looks for musical aptitude in assigning students to Chinese.

This page intentionally left blank.

Appendix B: Equivalence of performance on DLPT IV and DLPT 5

Background

Performance at the Defense Language Institute Foreign Language Center is gauged by the Defense Language Proficiency Test. Scoring for the DLPT is based on the guidelines of the Interagency Language Roundtable, and the test results are expressed as levels 0+ through 4 for some languages. Current course goals require students in the basic courses to achieve a proficiency level of 2 in listening/comprehension (L) and reading (R) and a level of 1+ in speaking (S) for most languages.[1]

The current version of the DLPT (DLPT 5) has undergone major revisions from the earlier version (DLPT IV) of the test. The DLPT 5 is more difficult than the earlier version, and it includes more text synthesis and interpretation items. Because the DLPT 5 uses a new score scale, results cannot be directly compared with the previous version.

Methodology

At the time of the data collection (end FY 2006), proficiency testing for all but three languages (Chinese, Spanish, and Russian) used either DLPT IV or DLPT 5. Given this testing pattern, we were not able to make meaningful DLPT form-wise comparisons of performance for most individual languages. Furthermore, because the score scales underlying the two forms of the DLPT were never equated, we believe that it is not advisable to "pool" the data across forms for the analysis of performance differences as they relate to demographics and other factors or policy variables.

1. The speaking test is not required for some languages.

As an alternative to pooling the data across forms, we analyze the effects of these variables for those tested with DLPT IV and DLPT 5 separately. We also determine if any given policy or demographics variable differentially affects DLPT IV and DLPT 5 scores. We use (logistic) regression models that include an interaction term between the policy variable and an indicator variable for the DLPT form. If the interaction is statistically significant, this would indicate that the regression lines are not parallel, implying that the effect of the policy or demographic variable is not the same across DLPT forms.

In these analyses, we pooled the data across specific language courses, thereby comparing performance at the DLPT form level. Because there is a strong correlation between performance (DLPT pass rates, as well as GPA) and the Defense Language Aptitude Battery (DLAB), we hold constant DLAB at the mean of those represented in the data. Thus, comparisons of DLPT form effects are independent of language ability as measured by DLAB.

Results

Overall pass rates by DLPT form

We compare the pass rates, adjusted for differences in DLAB scores, for the three language courses represented in the data (table 27). The pass rates were considerably lower for those tested with DLPT 5.

Table 27. DLIFLC performance for languages tested with both DLPT forms

Language	Form	# obs	Pass rate[a] (L2/R2 S1+)	DLAB	GPA
Chinese Mandarin	DLPT IV	630	0.99	117	3.3
	DLPT 5	105	0.52	117	3.3
Spanish	DLPT IV	470	0.92	105	3.3
	DLPT 5	44	0.29	102	3.3
Russian	DLPT IV	273	0.93	108	3.2
	DLPT 5	71	0.81	111	3.3

a. Adjusted for differences in DLAB scores, by holding DLAB constant at the mean of those represented in the table (111).

While only three languages were tested with both DLPT forms, performance for all languages tested with DLPT 5 was consistently lower than for languages tested with DLPT IV (45 percent vs. 82 percent average pass rate), as shown in table 28.

Table 28. DLIFLC performance for all languages tested by DLPT form

		Number of observations		Pass rate (R2/L2/S1+)	
Category	Language	DLPT IV	DLPT 5	DLPT IV	DLPT 5
I	FRENCH	231	0	0.70	n/a
I	ITALIAN	45	0	0.64	n/a
I	PORTUGUESE	32	0	0.93	n/a
I	**SPANISH**	**494**	**49**	**0.89**	**0.32**
II	GERMAN	50	0	0.91	n/a
II	HINDI	0	15	n/a	0.15
II	URDU	0	27	n/a	0.46
III	HEBREW MODERN	108	0	0.79	n/a
III	KURDISH	6	0	0.00	n/a
III	PERSIAN AFGHAN DARI	0	157	n/a	0.53
III	PERSIAN FARSI	627	0	0.78	n/a
III	PUSHTU AFGHAN	0	205	n/a	0.29
III	**RUSSIAN**	**296**	**96**	**0.84**	**0.82**
III	SERBIAN/CROATIAN	173	0	0.77	n/a
III	SORANI	0	11	n/a	0.00
III	TAGALOG	71	0	0.89	n/a
III	THAI	43	0	0.78	n/a
III	TURKISH	20	0	1.00	n/a
III	UZBEK	7	0	0.00	n/a
III	VIETNAMESE HANOI	2	0	1.00	n/a
IV	ARABIC MODERN	1,779	0	0.82	n/a
IV	**CHINESE MANDARIN**	**712**	**165**	**0.97**	**0.47**
IV	JAPANESE	55	0	0.58	n/a
IV	KOREAN	1,526	0	0.77	n/a

Effects of DLPT form and demographics on performance

In this section, we address the issue of differential effects of demographics and other variables on meeting course goals when performance is measured with DLPT IV or DLPT 5. We use a statistical model to test potential effects. The model includes interactive terms for the demographic variables with an indicator variable for DLPT form. A statistically significant coefficient for the interaction

term would indicate that the change in performance for those in the demographic group represented by the coefficient differed by form.

For example, we found that the pass rate for students not in the Proficiency Enhancement Program (PEP) was 16 percent less when tested with DLPT IV, but only 8 percent less when tested with DLPT 5. The estimates, and those for the other variables in the model were made at the overall mean DLAB score (112). This procedure was used to account for differences in language ability that might have existed for those tested with DLPT IV and DLPT 5. The logistic regression results are shown in table 29, and the DLAB-adjusted estimates of the pass rates are in table 30.

Table 29. DLIFLC performance for languages tested with both DLPT forms[a]

Predictor	Coef.	Std.Err.	Z	P>
dlab	0.025	0.002	10.72	0.000
some college	0.232	0.083	2.80	0.005
relang	-0.875	0.153	-5.72	0.000
motive	0.527	0.163	3.24	0.001
recycled	-0.251	0.073	-3.43	0.001
easy_lang	0.888	0.101	8.79	0.000
yos0	0.072	0.083	0.87	0.387
pep	0.214	0.079	2.72	0.006
dlpt 5	-1.595	0.449	-3.56	0.000
i_some college	-0.248	0.311	-0.80	0.426
i_relang	-0.744	0.661	-1.13	0.260
i_motive	0.243	0.440	0.55	0.581
i_recycled	-0.705	0.392	-1.80	0.072
i_easy_lang	-1.436	0.295	-4.87	0.000
i_yos0	-0.415	0.198	-2.10	0.036
i_pep	0.659	0.204	3.22	0.001
_cons	-2.683	0.310	-8.65	0.000

Y=Met L2/R2/S1+ goal
Pseudo-R2=.07
N= 7,000
Mean DLAB= 112.6

a. Interaction terms are notated with "i_" preceding the variable name.

Table 30. Estimated effects of DLPT form and demographics on DLIFLC performance

	Estimated "met goal" rate at mean DLAB score							
	DLPT 4			DLPT 5				
x-variable	x=1	x=0	Diff.	x=1	x=0	Diff.	DLPT effect	p < .05
Some college	0.67	0.34	0.33	0.73	0.38	0.36	-0.02	
Motive	0.68	0.57	0.12	0.35	0.17	0.18	-0.06	
Recycled	0.63	0.69	-0.06	0.19	0.35	-0.17	0.10	
easy_lang	0.82	0.66	0.16	0.35	0.34	0.01	0.14	*
yos0	0.70	0.68	0.02	0.28	0.38	-0.10	0.12	*
pep	0.75	0.67	0.08	0.44	0.28	0.16	-0.08	*

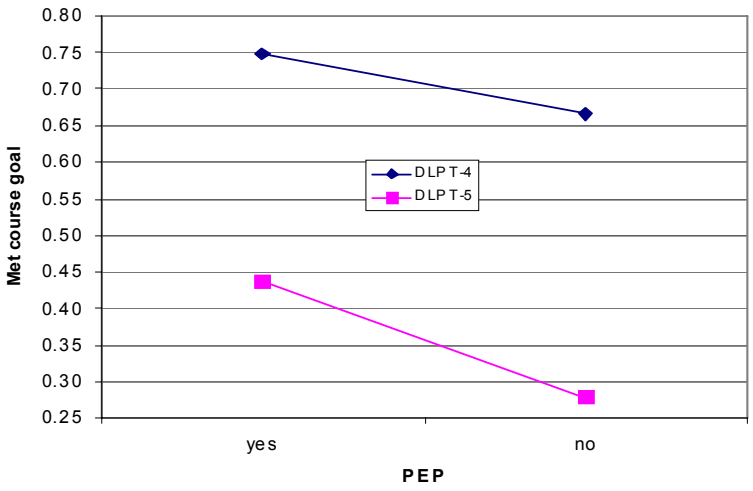
These results suggest that there are statistically significant interactions between the DLAB form a student took and other factors. For example:

- Those enrolled in Category I and II languages performed better with DLPT IV than with DLPT 5.
- Those in their first year of service (yos0) performed better with DLPT IV
- Those not in the PEP program performed worse with DLPT 5.

The Category 1/II language effect is likely an artifact of the different mix of languages tested with DLPT IV and DLPT 5. It is likely to dissipate as a wider range of languages are tested under DLPT 5.

The DLPT effect on PEP is illustrated in figure 22. The results suggest that DLPT 5 is a more difficult test, and those without the extra help afforded by the PEP program, suffer as a result.

Figure 22. Effect of PEP on meeting course goal when assessed by DLPT IV and DLPT 5

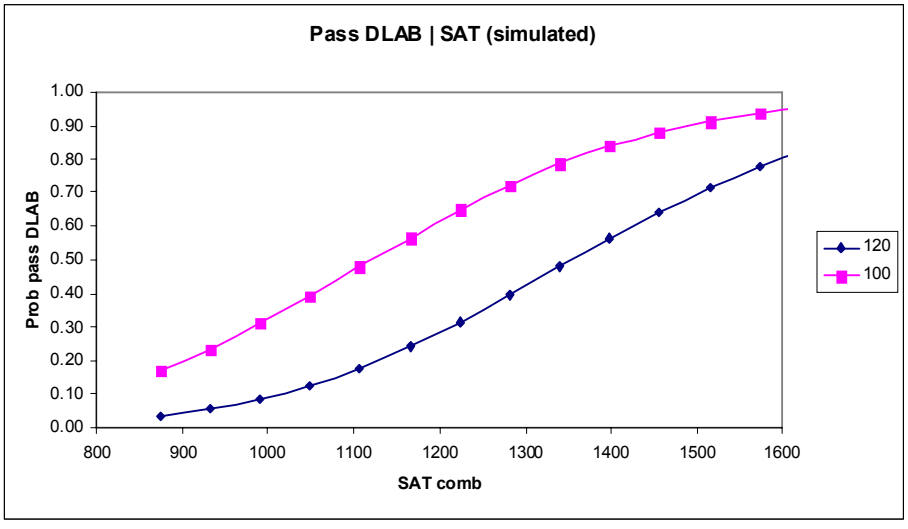


Appendix C: DLAB success rates for USMC officers

Data consisted of 2,500 officers who were tested on the Defense Language Aptitude Battery (DLAB) from 1999-2006. There were 1000 who had SAT, and GCT scores, as well as other background information. (Those missing SAT / GCT data had lower DLAB.) Other predictors that were evaluated included gender, accession source, and performance at The Basic School.

We estimated DLAB "success" rates from SAT / GCT (figure 23). However, note that these data are unlikely to be representative of the population, and motivation should be considered. Those who do not want to become a linguist may have lower scores.

Figure 23. Estimated DLAB success rates from SAT screen



Regression results

We ran a regression using the predictors as shown in table 31.

Table 31. Predicted DLAB score for Marine Corps Officers

Predictor	Coef.[a]
sat_comb	0.011
gct_tot	0.813
NROTC	7.520
TBS_academic	0.475
USNA	7.113
TBS_leadership	-0.884
male	-13.247
Constant	29.687
R-sq = 0.24	

a. $p < .05$.

We found that SAT/GCT composites were better predictors than their components. Females, who only made up 6 percent of the sample, have higher DLAB score (13 points). Additionally, NROTC and USNA had higher DLAB scores (7 points) compared to other accession sources (table 32).

Table 32. USMC DLAB/SAT by accession source

		Means		DLAB pass rate	
Accession source	% sample	DLAB	SAT	100	120
Enlisted Commissioning Program	8	87	1117	0.31	0.06
Meritorious Commissioning Program	6	84	1132	0.27	0.09
Officer Candidate Course	21	96	1075	0.45	0.19
Platoon Leaders Course	24	91	1054	0.37	0.12
Marine Corps Enlisted Commissioning Program	26	86	1132	0.27	0.08
NROTC	8	110	1174	0.69	0.38
USNA	5	113	1263	0.77	0.41
Other	1	101	987	0.67	0.08

Appendix D: Service-specific regressions for the probability of taking the DLAB test

Tables 33 through 36 are the service-specific estimates of the probability that a recruit with an Armed Forces Qualifying Test (AFQT) score of 50 or above would be tested with the Defense Language Aptitude Battery (DLAB) during FY 2008. The regressions were estimated using a logistic formulation of the probability of testing.

Table 33. Probability of taking the DLAB for the Army

took_dlab	Coef.	Std. Err.	Z	P>|z|
afqt	0.071	0.002	30.27	0.000
Male	0.078	0.086	0.91	0.363
some_college	0.083	0.070	1.18	0.240
Single	0.520	0.088	5.9	0.000
Age	-0.001	0.008	-0.19	0.850
pg1	-1.541	0.088	-17.45	0.000
dist_to_meps	0.000	0.001	-0.71	0.477
Perdiem	0.000	0.001	0.54	0.592
ad_gap	-0.001	0.000	-1.62	0.106
_cons	-8.804	0.292	-30.11	0.000

Number of obs = 39,683
LR chi2(9) = 2078.14
Prob > chi2 = 0.0000
Log likelihood = -5406.7822
Pseudo R2 = 0.1612

Table 34. Probability of taking the DLAB for the Air Force

took_dlab	Coef.	Std. Err.	Z	P>|z|
afqt	0.114	0.003	38.42	0.000
Male	-1.143	0.067	-17.16	0.000
some_college	0.227	0.087	2.6	0.009
Single	0.043	0.096	0.45	0.652
Age	0.044	0.015	3.04	0.002
pg1	0.068	0.069	0.98	0.326
dist_to_meps	-0.001	0.001	-2.48	0.013
Perdiem	0.002	0.001	2.98	0.003
ad_gap	0.003	0.000	9.33	0.000
_cons	-12.495	0.439	-28.45	0.000

Number of obs = 21,657
LR chi2(9) = 2586.72
Prob > chi2 = 0.0000
Log likelihood = -4300.0723
Pseudo R2 = 0.2312

Table 35. Probability of taking the DLAB for the Navy

took_dlab	Coef.	Std. Err.	Z	P>|z|
afqt	0.082	0.003	24.99	0.000
Male	-1.579	0.076	-20.65	0.000
some_college	0.452	0.121	3.72	0.000
Single	0.341	0.142	2.4	0.016
Age	0.003	0.013	0.23	0.815
pg1	0.373	0.083	4.52	0.000
dist_to_meps	-0.001	0.001	-0.75	0.451
Perdiem	0.000	0.001	-0.1	0.916
ad_gap	-0.002	0.000	-4.85	0.000
_cons	-9.046	0.460	-19.68	0.000

Number of obs = 23,704
LR chi2(9) = 1134.40
Prob > chi2 = 0.0000
Log likelihood = -3225.8025
Pseudo R2 = 0.1495

Table 36. Probability of taking the DLAB for the Marine Corps

took_dlab	Coef.	Std. Err.	Z	P>|z|
afqt	0.095	0.004	23.81	0.000
Male	-1.399	0.120	-11.65	0.000
some_college	0.208	0.186	1.11	0.265
Single	0.115	0.232	0.49	0.621
Age	0.012	0.022	0.56	0.575
pg1	-0.065	0.094	-0.69	0.489
dist_to_meps	-0.004	0.001	-3.86	0.000
Perdiem	0.000	0.001	0.21	0.833
ad_gap	0.001	0.000	1.63	0.104
_cons	-10.038	0.653	-15.38	0.000

Number of obs = 23,704
LR chi2(9) = 893.50
Prob > chi2 = 0.0000
Log likelihood = -2351.4711
Pseudo R2 = 0.1597

This page intentionally left blank.

References

[1] U.S. Department of Defense. *Defense Language Transformation Roadmap,* Feb 2005

[2] Defense Language Institute Foreign Language Center. *General Catalog 2006-2007* (DLIFLC Pamphlet 350-8)

[3] Eric Surface, Erich Dierdorff, and Jack Donnelly. "Modeling Second Language Proficiency Change for US Special Operations Personnel." *Proceedings of the 19th Annual Conference of the Society for Industrial and Organizational Psychology,* 2004

[4] Dan Segall. *The Utility of the Armed Forces Qualifying Test as a Predictor of Language Aptitude and Proficiency,* 2007 (Defense Manpower Data Center Information Paper)

[5] Jay Silva and Leonard White. "Relation of Cognitive Aptitudes to Success in Foreign Language Training." *Military Psychology,* 1993, Vol. 5

This page intentionally left blank.

List of figures

List of tables