ABSTRACT

Title of Document: THE INFLUENCES OF APTITUDE, LEARNING CONTEXT, AND LANGUAGE DIFFICULTY CATEGORIZATION ON FOREIGN LANGUAGE PROFICIENCY.

Thomas R. Wagener, Doctor of Philosophy, 2015

Directed By: Dr. Steven J. Ross, Second Language Acquisition

This research investigates the ability of predictive measures to differentiate level of language proficiency among learners across languages, language categories, and learning contexts. It fills a gap in the literature pertaining to language categorization and demonstrates differential predictive ability of language learning aptitude measures depending on the language being learned. In addition, it challenges a default assumption that aptitude and other individual difference measures ought to be context independent. This is done through an analysis of the effects of context on the predictive ability of individual difference measures where results show the differing predictive patterns between a foreign language classroom, a domestic intensive instruction setting, and a study abroad program. Finally, several individual difference measures that have shown some past success in differentiating foreign language outcomes for learners are examined to analyze incremental predictive

validity. Measures that demonstrate incremental predictive validity are useful in developing selection protocols for language learning programs. Additionally, measures that show differential incremental predictive validity across language categories and contexts may indicate a potential for aligning learners within a category and context to benefit learner outcomes.

This research provides evidence to support claims that suggest an interactive role between the learner and context leading to differential learning outcomes based on individual differences. It highlights the fact that predictive models of proficiency are not consistent within language category, nor are they consistent across language category boundaries. It shows that a measure of general cognitive memory may be the best indicator of long term language learning success across languages. Finally, it replicates earlier findings that the Defense Language Aptitude Battery (DLAB) provides incremental predictive validity in the face of other individual difference measures indicating that it remains a useful predictor of language learning performance.

THE INFLUENCE OF APTITUDE, LEARNING CONTEXT, AND LANGUAGE DIFFICULTY CATEGORIZATION ON FOREIGN LANGUAGE PROFICIENCY

By

Thomas R. Wagener

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Steven Ross, Chair
Professor Michael Long
Professor Cathy Doughty
Dr. Amber Bloomfield
Dr. Donald Bolger

Acknowledgements

Special thanks go out to the committee members for their time, effort, assistance, and patience throughout this process. Specifically, I would like to thank Dr. Steven Ross for directing this research; his suggestions and contributions were always timely and helpful. His guidance and supervision were invaluable. Dr. Michael Long was instrumental in keeping me on track, and he has always been a true inspiration and mentor during my time at the University of Maryland. Dr. Cathy Doughty and her research have provided a wealth of knowledge and ideas for this project. Dr. Amber Bloomfield broadened my mindset and allowed me to explore additional avenues leading to a more meaningful study. Finally, I would like to thank Dr. Donald Bolger for his time and for sharing his wisdom on aptitude.

In addition, I would like to acknowledge the SLA faculty, staff, and fellow students for their continued encouragement and friendship over the past five years. In particular, Dr. Robert DeKeyser, Dr. Kira Gor, Dr. Nan Jiang, Stephen O'Connell, Susan Benson, Assma and Buthainah Al Thowaini, Man Li, Megan Masters, Ilina Stojanovska, Yuichi Suzuki, and Payman Vafaee. A special thank you to Beth Mackey for her inspiration and assistance during this project. Also, I would like to thank Jim Child and Pardee Lowe for taking the time to meet with me and advise me on this line of research.

Finally, none of this would have been possible without the support of my family. My wife, Yabi, and my children, Gabriela, Vanessa, and Nathan carried a large portion of the burden in what was truly a family effort.

Table of Contents

This Table of Contents is automatically generated by MS Word, linked to the Heading formats used within the Chapter text.

List of Tables

List of Figures

Chapter 1: Literature Review

1.1 The Defense Language Aptitude Battery (DLAB)

The Defense Language Aptitude Battery was designed by Petersen and Al-Haik to serve as the primary selection tool for military personnel to train at the Defense Language Institute (DLI). The DLAB was produced to replace the older Defense Language Aptitude Test (DLAT). The idea was to design a test to meet the following objectives (Peterson & Al-Haik, 1976):

1) Meet or exceed the predictive validity of concurrently available commercial foreign language aptitude tests.
2) Examine the possibility of differential prediction of success by language or language family.
3) Test other predictors which might add incremental predictive validity or clarify correlational relationships.

Petersen & Al-Haik (1976) describe that one of the major considerations in the development of a foreign language aptitude test is to look at the type of curriculum that the student will encounter. Thus, they clearly state that the DLAB is designed for use at DLI, where the training is intensive and has an audio-lingual orientation. Language programs at DLI vary in length, depending on the DOD language category.

Length of Class	Language Category	Languages
64 weeks	IV	Examples: Arabic, Chinese
47 weeks	III	Examples: Russian, Tagalog
36 weeks	II	Examples: German, Indonesian
26 weeks	I	Examples: French, Spanish

As seen above, currently the duration of the programs are 64 weeks, 47 weeks, 36 weeks, and 26 weeks for Category IV, III, II and I languages, respectively. Graduation criteria are consistent across language categories and require that candidates reach Interagency Language Roundtable (ILR) level 2 in Listening and Reading on the Defense Language Proficiency Test (DLPT) and Level 1+ on the Oral Proficiency Interview (OPI). The method of instruction used at DLI at the time of DLAB development followed the "Army Method" described by Carroll (1963), and had four basic characteristics (Petersen & Al-Haik, 1976, p.370):

1) The spoken form was presented and learned before the written form.

2) The method used contrastive analysis of the learner's native language and the foreign language.

3) Overlearning through "pattern practice" was stressed.

4) The desirability of simulating "real life" communication situations was employed.

It is particularly important to note the method of instruction because the DLAB was specifically designed to predict success in this environment, which later in this paper will be described as Intensive Instruction (INI). In the next section on Individual Differences, the DLAB is described as sacrificing construct validity for predictive validity. If this is the case, its predictive validity may be at risk when taken out of context.

To develop the DLAB, Petersen and Al-Haik (1976) used factor analysis to determine the best predictive combination of items from Horne's Assessment of Basic Linguistic Abilities (HABLA) and the Al-Haik Foreign Language Auditory Aptitude Test (AFLAAT). From HABLA and AFLAAT three factors were retained and were grouped as follows:

1) Foreign language grammar rules.
2) Recognition of stress patterns and noun/adjective agreement.
3) Foreign language possessive forms.

Petersen and Al-Haik (1976) cautioned, however, that their results do not allow for a definitive interpretation. Items in the analysis had to have a loading on a factor of 0.20 or greater to be scored on that particular factor. If the item had a loading of that magnitude or greater on more than one factor, it was grouped with the factor where it had the highest loading. The end result was a 90-minute, 119-item auditory multiple-choice test requiring candidates to learn an artificial language (Silva & White, 1993).

The dependent variable for the DLAB development study was course grades, which were converted to standard scores within language before the correlations were computed. Here it is important to note that grades were used as the outcome measure, rather than a more standard measure of language proficiency. This could be problematic if current DOD standards seek to predict proficiency as rated by measures like the Defense Language Proficiency Test (DLPT) rather than course grades. One could expect large correlations between foreign language grades and standard foreign language proficiency measures, but the measures are not identical, and predictive errors could emerge if alternate outcome measures are used.

Fortunately, Silva and White (1993) tested the incremental predictive validity of the DLAB using the DLPT as the outcome measure. The outcome of the study demonstrated that the highest incremental predictive validity for DLAB, using DLPT scores as the outcome measure, was when it was added to a model with g (Silva and White, 1993). So, the factor analysis shows that DLAB is multidimensional in addition to the large g loading. To quickly touch on the outcome measure, the DLPT standardized testing program started in the 1950s, and is currently in its 5th generation of exams. The DLPT program provides testing in three areas: reading, listening, and speaking. The speaking exam is generally referred to as an Oral Proficiency Interview (OPI). For all of the modalities, there are lower range (ILR scale 0-3) and upper range (ILR scale 4-5) exams. The analyses in the current research use scores from the lower range exam. The lower range reading test includes 60 multiple-choice questions based on 36 authentic passages with up to four questions on each passage. The lower range listening test contains 60 multiple-choice questions based on 40 authentic passages, with up to two questions on each passage. In terms of the DLPT, a passage is a short excerpt typically from a news report or an interaction between native speakers of the tested language. The speaking test consists of four parts with 26 questions, including answering personalized questions, narrating events, speaking on selected topics, and role play. DLPT scores are reported using the Interagency Language Roundtable (ILR) Skill Level Descriptions (1985).

In the study mentioned above, Silva and White (1993) showed that multiple correlation increments of the DLAB over a measure of general intelligence (g) and the 10 ASVAB aptitude components were significant for reading, listening and speaking proficiency in all language categories. That was the intent of the study, but they also noted small variations in predictive patterns across language categories. For example, the DLAB had its greatest incremental value for Category II languages in reading and listening proficiency and for Category I and II in speaking proficiency. In any case, the study provides evidence to support the ability of the DLAB to predict outcomes on a language proficiency measure other than simply foreign language grades. The learning context, however, was restricted to intensive immersion.

Thus, the Silva & White (1993) study demonstrates incremental predictive validity and leads them to the claim that “the DLAB may be viewed as measuring the existence of strategies to extract and organize the semantic, syntactic, and phonetic structure of language, consisting of a specific kind of crystallized ability with predictive power beyond that of g” (Silva & White, 1993, p.91). They suggest future research using alternative models with a variety of predictors and individual differences in learning rates at various points in the language-learning process. The current research accomplishes this by using several measures of individual differences and including a time element, where DLPT scores were collected across several years for each of the participants.

1.2 Individual Differences

Researchers have shown that individual differences can be very useful in helping to predict language learning success. Dörnyei (2005) points out that this line of research in SLA can be traced back to educational psychology where personality, ability/aptitude, and motivation "are invariably seen as principle learner variables" (Dörnyei, 2005, p.7). The difficulty, however, comes in trying to define and measure the constructs involved in predicting that success. The following discussion will illustrate key details in the literature about the individual differences that have shown some success in predicting foreign language learning.

Numerous researchers have published studies demonstrating that measureable foreign language aptitude and native language skill indicators can be successful in predicting foreign language proficiency levels for learners (Skehan, 1991; Carroll, 1973; Sparks and Ganschow, 1991; 1993a; 1995a; Humes-Bartlo, 1989; Pimsleur, 1966a, 1968). Carroll (1962) claimed that the language skills forming the basis for foreign language aptitude are:

- phonetic coding ability
- grammatical sensitivity
- inductive language-learning ability.

Sparks et al. (1998) claim that "differences in the oral and written aspects of foreign-language learning are likely related to students' level of native-language skill." They went on to say, "Students with higher levels of both oral and written and both expressive and receptive proficiency in a foreign language achieved significantly higher scores on the MLAT than students who achieved lower levels of proficiency." This suggests that "a standard measure of foreign-language aptitude may provide a relatively good indicator of how proficient one may become in a foreign language, at least after two years of studying that language" (Sparks et al., 1998, p. 207-208). These claims propose that individual differences in foreign language aptitude measures and native-language abilities are good indicators of potential success in second language attainment by an individual learner.

Dörnyei & Skehan (2003) make the argument that "foreign language aptitude and motivation have generated the most consistent predictors of second language learning success since aptitude and motivation do not show particularly high correlations with one another, and they combine to yield multiple correlations which are frequently above 0.50" (Dörnyei & Skehan, 2003, p. 589). In their article, they discuss the development of the Modern Language Aptitude Test (MLAT) by Carroll and Sapon (1959). In Carroll's (1965) description of the test development, he clearly demonstrates that although the goal was to produce a construct valid measure of foreign language aptitude based on the proposed components, seen in Table 1 (below), predictive validity was the true focus of the measure.

Table 1. Carroll’s Four-component Model of Aptitude

Carroll’s Four-Component Model of Aptitude	
Component Name	**Nature and Function**
Phonemic coding ability	Capacity to code unfamiliar sound so that it can be retained over more than a few seconds and subsequently retrieved for recognized
Grammatical Sensitivity	Capacity to identify the grammatical functions that words fulfill in sentences
Inductive Language Learning Ability	Capacity to extract syntactic and morphological patterns from a given corpus of language material and to extrapolate from such patterns to create new sentences
Associative Memory	Capacity to form associative bonds in memory between L1 and L2 vocabulary items

Dörnyei & Skehan (2003) also make the point that the underlying components of the MLAT show an attempt at building construct validity, but that the effort was sacrificed in favor of producing a more predictive measure (Dörnyei & Skehan, 2003, p. 593). Importantly, however, Carroll (1973, 1979, 1981, 1991) clarifies the relationship of the four components of aptitude and aligns them with individual traits of learner memory and language processing abilities. For Carroll, phonemic coding ability meant more than just the ability to perceive and discriminate sounds, but also the ability to code the sound into memory. Grammatical sensitivity and inductive language learning ability are related to language processing, and associative memory concerns the linkages formed between memory items. All of these components, however, allow for individual differences among learners in foreign language aptitude.

During the 1960s, Paul Pimsleur was also studying aptitude and aptitude testing. Pimsleur (1966) produced the Pimsleur Language Aptitude Battery (PLAB).

His idea in producing the battery was to target differences in learner achievement in high school students. Different from the MLAT, the PLAB places greater emphasis on auditory factors and less on memory. Pimsleur proposed that the PLAB could identify remediable learning difficulties, at which point language instruction could then be adapted to meet the needs of the learner, thus indicating that individual differences could be addressed by varying instructional methods.

As mentioned previously, in section 1.1, the DLAB was another early attempt to produce a language aptitude battery. The U.S. Department of Defense (DOD) thought that the language aptitude batteries at the time, particularly the MLAT, did not discriminate well at the higher levels of language aptitude that were required of military members and DOD employees. Petersen & Al-Haik (1976) attempted to meet that need with the DLAB.

The DLAB's component measures were determined by factor analysis. The first factor is measured by a test that uses pictures described by an artificial language. Test takers are required to generalize new combinations of expression in the artificial language. The second factor is measured by the ability to detect stress patterns. The third factor is tested by the ability to apply grammar rules in an artificial language. But, as Dörnyei & Skehan (2003) explain, the DLAB still does not measure pure aptitude sub-components, but rather continues to seek the most predictive combinations of sub-tests.

As a result, Dörnyei & Skehan (2003) claim that these early attempts at developing aptitude measures fail to develop the appropriate construct components adequately and properly link them to their theoretical underpinnings. To further the work of Skehan (1998), who proposed that the components of aptitude could be linked to certain stages of language processing, Dörnyei & Skehan (2003) attempt to link the specific aptitude constructs to the different stages of second language acquisition as seen in Table 2, below (Dörnyei & Skehan, 2003, p.597).

Table 2. SLA Stages and Aptitude Constructs

Second Language Acquisition Stages and Aptitude Constructs	
SLA Stage	**Corresponding Aptitude Constructs**
Input processing strategies, segmentation	Attentional control, *Working memory*
Noticing	Phonemic coding ability, *Working memory*
Pattern Identification	Phonemic coding ability, *Working memory*, Grammatical sensitivity, Inductive language learning ability
Pattern restructuring and manipulation	Grammatical sensitivity, Inductive language learning ability
Pattern control	Automatization, Integrative Memory
Pattern integration	Chunking, Retrieval memory

Dörnyei & Skehan (2003) discuss how individual differences can be found at each of the stages. These individual differences contribute to overall differences in aptitude for individual learners.

Understanding how individual differences in aptitude permeate each stage of second language acquisition is an important step in determining which components of language learning aptitude are most effective in predicting language learning performance. Deconstructing aptitude may also allow researchers to see how some individuals adapt to overcome deficiencies in one area using strengths in another. For example, VanPatten (1996) offers that some learners may be better at segmenting the incoming sound stream, while other researchers suggest that learners differ in such abilities as working memory storage (Miyake & Friedman, 1999; Sawyer & Ranta, 2001; Walter, 2000). As proposed in the work of Wesche (1981), if learner abilities are matched to the appropriate instructional methods, learners can benefit and perform better.

In a more recent study, Linck et al. (2012) also look to detail and analyze the components of language aptitude. The researchers chose a logistic regression model to test which discriminating cognitive factors differentiate language learners with demonstrated high-level language proficiency from others who do not possess such proficiency. In this way, they could identify potential components of high-level language aptitude. Their study examined three groups of individuals; a high-attainment group, a mixed-attainment group, and a non-language group. The high-attainment group included individuals who tested at or above ILR level 4 in any language, worked two or more job assignments that were characterized as ILR level 4, or tested ILR level 3 or better in two or more languages. The mixed-attainment group included individuals who had extensive language training, but did not meet any of the criteria for the high-attainment group. The non-language group did not study a foreign language for more than three semesters in college, they did not study a language at the Defense Language Institute, and they did not live abroad in a non-English speaking country for more than six months. Also, members of the non-language group reported that they did not have extensive experience learning a foreign language.

Linck et al. (2012) state, "The purpose of this study was to obtain empirical evidence of the ability of the High-Level Language Aptitude Battery (Hi-LAB; Doughty, et al., 2007) to distinguish very successful language learners from other individuals" (p.2). The first step was to pinpoint the cognitive components that could make up the construct of high-level aptitude. To do this, the authors probed the literature to see what cognitive components are generally associated with language learning. Then, they defined the various measures for each hypothesized component construct. Once the components and their measures were identified, the authors could use them in a logistic regression model to discriminate the categorical grouping of the tested learners. The Hi-LAB constructs and test components for the study are located in Table 3, below. The Associative Memory measure, Paired Associates, was the only significant predictor in differentiating across all three groups of participants. It was also able to differentiate outcomes in each of the three comparison models (Listening, Reading, and Any-Skill). Implicit learning, as measured by Serial Reaction Time, was also able to single out the high attainment learners from the other two groups in the Listening Analysis. Since the primary intent of the study was to differentiate high aptitude learners, the focus of the remainder of the article shifts away from lower level aptitude. The authors need to make the shift since the currently available aptitude measures seem to only differentiate lower level proficiency learners (Li, 2015). Before leaving the discussion of the Linck et al. (2012) study, however, some other points of interest can be drawn from the results that pertain to individual differences in language learning.

Table 3. Hi-LAB constructs and test components.

Construct	Test
Working Memory	
Executive Functioning	
Updating	• Running Memory Span
Inhibitory Control	• Antisaccade • Stroop
Task Switching	• Task Switching Numbers
Phonological Short-term Memory	• Letter Span • Non-Word Span
Associative Memory	• Paired Associates
Long-term Memory Retrieval	• ALTM Synonym
Implicit Learning	• Serial Reaction Time 12
Processing Speed	• Serial Reaction Time 12
Auditory Perceptual Acuity	• Phonemic Discrimination: Hindi, English Pseudo-Contrastive • Phonemic Categorization Russian

For example, Inhibitory Control successfully distinguishes between the Non-language and the High-attainment groups, but not between the Mixed and High-attainment groups. This may indicate that bilinguals show better inhibitory control than monolinguals, but it also shows that individual differences in inhibitory control may not predict differences in overall language achievement. Also, Phonological Short Term Memory distinguishes between High and Mixed attainment individuals, but does not distinguish between High-attainment and Non-language groups. This result may indicate that the parameter is being masked by other predictive measures, but it could also indicate that this measure is specific to differentiating learner levels in second language learning. In either case, the result may indicate that the measure cannot predict levels of attainment prior to onset of language learning. This is beyond the scope of this paper, but is valuable for high aptitude research. In summary, Linck et al. (2012) use a very different operationalization of foreign language aptitude than has been used in the past. Linck et al. (2012) state, “High-level aptitude is distinct from the more traditional conceptualizations of language aptitude (e.g., Schneiderman & Desmarais, 1988), which typically distinguish rate of learning at lower levels of proficiency within language classroom contexts" (Linck et al., 2012, p. 2). But, the study clearly demonstrates that associative memory is involved in second language learning across proficiency levels. It also provides evidence that implicit learning is important for listening comprehension gains. More importantly, the study emphasizes construct validity while maintaining the predictive validity of the measures. It can also be argued that the study provides evidence for the componential nature of language learning aptitude, which is in line with differential aptitude theory.

Exploring differential aptitude theory in a second language learning context is important for two main reasons. First, it demonstrates the need for a measure of language learning aptitude in addition to more general measures of intelligence (Silva & White, 1993). And second, it expresses a componential nature of language aptitude, allowing that different constructs within a broader aptitude domain can interact in a variety of ways to yield learning results. This complements the work of Pimsleur (1966), Wesche (1981), VanPatten (1996), and other researchers who have maintained that learners can draw on their strengths to achieve a learning outcome. It could also allow for singling out aspects of aptitude that lead to development in specific modalities, as indicated by the implicit learning measure in Linck et al. (2012) and suggested by Lowe (1998), or possibly increased abilities in a specific language category (discussed below). In any case, it calls for the validation of language learning aptitude measures by assessing their predictive validity in the presence of other aptitude measures. It also allows for increased predictive validity if individual difference measures can be paired with particular learning programs to yield the desired outcomes. Lett and O'Mara (1990) also concede this point when discussing the DLAB. “The purposes to which DLAT/DLAB data have been put include both selection for language training and assignment to particular categories of languages. Of course, as more than two decades of research have shown, cognitive ability, even if defined and measured with reference to specific learning domains, is by no means the only learner characteristic that can be meaningfully linked to learning outcomes” (Lett & O’Mara, 1990, p.222).

This, then, directs researchers to examine the literature to seek out other measures that have shown some correlation with foreign language learning in addition to specific measures of foreign language aptitude. These include indicators of native language skills. Sparks et al. (1998) specifically showed that groups with differing second language proficiency levels also scored differently on several native language achievement tests, including the Iowa Test of Basic Skills (ITBS), suggesting some relationship between native language skills and foreign language learning. Feyten (1991) used the Watson-Barker Listening Test (WBLT) to show a relationship between general listening ability and overall foreign language proficiency. Vandergrift (2006) demonstrated that L1 listening comprehension abilities and L2 proficiency are both predictors of L2 listening comprehension abilities. This leads to the claims that developing L2 vocabulary knowledge is critical for L2 listening comprehension, and that with sufficient L2 vocabulary knowledge, learners can transfer listening comprehension abilities from the L1 to the L2, at least to some extent. Carson et al. (1990) showed a relationship between L1 and L2 reading and writing skills for Chinese and Japanese learners of English, but also showed that the relationship depended on L2 proficiency. These studies directly link the native language skills to foreign language proficiency, but they do not take foreign language aptitude into account. In other words, they do not look at incremental predictive validity or the componential nature of aptitude in language learning. That said, it is very difficult to find a study that includes native language skills and foreign language aptitude as predictors of success in a foreign language or in language learning outcomes. In an unpublished study (Wagener, 2014) undergraduate GPA (in the L1) and the Graduate Record Examination (GRE) quantitative scores predicted DLI GPA as the outcome measure of foreign language learning in the presence of the DLAB. The correlation between undergraduate GPA and DLI GPA may be due to a similarity in other factors being measured such as motivation, perseverance, study habits, etc., rather than a correlation between L1 and foreign language learning, but it is one of the few attempts to link the two while accounting for foreign language aptitude. Meanwhile, the correlation with GRE quantitative scores may indicate that general cognitive abilities are at work in language learning.

In summary, past foreign language aptitude measures have focused on predictive over construct validity (Dörnyei & Skehan, 2003). In more recent literature, research has been conducted to isolate specific cognitive constructs that make up second language learning aptitude, but with mixed results. Silva and White (1993) showed evidence of the incremental predictive validity of the DLAB over more general cognitive measures, providing some evidence in support of differential aptitude theory. Additionally, some researchers have inserted native language ability into their methodologies and have discovered a possible relationship between native language skills and second language learning which also may have differential effects on language learning (Sparks et al., 1998; Surface et al., 2004). If language learning aptitude is componential in nature and both general and specific cognitive abilities contribute to language learning, then the DLAB should still be a significant predictor of language learning success in the presence of other measures of individual differences in aptitude and achievement.

1.3 *Learning Contexts*

The focus of research on second language learning contexts has typically been on study abroad or the foreign language classroom in a domestic environment. Of these two contexts, the assumption has often been that study abroad is a quick and perhaps effortless way of improving language proficiency. But, as DeKeyser (2010) argues, "the more nuanced picture that emerges from the literature of the past couple of decades is that accuracy tends to improve little, but fluency more. Even these modest advantages of study abroad are far from firmly established, however" (p. 80). As researchers continue this debate, a third context has entered the discussion, that of domestic immersion. Freed et al. (2004) show that students in an intensive domestic immersion program outperform those in both study abroad and the foreign language classroom in oral performance. Like other researchers, e.g., Martinsen et al. (2010), however, they argue that far too little research has been conducted on domestic immersion programs to make any claims about them.

Before discussing the different contexts further, it is important to define the current operationalization of the three contexts addressed in the current research. The foreign language classroom (FLC) is operationalized as the traditional classroom environment, where students are enrolled in an academic institution seeking a degree or certificate, not necessarily in a foreign language, but take formally instructed classes in a foreign language. These classes typically meet several times a week for an hour or a standard instructional period. At the United States Naval Academy (USNA), foreign language classes teach pronunciation, spelling, vocabulary, and grammar in the basic courses. Intermediate courses expand on the basics and move toward reading, writing, and communicative skills. Finally, advanced courses are taught exclusively in the second language and focus on literary and cultural aspects as well as history and current events of the countries where the language is spoken. Intensive instruction (INI) is an intensive course dedicated to a specific foreign language where students study the language full time. This type of training typically involves four or more hours of instruction on a daily basis, and students are typically encouraged to participate in language learning outside of the classroom. DLI is considered intensive instruction in the current study. At DLI, the classes and type of instruction are similar to the courses taught at USNA, but they are taught at a more intensive pace, and students focus solely on the foreign language. Lastly, study abroad (SA) is a course of instruction where students study a foreign language in another country where the target language is spoken.

The DOD currently offers programs in all three of these learning contexts. Considering all the resources expended training individuals, however, little research has been done to validate predictive measures used in the selection of candidates for the various programs. Many DOD programs seem to have simply borrowed the research done for DLI students and assumed that a predictive measure like the DLAB would be sufficient. This fails to account for the fact that generalization of the measure may be problematic when taken out of context, especially when the construct validity of the measure is disregarded.

Here, a brief discussion of predictive validity and construct validity is also warranted. Predictive validity is involved when a measure intends to predict an outcome on a particular criterion whereas construct validity is involved whenever a measure is to be interpreted as a measure of some attribute or quality (Cronbach and Meehl, 1955). Cronbach and Meehl (1955) describe predictive validity as a correlation coefficient between the predictive measure and the outcome on a criterion where the experimental and sampling conditions are adequately described. In the case of language learning, then, participants would be scored on the predictive measure, followed by the passage of time through some experimental or language learning condition, and eventually leading to a score on the criterion measure. On the other hand, a measure that is construct valid is more concerned with an attribute at the point of departure, somewhat alleviating the need to tie it to the process. Therefore, in order for the DLAB to serve as a predictive instrument for all contexts alike, the assumption must be made that the language learning process is very similar, regardless of the context, or that the DLAB is construct valid. Unfortunately, the literature seems to suggest that neither is the case. To the author's knowledge, DOD has never tested either assumption. The current research will explore aptitude and achievement measures as well as proficiency outcome measures for each of these contexts.

Freed et al. (2004) compared various dimensions of fluency of French students studying in three different contexts; study abroad (SA), intensive instruction (INI), and the at-home foreign language classroom (FLC). The main findings were that the INI group made significant gains in oral performance, outperforming both the SA and the FLC groups. The oral performance measures included total words spoken, length of turn, rate of speech, and a composite fluidity measure. The authors indicate that the INI group reported that they spoke and wrote significantly more in the L2 than the other two groups, and analyses showed that hours spent writing outside of class were significantly associated with oral fluidity gains. This finding suggests that an intensive instruction program may encourage students to spend more time-on-task and therefore allow for more language gains than in the other two contexts. Other researchers, however, were unable to find relationships between time-on-task and the development of speaking proficiency (Ginsburg and Miller, 2000). So, what is it about the learning context of the Freed et al. (2004) study that allowed for time-on-task to differentially affect language gains? Ginsburg and Miller (2000) offer that "we must dig deeper into the qualities and specifics of student experiences, we must understand what students bring to them and how they use them for learning" (p.256). This claim would indicate that it is not the context that differentiates learning gains, but rather the individual learners within the context.

Long (1997) would agree as he clearly points out that although learning takes place in context, it is within the individual learner that mental representations are formed. He adds, “the goal of research on SLA, qualitative or quantitative, inside or outside the classroom, in the laboratory or on the street, is to understand how changes in that internal mental representation are achieved, why they sometimes appear to cease (so-called "fossilization"), and which learner, linguistic, and social factors (and where relevant, which instructional practices) affect and effect the process” (p.319). In other words, Long (1997) recognizes that context is relevant since context affects the type of L2 input and processes (explicit/implicit, attention to form, etc.), but his focus is the learner. DeKeyser (1991) makes a similar point when he discusses the differences between two of the participants in his study, Tim and Paul. He describes Tim as the stereotypical learner who carries a big dictionary around with him wherever he goes, but he says Paul used the language as a “cloak” to attempt to mask himself as a native. Indeed, as the study played out, Paul did sound almost native; Tim, on the other hand, was very tiresome to listen to, since he constantly self-monitored and self-corrected, resulting in very broken speech. DeKeyser (1991) comments, “The main conclusion of this study is that the group differences were far less important than the individual differences” (DeKeyser, 1991, p.115). DuFon and Churchill (2006) echo these observations concerning individual differences, but mention that context still plays a role.

> For the second language acquisition (SLA) researcher, there are perhaps few contexts as potentially rich and complex as study abroad. On the one hand, concentrated time enjoyed by learners in the host context would appear to facilitate significant linguistic gains. On the other hand, pre-departure individual differences interact in complex ways and are affected by the study abroad context, itself conditioned by cultural norms and factors related to program design. Given these interactions, it is not surprising that within-group differences are just as frequently reported as between-group differences...(p.1)

This research suggests that it is the interaction of learner individual differences nested within a learning context which may differentiate learning outcomes. If this is the case, we would expect differential effects of context on the DLAB's ability to predict learner outcomes.

In Collentine and Freed's (2004) summary of the literature to date on learning context, they are particularly aware of the debate about where the focus of SLA research should be. They mention Ellis (1994), whose primary focus is on the development of cognitive accounts of SLA, but claims that language acquisition is powered by the internal and external pressures on the learner that come from context. This is different from Batstone's (2002) definition of context, but he makes an important point regarding context that is pertinent in the current research. Batstone's (2002) definition of context is spelled out here:

> Communicative contexts require that the learner use the L2 as a tool of sorts for exchanging information and participating in important social and interpersonal functions. Learning contexts are those in which input and learner output are fashioned (normally with the assistance of a teacher) so that learners will attend to form and take risks toward the ultimate goal of improving their linguistic expertise. (Collentine and Freed, 2004, p.155)

Batstone (2002) says that in communicative contexts, learners may be focused more on meaning and less on form, and therefore may not be as interested in furthering their linguistic development. This context would more closely represent a study abroad setting for most learners. Then, as Batstone (2002) describes the learning context, it would more closely resemble the typical language classroom (FLC). Intensive instruction (INI), however, would find itself more in the middle. All of the contexts; FLC, SA, and INI, have communicative elements and learning elements, and, depending on the specific study program, may shift as to where they fall on a Batstone (2002) continuum. This may mean that certain components of foreign language learning aptitude may play more important roles in one context or another.

Collentine and Freed (2004) make several additional points about FLC, SA, and INI. First, they comment that “although pedagogues have made great strides in creating tasks in which formal classroom learners use the L2 as a communicative tool, it would be difficult to argue that such learners regularly confront the affective variables that are built heavily into social and interpersonal functions of their L2” (p.155). Second, in the intensive instruction context, the attempt may be to imitate a study abroad context, but the surrounding culture is the L1, so the language is not embedded in authentic cultural situations. Third, in the study abroad context, learners negotiate the communicative contexts, and attempt to use explicit knowledge attained in the classroom, but the communication may lack some of the risk-taking behavior that only emerges after interpersonal relationships develop. In other words, limits are placed on learning in all three contexts, but according to Collentine and Freed (2004), the learners’ willingness to take risks in conversation, is particularly apparent in the SA context.

The question remains, then, whether the context is differentially beneficial for a learner with particular characteristics or skill sets. Many educators have long seen study abroad as the ideal learning context across the board. Kinginger (2008) spends some time discussing the history of research on language study abroad. She references a national assessment done by Carroll (1967), who makes a very positive claim about study abroad.

> Time spent abroad is clearly one of the most potent variables we have found, and this is not surprising, for reasons that need not be belabored. Certainly our results provide a strong justification for a "year abroad" as one of the experiences to be recommended for the language majors. Even a tour abroad, or a summer school course abroad is useful, apparently, in improving the student's skill. (Carroll, 1967, p.137)

Freed (1998) summarized SA research to date and made a generalization about study abroad students: "Those who have been abroad appear to speak with greater ease and confidence, expressed in part by a greater abundance of speech, spoken at a faster rate and characterized by fewer dysfluent-sounding pauses. As a group, they tend to reformulate their speech to express more complicated and abstract thoughts, display a wider range of communicative strategies and a broader repertoire of styles" (Freed, 1998, p.50). Lafford (2006), also, mentions that study abroad has always been thought to provide the best learning environments for acquiring a second language and learning about other cultures. But what is the true value of study abroad as a language-learning tool? Kinginger (2008) points out that the number of students from the United States enrolling in study abroad programs is increasing, but "the relationship between study abroad and language learning is highly complex and changing." (Kinginger, 2008, p.2).

DeKeyser (1991) observed that second language acquisition in a study abroad context is complex, meaning that there is more at work than just context, and reiterating that learner individual differences play a significant role in linguistic development in addition to the context. DeKeyser (2013) argues that study abroad does not necessarily produce better results than domestic immersion programs, nor does it necessarily produce measureable language gains in a number of areas. He does mention, however, that when gains are made it is normally in fluency, not accuracy or complexity. Lafford (2006) suggests that due to the variety of methodological design features of studies showing an advantage for SA over FLC in one area or another (Segalowitz & Freed, 2004; DeKeyser, 1986; Stevens, 2001; Collentine, 2004; Lafford, 1995, 2004; Rodriguez, 2001; Torres, 2003), it is difficult to generalize findings. She invites a reexamination of the factors involved in the process of acquiring a second language. Collentine and Freed (2004) come to a similar conclusion and say that no one learning context (SA, INI, FLC) is "uniformly superior to another for all students, at all levels of language learning and for all language skills" (p.164). Martinsen et al. (2010) say, "Interaction with native speakers is one of the most widely studied variables relating to improvement in oral language skills in study abroad (Brecht, et al., 1993, Freed, Segalowitz & Dewey, 2004; Keating, 1994; Segalowitz & Freed, 2004), though there is little consensus as to its benefits" (p.47). All of this indicates that learner differences may moderate their language gains, but the interaction of individual differences with context may also affect language learning.

Realizing that study abroad may not work for everyone, and with the increased costs associated with SA, many schools also offer intensive instruction programs. Considering the above arguments and study results, some may see INI as a replacement for SA; but, does INI measure up to SA? Collentine and Freed (2004) say that INI programs often meet or exceed the language gains found in SA programs.

> Among the studies reported here [SSLA 2004], perhaps the greatest surprise derives from the fact that students in the SA context do not emerge as those with strengths superior to those who spend periods of time in an [INI] context. Many [INI] students tend to make greater

> gains in the areas studied—in both the oral and literate domains—than do their SA counterparts. (Collentine and Freed, 2004, p.164)

Dewey (2004), for example, showed that scores between SA and INI learners of Japanese only differed on a self-assessment of reading ability, but not on measures of comprehension/free-recall and vocabulary knowledge. Dewey (2008) showed that INI students outperformed SA and FLC students in producing words in complete sentences; INI students also showed a greater knowledge of less frequent words than SA and FLC students; and, INI students showed overall similar performance to SA students in vocabulary acquisition. Martinsen et al. (2010) compared three groups of students looking at language gains. Their three groups included the typical SA group, a service-oriented SA group, and an intensive instruction group in a foreign language house on campus. The typical SA group requires no further explanation, but the service-oriented SA group and the domestic foreign language housing group merit a quick description. The service-oriented SA group spent time in a SA setting, but also were required to study a particular academic discipline and then engage in some form of service related to that discipline to benefit the members of the local community. Foreign language housing (FLH) in an intensive instruction program "is a language learning context in which students (1) live together in an area designated as foreign language housing, (2) commit to speaking exclusively in the target language while in the foreign language housing, and (3) are often encouraged or required to participate in certain activities designed to increase use of the target language or understanding of the target culture such as preparing and eating dinner together and/or participating in cultural or social activities" (Martinsen et al., 2010, p. 47).The researchers showed that all three groups showed similar gains on all language measures. Finally, Freed et al. (2004) showed that their INI group outperformed their SA and FLC groups. Looking at this evidence, then, it would appear that an INI program is at least as beneficial as a study abroad program.

Most researchers, however, would caution against this conclusion for several reasons. First, most would admit that the limited literature available on INI programs does not allow for such drastic claims to be made. Freed et al. (2004), for example, state several times that there have only been a handful of studies comparing SA programs to INI programs, and that there are only a small number of qualitative studies that explore learning in the INI setting. Second, as previously mentioned, the research results are inconclusive; sometimes showing more linguistic gains for learners in the INI setting, and sometimes showing greater advances for learners in the SA setting. Third, many researchers, particularly in sociolinguistics, would be concerned about the lack of culturally based language learning. Even Collentine and Freed (2004) say that students studying the L2 in a context where the surrounding culture is their L1 will find that the interaction "may not be totally natural, given that it does not always involve contact with native speakers nor is interaction embedded in authentic target culture situations" (p.156). Lantolf and his colleagues (Lantolf, 1994, 2000; Lantolf & Appel, 1994) would argue that form-meaning associations made by learners are informed by the situational and cultural phenomena surrounding them; therefore, associations formed in SA programs would differ from those formed in INI programs. Also, as with SA and FLC, INI program designs are numerous, so researchers must be cautious when claiming a particular setting is superior, and as Martinsen (2010) makes quite clear, very little research has been done on these types of programs, at least in comparison to SA and FLC programs. Once again, however, researchers recognize that results within contexts vary, which could indicate that learner differences are at play.

The last context, the foreign language classroom (FLC), is often used as a baseline in comparative studies of language gain, even though the variety of foreign language classroom programs should preempt researchers from concluding that the programs are similar enough to establish a baseline in the first place. But assuming there were such a thing as a stereotypical FL classroom, Freed (1991) claims that the role of instruction in this context includes "the teaching of specific structural features, the respective roles of grammar and communication, the role of error correction, and the role of environmental factors of the classroom" (Freed, 1991, pp.12-13). This description is very much in line with Tarone and Swain's (1995) claims that learners in the FLC context are mostly exposed only to academic/formal registers, although, to their credit, foreign language classrooms have evolved considerably since then to include more communicative methods. Lafford (2006) provides a concise summary of the impoverished input in FL classrooms in the following excerpt.

> The input received by classroom learners has traditionally been limited to NNS or NS teacher talk and NNS peer language, with input modified through the negotiation of form or meaning. With the current wide availability of authentic materials from target language/culture videos, DVDs and the Internet, students are now able to be exposed to more authentic language input. However, this exposure is very often sporadic and classroom learners normally have little chance to hear/read frequently the same vocabulary items in various contexts to create multiple links among sensory experiences. (Lafford, 2006, p.5)

The conclusion appears to be that classroom learners are disadvantaged by far less authentic input than learners in an immersion program. Additionally, the average classroom exchanges tend to be at the sentence level or below (Lafford, 2006), and each individual learner would appear to have less opportunity to participate in authentic conversation. So far, these researchers are focusing on the negative aspects of classroom learning. But, Lafford (2006) does acknowledge that "processing of input is facilitated in classroom contexts, due to the fact that the student's working memory is not overtaxed with too much target language input to retain and process while formulating a response to his/her interlocutor" (Lafford, 2006, p.5). Additionally, she comments that the FLC context allows both the learners and instructors to concentrate on learner comprehension and output and on the development of the learners' L2 systems. Finally, Lafford (2006) comments that the additional time to focus on form and meaning allows learners' to notice gaps between their own interlanguage and the target language. Other researchers also concede that the FLC context offers valuable learning opportunities and even outperforms SA in certain aspects (Collentine and Freed, 2004). Similarly, DeKeyser (1991) states, "The results of our study, then, do not suggest a strong dichotomy between language learning in the classroom and picking it up abroad, or between grammar and oral proficiency" (DeKeyser, 1991, p.115), indicating that context, by itself, does not explain differences in language gains between FLC and SA learning.

Kinginger (2008) explains that despite the context of learning, it is the quality of the learning experience that helps to bring about language gains. However, as mentioned earlier, context cannot be totally disregarded. Context still does play a role in SLA, in that it allows learners a variety of opportunities to build form-meaning associations, and depending on the specific context, these form-meaning representations may be biased. “Atkinson (2002), taking a connectionist perspective, proposes a sociocognitive approach to the study of SLA in which it is recognized that language in the brain is interconnected with the experiences and emotions from the context in which it is acquired” (Lafford, 2006, p.3). Tarone (2007) claims that there is empirical evidence to support a model of “the relationship between social context and second language use and acquisition, which shows that learners' second language (L2) input and processing of L2 input in social settings are socially mediated, that social and linguistic context affect linguistic use, choice, and development, and that learners intentionally assert social identities through their L2 in communicating in social contexts” (Tarone, 2007, p.837). Selinker and Douglas (1985), also, suggest that adult L2 learners establish internal discourse domains that are derived from the particular forms and structure based on their perceptions of the social setting in which they find themselves. Form refinement may also be affected by context, as is evidenced by better development of phonetic and phonological abilities in study abroad contexts (Diaz-Campos, 2004; Simões, 1996; Stevens, 2001). Lafford (2006) herself states: “In both classroom and study-abroad contexts, the purpose of a given communication, and a concomitant focus on either form or meaning, may shift dynamically according to changing learner and interlocutor needs within a conversation in either context” (Lafford, 2006, p.8). This evidence tends to indicate that context affects learning or at least what is learned. The sociocultural literature may point more to the macro-effects on language and the subsequent effects on the language learner while the cognitive literature may focus more on the micro-effects inside the minds of the learners, but in either case an argument can be made that context plays some role in language learning, no matter how small.

The discussion to this point demonstrates the complex nature of language learning and the interactive role between the learner and context. If language learning truly differs based on the individual nested within a particular context, then one would expect that the ability of aptitude measures to predict learning gains would also vary between contexts. Stanhope & Surface (2014) suggest the importance of predictor-criterion alignment in different learning contexts and state that “it is reasonable to expect individuals with specific abilities that align with training content to have a higher likelihood of success” (Stanhope & Surface, 2014, p. 152). This interpretation would also be in line with Pimsleur’s (1966) argument when he was developing the Pimsleur Language Aptitude Battery (PLAB). The intent of the PLAB is to discover learner strengths and weaknesses and to adapt teaching methods to better align with learner abilities. The biggest assumption in this argument is that aligning teaching methods to learner strengths would indeed increase learning for a particular individual. If Pimsleur (1966) is correct, the DLAB and other measures of individual differences will vary in their ability to predict language gains depending on context. The current set of studies will also attempt to validate the claims of Linck et al. (2012) that the DLAB differentiates the rate of language learning in FLC contexts specifically, but does not necessarily distinguish language learning rates in other contexts. No study, to this author’s knowledge, has directly looked at possible differential effects of context as defined in this study (FLC, INI, and SA) on the predictive ability of individual differences. Therefore, this study probes these three contexts to analyze and test the default assumption that aptitude and other individual difference measures ought to be context independent.

1.4 Language Categorization

Lett & O'Mara (1990) clearly state that one of the uses of the DLAB is to determine probable success of learners in a particular category of languages. This notion has been challenged by authors such as Child (1998) and Lowe (1998). Child (1998) mentions that items appearing on the DLAB are confined to word and phrase segments roughly similar to English in length and part-of-speech category, thus making the DLAB the preferred foreign language aptitude measure for category I and II languages. He claims VORD, on the other hand, is tailored to predicting success in languages that have far different syntactic patterns and structures than English, making it the better measure for category III and IV languages. Lowe (1998, and in personal conversation) makes the point that language categories were formed for practical purposes, where time to train was the most important factor. If a language aptitude measure can be tailored to specific linguistic features that vary across languages, then the possibility exists that certain aptitude components could play a larger role in one language category than another. Although in many cases, a 'less is more' interpretation may be preferred; especially since Child (1998) points out that the "distances" from English can vary with time. However, a broad categorical division that aligns with a particular aptitude component may still allow for increased predictive validity. This would indicate that a measure like DLAB (and other individual difference measures) could certainly vary in its predictive ability across language categories, as suggested by Child (1998). Therefore, the current language categorization system and predictive measures used to select learners for study in a particular category warrant a closer look.

Simply stated, language categorization is the division of languages into groups. Lowe (1998) explains that the current language categorization system of the Defense Language Institute Foreign Language Center (DLIFLC) "aids in planning training, but it clusters together languages whose common features may cause Americans difficulties in learning, yet whose nature can differ radically in structure and thought patterns from language to language" (Lowe, 1998, p.17). He points out that the system is efficient in establishing schedules based on time to train, but does not necessarily group languages according to the types of difficulties they involve for learners. Lowe (1998) comments, that for native speakers of English, the most difficult languages to learn are Arabic, Chinese, Japanese, and Korean. These languages do share the difficulty of different writing systems, but Lowe (1998) says "from that point on there are more divergences than commonalities." These four languages make up Language Category IV, along with the recent addition of Pashto. Appendix A shows a list of languages by DLIFLC language category. Category I languages are considered the easiest to learn for native speakers of English, followed by Category II, III, and then IV. As mentioned earlier, current lengths of study for programs are 64 weeks, 47 weeks, 36 weeks, and 26 weeks for Category IV, III, II and I languages, respectively. Required DLAB scores to study a language in a particular category are identified in figure 1, below.

Qualifications		
Categories	**Score**	**Language Description**
Category I	95	Dutch, French, Italian, Norwegian, Portuguese, Spanish
Category II	100	German, Malay, Indonesian, Romanian
Category III	105	Czech, Farsi, Polish, Russian, Serbo-Croatian, Tagalog, Thai, Turkish, Uzbek, Vietnamese
Category IV	110+	Arabic, Chinese, Japanese, Korean, Pashto

Figure 1. Required DLAB scores by category

Child (1998) agrees with Lowe (1998), stating that language categories were created more for practical reasons than for actual commonalities among the languages within the categories. Although no claim has ever been made as to the construct validity of the language categorization system, Child (1998) argues the need for further discussion as to matching language difficulties based on phonology, with provision made for written representation, grammatical system covering morphology and syntax, and semantics. He says each of these three should be rated with respect to their distance from English, and then languages should be compiled into their language category based on these distances. Next, Child (1998) discusses how "learning difficulty is tied to the degree in which the object of learning resembles something already known" (Child, 1998, p.6). He suggests that aptitude tests can then be designed to predict success in a specific language or language category. He states that aligning the predictor to the criterion is the best way to identify aptitude for a particular language category. In this way, he appears to be seeking a better linkage between aptitude and the domain where it intends to predict outcomes. He states that the DLAB may be better suited for certain languages like German; whereas VORD may be better suited for languages like Japanese. His discussion may shed some light on why Lett & O'Mara (1990) found considerable variation in DLAB scores for learners within language categories for learners who were successful. If languages within a category are substantially different, and aptitude measures pick up on certain traits within a particular language that make it easier for a certain learner, then variation within this complex environment would also be substantial. Child (1998) is adamant about the importance of aligning languages in a more meaningful way.

> The entire "language aptitude" enterprise could falter in the absence of a comprehensive overview of similarities and differences among the major languages of the world. There have been over the years a number of attempts to categorize languages in terms of their presumed difficulty; which is to say, how hard they are to learn for native speakers of English. Several of these efforts have in fact been officially blessed within a number of government agencies because they have a certain face validity and have proved useful as general guidelines. (Child, 1998, p.15-16).

Lett & O'Mara (1990) showed that higher DLAB scores indicate an increased probability of success in a course of study at DLI for all four of the language categories. Success in their study was meeting the graduation requirements at DLI. Once again, graduation criteria are consistent across language categories and require that candidates reach ILR level 2 in Listening and Reading on the Defense Language Proficiency Test (DLPT) and Level 1+ on the Oral Proficiency Interview (OPI). Lett & O'Mara (1990), however, did not find evidence for the ability of the DLAB to predict learning outcomes within language categories. In other words, students had a higher chance of graduating if they met the minimum score guidelines for the particular language category, but a student who scored higher did not necessarily reach higher levels of proficiency within that category.

In Wagener (2014), the DLAB did predict success within language category, but only for reading proficiency as measured by the DLPT, and only for students of DLI at graduation. The DLAB was unable to predict proficiency gains later in learning for these same students, although by that point they were in the higher proficiency ranges (above an ILR scale score of 2). Silva & White (1993) were able to show incremental predictive validity for the DLAB, again for DLI students, above other aptitude measures within categories. This indicates that the DLAB should predict proficiency gains since their outcome measure was the DLPT. As mentioned earlier, there was variation in the incremental validity patterns across language categories. This may demonstrate differential predictive patterns for the DLAB across languages. This is in line with Child (1998) and Lowe (1998), and may indicate different aptitude components are differentially utilized among languages. Ideally, then, if languages were appropriately categorized by distance from English in each of the aptitude components, then sorting tools could be created to more efficiently predict success among learners.

Surface et al. (2004) comment that their results show that language difficulty had a significant negative relationship with initial proficiency and proficiency growth, demonstrating that language difficulty has a larger effect than cognitive ability (as measured by the ASVAB) on language proficiency. This could indicate that the cognitive measures used do not differentiate well within category. Once again, if aptitude components could be identified that predict proficiency gains within language categories, then measures could be created to more effectively predict learner success. That, in fact, was one of the goals of Child (1998): to develop an aptitude measure that could be used for specific language categories and, possibly, higher attainment learners. That said, looking at the results discussed above, the expected finding is that DLAB will continue to predict success across categories, but differences of languages within category will yield the variation discovered by Lett & O'Mara (1990).

Research and discussion in the literature beyond Lett & O'Mara (1990), Silva & White (1993), Child (1998), and Lowe (1998) on language categorization is sparse. Much of the research has turned to higher attainment aptitude measures, but has neglected language category. This could be allowing unneeded variation to enter foreign language aptitude studies. The current study refocuses on the issue of language categorization to uncover predictive patterns among various aptitude and achievement measures, in order to further the research of Child (1998) and identify possible indicators of success within and across language categories. This research lays the foundation for aligning predictor and criterion within language categories.

1.5 Conclusion

This review identifies several areas of focus that are absent from the current literature on aptitude measures in second language acquisition. For example, if aptitude measures predict differential outcomes based on context and language categorization, high stakes Department of Defense and other programs should be made aware and account for this in selection of their candidates. Additionally, if other aptitude measures add predictive validity in certain circumstances, then selection protocols should incorporate them as well. These complexities will be investigated in the chapters that follow.

Chapter 2: Purpose of the Study

For many years, the Department of Defense (DOD) has sought to train professionals in world languages. More recently, the desire for language professionals and analysts in the DOD has seemingly grown exponentially. Current U.S. involvement in world affairs has DOD linguists stretched across the globe. Additionally, DOD has sought to train individuals in more specialized languages, which has added pressure to the current training system. In order to streamline the system, and avoid unnecessary expenditures, the DOD has made an effort to identify candidates with the highest probability of success, based on prediction models. Although there is often a high degree of overlap in program training, selection criteria may vary, based on specific program needs and desired outcomes, but the central predictor for nearly all of the DOD selection models is the Defense Language Aptitude Battery (DLAB).

The DLAB was developed in 1976 by Petersen and Al-Haik to select candidates for the Defense Language Institute (DLI). DLAB scores are used to identify individuals who have an increased probability of success in language learning, and the scores also help to determine into which language categories individuals should be placed (Lett & O'Mara, 1990). The assumption is that the higher the DLAB score, the better the chance of overall success and the better the chance of successfully learning more difficult languages. But, Lett & O'Mara (1990) caution that although the DLAB is a valuable predictor of success, it is not the best predictor of language proficiency gains within a language category. Linck et al.

(2012) suggest that the reason that the DLAB may not be the best predictor of language proficiency gains is because measures like the DLAB "distinguish rate of learning at lower levels of proficiency within language classroom contexts" (Linck et al., 2012, p. 2). So, where Lett & O'Mara (1990) say the DLAB's predictive ability comes into question based on language category, Linck et al. (2012) also limit its ability based on proficiency level and context.

Interestingly, and in contrast to the statement of Linck et al. (2012), several DOD language training programs do not see the value of the DLAB at predicting success in lower proficiency learners and are using other aptitude measures for program selection decisions. In one program, educators attempted to rely on other individual difference measures for selection into foreign language studies rather than the DLAB. Based on the results of Surface et al. (2004), the program required a score of 620 or higher on the SAT math section in order to study Chinese or Arabic, while there was no requirement for DLAB at the time. Another DOD program still relies on a variety of aptitude and achievement measures to make its selections rather than trusting the DLAB as a sole predictor of language learning success. This approach is in line with the findings of Silva and White (1993) that suggest an incremental predictive validity when several aptitude measures are used to predict language learning success. Additionally, this program chooses not to follow the guidelines established at the Defense Language Institute Foreign Language Center (DLIFLC) for recommended minimum scores based on language category, but rather, allows candidates to self-select any language as long as they have a minimum DLAB score of 95. Doubt as far as assigning students to a language category based on DLAB score may come from comments made by authors such as Child (1998) and Lowe (1998), who state that the current language categorization system is based on time to train rather than actual “language distances” from English. Child (1998), for example, mentions that an alternate language aptitude measure, VORD, may be more appropriate than DLAB for predicting success in some languages.

These concerns with the DLAB may have found their roots in the overgeneralization of the DLAB’s capabilities. Perhaps researchers like Lett and O’Mara (1990) and Linck et al. (2012) are indirectly stating that a reassessment of the DLAB is necessary, where research into the limitations of its predictive ability will help guide future selection criteria for many DOD programs. That is one goal of this current research project. Therefore, taking into account the claims of Lett and O’Mara (1990), Linck et al. (2012), the findings of Silva and White (1993), and the assertions of Child (1998) and Lowe (1998), there are four major questions to be answered. First, does the DLAB predict foreign language learning success for lower proficiency learners? Second, how does the DLAB compare with other individual difference measures in predicting foreign language learning? Third, does learning context play a role in how well the DLAB predicts success? Finally, how well does the DLAB align learners with DOD language categories, and how well does it predict learning success within those categories?

This study examines each of these questions in turn. The DLAB is the central focus of the study, but equally important for DOD program selection criteria is the focus on other individual difference measures and how they play a role in predicting language learning success. Additionally, this study considers the flexibility of the

DLAB and other individual difference measures in predicting success across language learning contexts and learner proficiency levels, and shows that selection protocols may need some modification based on the language and context. Finally, the study examines DOD language categorization, which Child (1998) explains has a certain face validity, but little empirical evidence to support it.

This research seeks to provide crucial evidence for predictive measures of language proficiency growth and fill a gap in the literature pertaining to language categorization. As the DOD continues to seek out efficient measures for use in selecting the best candidates for language training, this research adds important evidence to help design appropriate decision matrices. Additionally, this research identifies critical similarities and differences in language categories that will help the DOD to better align candidates with languages where they can truly excel. The importance of finding where the DLAB has predictive value cannot be understated, considering the resources the DOD spends to train individuals in a variety of settings and schools.

Chapter 3: Research Questions and Hypotheses

3.1 Research Questions

Silva and White (1993) showed evidence of the incremental predictive validity of a foreign language aptitude measure (DLAB) over more general cognitive measures, providing evidence in support of differential aptitude theory. Additionally, other researchers have discovered a possible relationship between native language skills and second language learning (Sparks et al., 1998; Surface et al., 2004). This may indicate that language learning aptitude is componential in nature and both general and specific cognitive abilities contribute to language learning. This complex nature of language learning is further complicated by the learning environment and instructional methods to which the learner is subjected. Stanhope & Surface (2014) suggest the importance of predictor-criterion alignment in different learning contexts and state that "it is reasonable to expect individuals with specific abilities that align with training content to have a higher likelihood of success" (Stanhope & Surface, 2014, p. 152). If language learning differs based on the individual nested within a particular context, then one would expect that the ability of aptitude and achievement measures to predict learning gains would also vary between contexts. This dissertation research directly investigates the componential nature and the predictive ability of individual differences on foreign language learning in several different learning contexts (FLC, INI, and SA). Additionally, this research probes these learning contexts to analyze and test the default assumption that aptitude and other individual difference measures ought to be context independent. This will be achieved by looking at the predictive ability of several aptitude and achievement measures in order to answer the following research questions.

- Research Question 1: Do individual differences in verbal, quantitative, and foreign language aptitude and individual differences in L1 skills predict differences in foreign language proficiency in a foreign language classroom (FLC) environment?
- Research Question 2: Do individual differences in verbal, quantitative, and foreign language aptitude and individual differences in L1 skills predict differences in foreign language proficiency in an intensive instruction (INI) environment?
- Research Question 3: Do individual differences in verbal, quantitative, and foreign language aptitude and individual differences in L1 skills predict differences in foreign language proficiency after a semester study abroad (SA)?
- Research Question 4: Do the magnitudes of the coefficients in predictor models of language success vary for the independent variables (L1 skills, measures of verbal, quantitative, and foreign language aptitude) in different learning environments?

Additionally, this dissertation research refocuses on the issue of language categorization to uncover possible predictive patterns among various aptitude and achievement measures. This refocusing is in an effort to further the research of Child

(1998) and identify possible indicators of success within and across language categories and to lay the foundation for aligning predictor and criterion within language categories. This will be achieved by looking at the following research questions.

- Research Question 5: Do individual differences in verbal, quantitative, and foreign language aptitude and individual differences in L1 achievement differ in how they predict foreign language proficiency based on language category?
- Research Question 6: Do the magnitudes of the coefficients in predictor models of language success vary for the independent variables (L1 skills, measures of verbal, quantitative, and foreign language aptitude) within the same language category?
- Research Question 7: Do the patterns in the magnitudes of the coefficients of achievement measures, measures of verbal, quantitative, and foreign language aptitude and foreign language proficiency vary across language categories?

3.2 Expected Findings

Table 4 displays the results of the Wagener (2014) study, which looks at language proficiency growth for a group of students as they progressed through a three-year DOD language training program. The students in the program trained at DLI in a particular language and then continued training in a study abroad setting for an additional two years. The individual difference measures that had a significant impact on the growth model for each one-year period are identified by an asterisk.

Table 4. Predictors of language proficiency growth.

	Defense Language Institute Proficiency Growth Models				Study Abroad Foreign Language Proficiency Growth Models					
	DLI GPA	Listening DLPT	Reading DLPT	OPI	Listening Growth			Reading Growth		
Predictors					Year 1	Year 2	Both Years	Year 1	Year 2	Both Years
GPA	*									
GREQ	*		*[1]		*	*[2]				*
GREV						*				
DLAB			*							
Language Cat.		*	*							√

√ - marginally significant predictor (P < 0.10)
* - significant predictor (P < 0.05)
1 - becomes significant with the addition of Language Category to the model
2 - becomes significant with the addition of GREV to the model

Based on these results and the literature presented in this proposal, the expected findings for the current research are presented in Tables 5 & 6. In Table 5 the larger column headings represent the learning contexts, and the rows are labeled with the individual difference measures used to predict language learning. Within the context headings, the individual columns depict the various outcome measures. In Table 6 the larger column headings show the language categories, and the rows once again display the individual difference measures. Each column within a particular language category is labeled with the outcome measures. Expected predictors of success are identified with an asterisk.

Table 5. Expected findings for predictors of success in second language learning.

	Foreign Language Classroom			Intensive Instruction			Study Abroad		
	FL GPA	DLPT-R	DLPT-L	FL GPA	DLPT-R	DLPT-L	FL GPA	DLPT-R	DLPT-L
Predictors									
SAT Math	*	*		*	*		*	*	
SAT Verbal									*
Undergrad GPA	*			*			*		
DLAB		*			*			*	

As mentioned in the literature review, the default assumption is that aptitude and other individual difference measures ought to be context independent. The expectation is that the results of this study will provide evidence to support the default assumption; however, due to the increased intensity in the listening modality in a study abroad context, the measure of verbal ability is expected to differentiate learner listening proficiency.

Table 6. The expected effect of language category on the predictive validity of ID measures.

	Language Category 1				Language Category 2				Language Category 3				Language Category 4			
	FL GPA	OPI	DLPT-R	DLPT-L	FL GPA	OPI	DLPT-R	DLPT-L	FL GPA	OPI	DLPT-R	DLPT-L	FL GPA	OPI	DLPT-R	DLPT-L
Predictors																
DLAB	*		*		*		*		*		*		*		*	
ASVAB Q (AR)	*		*		*		*		*		*		*		*	
ASVAB Q (MK)																
ASVAB V (PC)	*		*	*	*		*	*	*			*				
ASVAB V (WK)	*	*		*	*			*								

ASVAB Q (AR) = Arimetic Reasoning
ASVAB Q (MK) = Math Knowledge
ASVAB V (PC) = Paragraph Comprehension
ASVAB V (WK) = Word Knowledge

As mentioned by Child (1998) and Lowe (1998), distance from English is expected to play a role in the predictive ability of individual difference measures on second language learning for native speakers of English. These “distance” effects are primarily expected for measures of verbal ability since the measures used are specifically measures of English ability. Therefore, measures that differentiate learners in their native language, English, would also be anticipated to differentiate learners in languages that largely overlap with English. As the distance from English grows, however, the predictive ability of the measures is expected to taper off. Measures of more general cognitive abilities, on the other hand, would be expected to have similar predictive ability across language category boundaries. The measure, Arithmetic Reasoning (AR) for example, is anticipated to have predictive ability in all language categories since it is a measure of general cognitive reasoning ability. Math Knowledge (MK) is an unrelated measure of crystallized knowledge (Alderton et al., 1997); and therefore is expected to have no predictive ability concerning language learning. According to the ASVAB Career Exploration Program (2011), however, both Arithmetic Reasoning and Math Knowledge are measures of logical thinking. Alderton et al. (1997) in a factor analysis found that AR and MK both load on a math factor, but they showed that AR also loads nearly equally on a measure of non-verbal reasoning. They claim that their non-verbal reasoning factor may be an indicator of fluid intelligence. So, AR and MK are different in how they measure math abilities with AR demonstrating more of a non-verbal ability to reason. Both measures are used in this study as indicators of general cognitive ability (Surface et al. 2004). Finally, the DLAB was constructed as a measure of language learning aptitude to differentiate successful and non-successful learners at DLI. A large part of that success is determined by the learner's grade point average while studying at the Defense Language Institute. Additionally, the learner must score at the prescribed levels on the DLPTs and the OPI. The confounding factor is the time allotted for study depending on language category. Therefore further research is required to determine the ability of the DLAB to differentiate learners within category. The only expected result at this point is the DLAB's ability to differentiate learner reading proficiency levels since this has been demonstrated in the past.

Chapter 4: Current Study Overview

Since the Department of Defense is highly invested in the DLAB and is seeking the most efficient means to train military personnel, it is important to know the effectiveness of the DLAB in predicting foreign language growth, as compared to and in combination with other aptitude and achievement measures in a variety of learning contexts. As mentioned, Linck et al. (2012) suggest that the predictive ability of the DLAB may be confined to certain learning environments and proficiency levels. Robinson (2013) makes a similar suggestion. Silva & White (1993) show incremental predictive validity of the DLAB over other more general measures of aptitude. Surface et al. (2004) demonstrate that the long-range predictive effects of other aptitude measures outweigh the ability of the DLAB to predict growth in foreign language proficiency, specifically in certain language categories. Child (1998) and Lowe (1998) recognize that language categorization is based mainly on time to learn a language rather than language similarities. If predictive measures prove to be more effective for a particular language or language category, perhaps this will shed light on the particular components of language aptitude involved in learning that language or type of language. Also, if patterns are found in predictors for certain languages, then that may allow for improvements in how languages are categorized. With this information, the DOD can refine its predictive testing measures to better align with the desired language category and context for a particular individual. The goal is grandiose, but this research will help lay the foundation.

In many cases, training context will be decided by availability of training and resources. For example, a student at the Naval Academy will normally study in an academic classroom setting, or FLC as defined in this study. Since USNA students may now major in Arabic or Chinese, the Academy is seeking a selection tool to help assess the probability of success for students in these languages. To this point the Academy has allowed for self-selection into these majors, but relatively high failure rates are forcing the school to look for other measures. Additionally, USNA has some available resources to send selected students to study abroad, but because resources are limited, administrators are also seeking effective selection tools for these programs. On a larger scale, the U.S. Navy offers a variety of programs, including intensive instruction programs. As mentioned previously, the Navy has typically used the DLAB as a predictor of success and, therefore, as a selection tool, but research into the effectiveness of the DLAB as a predictor has mainly been limited to intensive instruction programs, and by default other programs use the DLAB, as well, without support from empirical studies. This study will investigate whether the DLAB is a valuable instrument for the U.S. Navy in selection of candidates to programs in all three contexts based on its ability to predict differences in foreign language proficiency outcomes (Foreign Language GPA and the DLPT). It will also explore whether the coefficients for the independent variables of predictive models of foreign language success will vary in magnitude across language learning contexts.

The outcome measures include Foreign Language GPA, the listening and reading Defense Language Proficiency Tests, and the Oral Proficiency Interview. Foreign Language GPA is simply the average grade given for all foreign language courses taken by the participant. But, more commonly, the DOD uses the Defense Language Proficiency Test (DLPT) and the Oral Proficiency Interview (OPI) to measure foreign language proficiency. The DLPT standardized testing program started in the 1950s, and is currently in its 5th generation of exams. The DLPT program provides testing in 2 areas; reading and listening. There are lower range (ILR scale 0-3) and upper range (ILR scale 4-5) exams. This analysis uses scores from the lower range exam. The lower range reading test includes 60 multiple choice questions based on 36 authentic passages with up to 4 questions on each passage (for example, see Appendix A). The lower range listening test contains 60 multiple choice questions based on 40 authentic passages with up to 2 questions on each passage. The OPI consists of 4 parts with 26 questions including answering personalized questions, narrating events, speaking on selected topics, and role play. For the analyses in this report, an equivalent ILR scale score is calculated. The equivalent score is calculated by multiplying the scale score by 10 and then adding 6 for a "+" scale score. For example, an ILR scale score of 3 is replaced with a value of 30; an ILR scale score of 2+ is replaced with a value of 26; an ILR scale score of 2 is replaced with a value of 20; and so on.

The analyses for the first study will use linear regression and examine the magnitude of the standardized β coefficients of the independent variables for the predictor models to determine the impact of learning context on the predictive validity of individual difference measures for several foreign language outcomes. The three different context groups include a FLC group, an INI group, and a SA group. Each group will have 120 participants. The first group will consist of students at the United States Naval Academy (USNA) studying a foreign language in the FLC context. The second group will also consist of USNA students, but this group will be those students who spent a semester studying abroad. They will be considered the SA group. The third group will consist of U.S. military officers who attend the Defense Language Institute (DLI). The third group will be the INI group. Linear regression will be used to determine the ability of the DLAB to predict language learning success for the FL GPA outcome measure in all three contexts.

Whether or not learning context is a factor in determining the ability of the DLAB to predict language learning success, it is also important to examine other factors that may play a role in determining language learning success in these contexts. This is important because the effects of other variables in language learning and/or their interactions with the effects of the DLAB scores may affect language learning outcome measures. Dörnyei (2005) claims that individual differences in personality, ability/aptitude, and motivation are all seen as principle learner variables from an educational perspective. These individual difference measures will include a native language verbal aptitude measure (SAT verbal/GREV), a quantitative aptitude measure (SAT math/GREQ), and a native language achievement measure (undergraduate grade point average). For this analysis, hierarchical entry of independent variables will be used. This method will be performed separately for each of the outcome measures (FL GPA, DLPT Reading, and DLPT Listening). Within group correlations will be analyzed to better understand the ability of the DLAB to predict foreign language gains in each of the three contexts. (See the Study 1 summary table, below).

Table 7. Summary of Study 1: Effect of context on predicting success in SLA.

Summary Table

Study 1: Effect of context on predicting success in SLA

Context	N-size	Predictors	Outcome 1	Outcome 2	Outcome 3
FLC	120	SAT Math			
INI	120	SAT Verbal, DLAB	FL GPA	DLPT-R	DLPT-L
SA	120	Undergrad GPA			

Analyses: -Linear Regression using FL GPA, DLPT-L, and DLPT-R as the outcome measures for each context group

-Hierarchical input of predictors (chronological order of occurrence)

-Analyze ΔR^2 to determine significance of impact of each predictor on the model

Another crucial aspect of understanding how DLAB predicts success is that of language categorization. Child (1998) suggests that differences in languages and their "distance from English" may be an important influence on how well an aptitude measure predicts proficiency gains. Child (1998) also suggests that although there have been a number of attempts to categorize languages according to their presumed difficulty, which on the surface have a certain face validity, insufficient empirical evidence exists to support the current system of language categorization. The second study, therefore, will look at 150 graduates of the United States Naval Academy; 75 graduates who majored in Chinese and 75 who majored in Arabic. The same methodology noted above will be used here. Since students entering these majors are not required to take the DLAB, this study will use SAT verbal, SAT math, and English Composition GPA as predictors, and DLPT scores and final foreign language GPAs will be used as outcome measures. (See the Study 2 summary table, below).

Table 8. Summary Table Study 2: Predicting success for category 4 language majors.

Summary Table

Study 2: Predicting success for students majoring in a category IV language

Major	N-size	Predictors	Outcome 1	Outcome 2	Outcome 3
Chinese	75	SAT Math, SAT Verbal	FL GPA	DLPT-L	DLPT-R
Arabic	75	English Comp.			

First Analysis: -Linear Regression using FL GPA as the outcome measure for each context group
-Hierarchical input of predictors (chronological order of occurrence)
-Analyze ΔR^2 to determine significance of impact of each predictor on the model

Second Analysis: -Logistic Regression using DLPT-L and DLPT-R as the outcome measures
-Outcomes scored as "1" for above average and "0" for below average
-Hierarchical input of predictors (chronological order of occurrence)
-Analyze ΔX^2 to determine significance of impact of each predictor on the model

The final study will present a more comprehensive analysis of languages within and across language categories by looking at the effects of a particular set of predictor and outcome variables for each of eight different languages. The analyses will look at two languages in each of the four language categories. Using linear and logistic regression, the proficiency growth in reading and listening for 200 DOD language specialists in each language will be analyzed to compare outcomes, as well as predictive aptitude measures. Linear regression will be used to evaluate outcomes at DLI graduation and for three additional DLPT annual assessments, while logistic regression will look at growth over time intervals. Latent growth curve models were considered, but the data are not suited for that approach. Additionally, the desire here is to analyze specific examination points along the growth curve. The data provided by DMDC will include DLAB and ASVAB scores, as well as four consecutive annual DLPT reading and listening scores for each individual. Growth over a time interval will be coded as “1,” and no growth or loss will be coded as “0”. This will allow for investigation of specific aptitude factors that predict proficiency growth over time. The analysis will provide evidence to help evaluate current language categorization, while looking at the best predictors of success for each of the various language categories. (See the Study 3 summary table, below).

Table 9. Summary Table Study 3: Language category effects.

Summary Table

Study 3: Language category effects on predicting outcomes in second language learning

Language	N-size	Lang Category	Predictors	Outcome 1	Outcome 2	Outcome 3	Outcome 4
French	200	1	ASVAB Quantitative (MK, AR), ASVAB Verbal (PC, WK), DLAB	FL GPA	DLPT-L (4 years)	DLPT-R (4 years)	Oral Proficiency Interview
Spanish	200	1					
Indonesian	200	2					
German	200	2					
Tagalog	200	3					
Russian	200	3					
Arabic	200	4					
Chinese	200	4					

Analysis 1: -Linear Regression using FL GPA, DLPT-L, DLPT-R, and OPI as the outcome measures for each lang. cat.
-Hierarchical input of predictors (MK, AR, PC, WK, DLAB)
-Analyze ΔR^2 to determine significance of impact of each predictor on the model
-Analyze standardized β coefficients for the predictors in the model

Analysis 2: -Linear Regression using FL GPA, DLPT-L, DLPT-R, and OPI as the outcome measures for each language
-Hierarchical input of predictors (MK, AR, PC, WK, DLAB)
-Analyze ΔR^2 to determine significance of impact of each predictor on the model
-Analyze standardized β coefficients for the predictors in the model

The DLAB was specifically developed to predict language learning success at the Defense Language Institute (Lett & O’Mara, 1990). As Dörnyei & Skehan (2003) point out, the DLAB uses the most predictive combinations of sub-tests to measure aptitude. Therefore, the DLAB is expected to predict proficiency levels at the completion of DLI for participants in this analysis. If language learning is more dependent on the individual cognitive ability of the learner as Long (1997) claims, and the DLAB is a measure of that cognitive ability as Dörnyei & Skehan (2003) suggests, then the DLAB should be successful in predicting proficiency levels upon completion of programs in all three contexts as well.

Additionally, many researchers claim that learners with greater abilities in their native language tend to have higher proficiency levels in their second language (Sparks, 1998; Skehan, 1991; Carroll, 1973; Sparks and Ganschow, 1991; 1993a; 1995a; Humes-Bartlo, 1989; Pimsleur, 1966a, 1968). Therefore, English composition grades should differentiate proficiency levels of foreign language learners. Cho and Bridgeman (2012) and Ayers and Quattlebaum (1992) showed some indirect evidence to support the claim that measures of verbal and quantitative ability may have some value in predicting success in a foreign language, as well. Surface et al. (2004) also showed that quantitative ability measures can predict foreign language proficiency growth, at least in some language categories. So, it is reasonable to expect some predictive influence of SAT scores and English Composition grades on proficiency outcomes.

Language category on the other hand is more of an unknown. Lett & O'Mara (1990) commented about the large variances of DLAB scores that were able to predict success within language category indicating that DLAB is not the best predictor of foreign language proficiency growth within category. Child (1998) explains that the current categorization is based on time to learn rather than similarities and differences (distance from English) of the languages. This could allow for the variation that Lett & O'Mara (1990) discovered. If languages are grouped according to the system suggested by Child (1998), aptitude measures like the DLAB may be more stable within language category. That said, the current study expects to replicate the large variances of Lett & O'Mara (1990) and show little to no success for the DLAB in differentiating learner language gains within category. The study also hopes to provide evidence for patterns among predictors that may help to shape language categorization in the future.

Chapter 5: Effect of Context on Predicting Success in SLA

5.1 Study Overview

Pimsleur (1966) developed the Pimsleur Language Aptitude Battery (PLAB) with the intent of discovering learner strengths and weaknesses and adapting teaching methods to better align with learner abilities. This idea would suggest that if teaching methods are aligned with learner strengths then learning would be increased for the individual. Stanhope & Surface (2014) suggest the importance of predictor-criterion alignment in different learning contexts and state that "it is reasonable to expect individuals with specific abilities that align with training content to have a higher likelihood of success" (Stanhope & Surface, 2014, p. 152). These claims suggest an interactive role between the learner and context that could lead to differential learning outcomes based on individual differences that align better with one context over another. If language learning truly differs based on the individual nested within a particular context, then one would expect that the ability of aptitude measures to predict learning gains would also vary between contexts. The intent of this chapter is to examine possible differential effects of context as defined in this study (FLC, INI, and SA) on the predictive ability of individual differences and test the default assumption that aptitude and other individual difference measures ought to be context independent.

5.2 Method

The analyses for this chapter use linear regression and examine the standardized β coefficients for the predictors of foreign language proficiency outcomes for three groups; a FLC group, an INI group, and a SA group. Linear regression is used to determine the ability of several individual difference measures to predict language learning success as measured by foreign language GPA, listening DLPT scores, and reading DLPT scores in the three contexts described in this study. Foreign language GPA is the grade point average attained for all of the foreign language courses taken by a participant at USNA or the Defense Language Institute. The DLPT scores used as the outcome measures in this study are equivalent DLPT scores calculated by multiplying the Interagency Language Roundtable (ILR) scale score by 10 and then adding 6 for a "+" scale score. For example, an ILR scale score of 3 is replaced with a value of 30; an ILR scale score of 2+ is replaced with a value of 26; an ILR scale score of 2 is replaced with a value of 20; and so on.

The individual difference measures used in this study are possible predictors of language learning success as indicated by the literature on the subject as described in the literature review in Chapter 1. The measures include a quantitative aptitude measure (SAT math/GREQ) as a proxy for general cognitive aptitude, a native language verbal aptitude measure (SAT verbal/GREV), a foreign language aptitude measure (DLAB), and a native language achievement measure (undergraduate grade point average). For the analyses, hierarchical entry into the models is used. Entry of predictors into the models is according to the chronological progression of occurrence for the participants. Where measures were taken simultaneously, the general cognitive measure is entered first.

The predictors were hypothesized to have varying degrees of impact on the different outcomes being assessed as discussed in Chapter 1. The default assumption is that aptitude and other individual difference measures ought to be context independent. The expectation was that the results of this study would provide evidence to support the default assumption. Additionally, due to the increased intensity in the listening modality in a study abroad context, the measure of verbal ability was expected to differentiate learner listening proficiency.

Based on the hierarchical entry procedure described, SATM/GREQ was entered first as a general cognitive measure and was expected to have a significant impact on the predictive models. General cognitive measures like SATM/GREQ have shown some correlation with L2 proficiency measures (Surface et al., 2004; Wagener, 2014). SATV/GREV was entered into the model following SATM/GREQ scores. Unlike SATM/GREQ, SATV/GREV was only expected to have a significant impact on L2 listening proficiency for the study abroad students since researchers have shown at higher proficiency levels there are correlations between native language verbal measures and foreign language proficiency (Vandergrift, 2006; Carson et al., 1990). The DLAB is entered third and was expected to have a significant impact on the models for all contexts based on its design to measure foreign language aptitude. Finally, undergraduate grade point average is entered into the model as an indicator of skill in the native language. Undergraduate GPA also measures some of the other intangibles of success that Dörnyei (2005) suggests may have an impact on language learning. As a measure of native language achievement, GPA was expected to have a significant impact on predictive models of foreign language success.

5.3 Participants and Data

The participants are midshipmen and young military officers who studied a foreign language. The first group consists of 111 midshipmen at the United States Naval Academy (USNA) studying a foreign language in the FLC context. The second group consists of 57 students who also attended USNA, but this group has spent a semester studying abroad in a country where the studied language is spoken. They are considered the SA group, and are different participants than those in the FLC group. The third group consists of 147 U.S. military officers who attended the Defense Language Institute (DLI). The third group will be the INI group.

The data for the midshipmen was provided by the Office of Academic Affairs at the United States Naval Academy. It was taken from admissions' and student academic data records. The data for the INI group was provided by the Defense Manpower and Data Center (DMDC). It was taken from DLI records and applicant data for military foreign language study programs.

The measures provided by USNA and DMDC include undergraduate GPA (CQPR), SAT math/GREQ, SAT verbal/GREV, and DLAB scores. The range of undergraduate grade point average is from a minimum of 0.0 to a maximum of 4.0. The SAT/GRE verbal and math scores can range from 200 to 800. The DLAB is scored on a 176 point scale. Foreign language GPA and DLPT scores were also provided where available. Students majoring in a foreign language at USNA complete 10 to 14 language courses in their chosen major language. The foreign language grade point average reported here is the average grade for these major language courses for each student. For the DLI students, the grade point average at DLI is considered the foreign language (FL) GPA. The listening and reading DLPT scores (discussed in detail in previous chapters) used in these analyses are in the respective foreign language for each participant. The descriptive statistics for the measures by group are reported in the following section.

5.4 Analyses

Foreign Language GPA

Linear regression with hierarchical entry of predictor variables is used in this analysis. This analysis uses the foreign language GPA of the participants as the outcome measure. The significance of each predictor is determined when it is added to the model by looking at the change in the R^2 term. The descriptive statistics for the predictor variables and the outcome measures by group is shown in Tables 10, 11, and 12 below.

Table 10. Descriptive statistics for the FLC group predictor and outcome variables.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
SATM	111	490	800	694.32	72.483
SATV	111	480	800	668.38	67.938
DLAB	111	58	150	119.41	16.769
CQPR	111	2.32	4.00	3.4738	.39332
FL_GPA	101	1.50	4.00	3.6370	.51128
DLPT_List	47	0	30	18.68	9.215
DLPT_Read	47	6	30	19.79	7.715
Valid N (listwise)	44				

Table 11. Descriptive statistics for the INI group predictor and outcome variables.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
GREQ	147	490	800	702.31	72.396
GREV	147	450	800	613.13	68.465
DLAB	146	95	154	127.85	12.167
CQPR	146	2.58	4.00	3.4738	.30153
FL_GPA	85	2.80	4.00	3.7671	.27141
DLPT_List	144	6	30	23.28	6.079
DLPT_Read	145	0	30	26.22	5.427
Valid N (listwise)	83				

Table 12. Descriptive statistics for the SA group predictor and outcome variables.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
SATV	57	480	800	676.84	69.312
SATM	57	510	800	682.81	62.041
DLAB	57	96	147	122.74	8.948
CQPR	57	2.34	4.00	3.4360	.30866
FL_GPA	55	3.08	4.00	3.7973	.24849
DLPT_List	57	0	30	15.02	10.283
DLPT_Read	57	0	30	16.53	9.566
Valid N (listwise)	55				

Beginning with the model for the foreign language classroom (FLC) group, verbal scores, the DLAB, and undergraduate GPA (CQPR) are all significant predictors of foreign language grade point average (see Table 13, below). The model with all four predictors is highly significant ($F = 22.225$, $P < 0.001$) and explains 48.1% of the variance. The standardized β coefficients for the four predictors are -0.182, 0.006, 0.376 and 0.539 for SATM, SATV, DLAB and CQPR, respectively.

Table 13. Foreign Language GPA Outcome Model for the Foreign Language Classroom Group.

Foreign Language GPA for FLC Group

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	SATM	0.019	1.885	0.173
2	SATM, SATV	0.055	3.733	0.027
3	SATM, SATV, DLAB	0.250	12.740	0.000
4	SATM, SATV, DLAB, CQPR	0.481	14.391	0.000

For the INI group, quantitative and verbal GRE scores are significant predictors of foreign language grade point average (see Table 14, below). The model with all four predictors is highly significant ($F = 6.174$, $P < 0.001$) and explains 23.8% of the variance. The standardized β coefficients for the four predictors are 0.364, 0.197, -0.003 and 0.089 for GREQ, GREV, DLAB and CQPR, respectively.

Table 14. Foreign Language GPA Outcome Model for the Intensive Instruction Group.

Foreign Language GPA for INI Group

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	GREQ	0.189	19.374	0.000
2	GREQ, GREV	0.232	4.591	0.013
3	GREQ, GREV DLAB	0.232	0.000	1.000
4	GREQ, GREV DLAB, CQPR	0.238	0.213	0.931

Finally, for the SA group, only undergraduate GPA proves to be significant predictor of foreign language grade point average (see Table 15, below). The model with all four predictors is highly significant (F = 4.841, P = 0.002) and explains 27.9% of the variance. The standardized β coefficients for the four predictors are -0.259, -0.056, 0.045 and 0.556 for SATM, SATV, DLAB and CQPR, respectively.

Table 15. Foreign Language GPA Outcome Model for the Study Abroad Group.

Foreign Language GPA for SA Group

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	SATM	0.008	0.447	0.507
2	SATM, SATV	0.042	1.846	0.168
3	SATM, SATV, DLAB	0.056	0.386	0.764
4	SATM, SATV, DLAB, CQPR	0.279	5.258	0.001

Listening DLPT

Switching to an arguably better indicator of foreign language proficiency, each of these groups will be examined using listening defense language proficiency test (DLPT) scores as the outcome measure. As in the previous section, linear regression with hierarchical entry of independent variables is used to determine the predictive ability of certain individual difference measures. For the FLC group, none of the predictors reaches significance. The model (see Table 16, below) with all four predictors does not reach significance either, and it only explains 9.8% of the variance (F = 1.142, P = 0.350). The standardized β coefficients for the four predictors are -0.177, 0.132, 0.216 and 0.118 for SATM, SATV, DLAB and CQPR, respectively. In a post hoc analysis, when DLAB is input as the only predictor in the model, it reaches marginal significance (F = 3.085, P = 0.086). Additionally, post hoc, foreign language GPA was substituted for undergraduate GPA. This resulted in a minor improvement in the explained variance (10.8% for the model), but the model still did not reach significance (F = 1.183, P = 0.333).

Table 16. Listening DLPT Outcome Model for the FLC Group.

Listening DLPT for FLC Group

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	SATM	0.000	0.000	0.990
2	SATM, SATV	0.044	2.025	0.144
3	SATM, SATV, DLAB	0.087	1.036	0.386
4	SATM, SATV, DLAB, CQPR	0.098	0.175	0.950

The predictors in the INI group model also fail to reach significance. In fact, the model (see Table 17, below) with all four predictors explains only 1.3% of the variance (F = 0.396, P = 0.811). GREV has the largest impact on the model explaining 1.2% of that variance. The standardized β coefficients for the four predictors are -0.053, 0.095, -0.078 and 0.023 for GREQ, GREV, DLAB and CQPR, respectively. In post hoc analysis, undergraduate GPA was again replaced by foreign language GPA which resulted in a large increase in the explained variance (13.8% for the model). This also resulted in a highly significant model (F = 5.678, P < 0.001).

Table 17. Listening DLPT Outcome Model for the INI Group.

Listening DLPT for INI Group

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	GREQ	0.001	0.174	0.677
2	GREQ, GREV	0.013	1.714	0.184
3	GREQ, GREV, DLAB	0.013	0.000	1.000
4	GREQ, GREV, DLAB, CQPR	0.013	0.000	1.000

Similarly, the model for the study abroad group does not present any significant predictors either (see Table 18, below). In the case of the Study Abroad group, however, undergraduate grade point average has the largest impact on the model explaining 4.9% of the variance. With all four predictors in the model, the model fails to reach significance (F = 1.069, P = 0.381). The standardized β coefficients for the predictors are 0.076, -0.138, 0.047 and 0.263 for SATM, SATV, DLAB and CQPR, respectively. In post hoc analysis, when undergraduate GPA is entered as the only predictor, the model reaches marginal significance (F = 3.456, P = 0.068). Additionally, as done with the other two groups, undergraduate GPA was replaced by foreign language GPA. This resulted in an increase in the explained variance (8.2% for the model), but the model still did not reach significance (F= 1.164, P = 0.338).

Table 18. Listening DLPT Outcome Model for the SA Group.

Listening DLPT for SA Group

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	SATM	0.019	1.045	0.311
2	SATM, SATV	0.019	0.000	1.000
3	SATM, SATV, DLAB	0.027	0.222	0.881
4	SATM, SATV, DLAB, CQPR	0.076	0.937	0.450

Reading DLPT

This section uses linear regression with hierarchical entry of independent variables to examine the predictive ability of the same individual difference measures used above. For the FLC group, none of the predictors reaches significance. The model (see Table 19, below) with all four predictors does not reach significance either, and it only explains 5.4% of the variance (F = 0.601, P = 0.664). The standardized β coefficients for the four predictors are 0.039, 0.142, 0.025 and 0.098 for SATM, SATV, DLAB and CQPR, respectively. In post hoc analysis, undergraduate GPA was replaced by foreign language GPA, and the variance explained is increased to 9.7%. The post hoc model is not significant (F = 1.051, P = 0.394).

Table 19. Reading DLPT Outcome Model for the FLC Group.

Reading DLPT for FLC Group

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	SATM	0.022	1.017	0.319
2	SATM, SATV	0.046	1.107	0.340
3	SATM, SATV, DLAB	0.047	0.023	0.995
4	SATM, SATV, DLAB, CQPR	0.054	0.106	0.980

The predictors in the INI group model also fail to reach significance. The model (see Table 20, below) with all four predictors explains only 2.4% of the variance (F = 0.396, P = 0.811). GREV has the largest impact on the model explaining 1.2% of that variance. The standardized β coefficients for the four predictors are -0.053, 0.095, -0.078 and 0.023 for GREQ, GREV, DLAB and CQPR, respectively. In post hoc analysis replacing undergraduate GPA with foreign language

GPA increases the explained variance to 5.0%. The model still does not reach significance (F = 1.864, P = 0.120), but a model with FL GPA as the only predictor is significant (F = 4.074, P = 0.045).

Table 20. Reading DLPT Outcome Model for the INI Group.

Reading DLPT for INI Group

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	GREQ	0.000	0.010	0.922
2	GREQ, GREV	0.000	0.000	1.000
3	GREQ, GREV, DLAB	0.018	1.301	0.277
4	GREQ, GREV, DLAB, CQPR	0.024	0.289	0.885

Once again, the model for the study abroad group does not present any significant predictors (see Table 21, below). As in the listening DLPT section for the Study Abroad group, undergraduate grade point average has the largest impact on the model explaining an additional 8.8% of the variance. With all four predictors in the model, the model fails to reach significance (F = 1.720, P = 0.160). The standardized β coefficients for the predictors are 0.088, -0.143, -0.031 and 0.351 for SATM, SATV, DLAB and CQPR, respectively. In post hoc analysis, when undergraduate GPA is entered as the only predictor, the model reaches marginal significance (F = 3.456, P = 0.068). Additionally, as done with the other two groups, undergraduate GPA was replaced by foreign language GPA. This resulted in a small increase in the explained variance (11.9% for the model), but the model still did not reach significance (F= 1.749, P = 0.153). With FL GPA entered as the only predictor, the model is significant (F = 4.564, P = 0.037) and explains 7.7% of the variance.

Table 21. Reading DLPT Outcome Model for the SA Group.

Reading DLPT for SA Group

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	SATM	0.028	1.570	0.216
2	SATM, SATV	0.028	0.000	1.000
3	SATM, SATV, DLAB	0.029	0.028	0.994
4	SATM, SATV, DLAB, CQPR	0.117	1.761	0.151

5.5 Summary of Results and Discussion

The method used in this study was linear regression with hierarchical entry of predictors. Brunner et al. (2009) say this method is useful even when independent variables are measured with error if the primary interest is to build a regression model for the purposes of prediction. They do caution, however, that this type of analysis has the potential of finding spurious significant coefficients due to measurement error. In future studies, the Bonferroni correction ($p < .05/k$, where k tests are conducted) could be used as an adjustment (Fullmann, 2005). Additionally, examining the standardized β coefficients across models allows for a better understanding of possible predictor model differences. Also, the order of entry of the predictors could change the coefficients. Once again, the main intent here was to compare across learning contexts, so the assumption would be that if independent variables are entered in the same order for each of the models, the primary investigation would not be effected.

In examining the models where the outcome measure is foreign language GPA, striking differences are readily apparent between the groups. First, the polarity of the standardized β coefficients indicates that high math scores tend to be detrimental to achieving a higher GPA in foreign language at USNA. The opposite is true for the graduates of the Defense Language Institute. This would seem to indicate differences in the language programs or language learning focus of the students between the two institutions. Second, the differences between the domestic programs and the study abroad group in predictor significance, magnitude of standardized β coefficients, and polarity of the coefficients for the L1 verbal aptitude measure may indicate a heavier reliance on L1 abilities in the domestic environments. Third, with the exception of the overlap in measures between undergraduate GPA and foreign language GPA for the USNA students, the predictive abilities of the independent variables are negated by study abroad. As suggested in the literature, this indicates that learning a second language in a foreign country is very different than learning a second language in a classroom. But, it also provides some evidence that aptitude and other individual difference measures may not be context independent. Finally, the fact that the DLAB is only significant in the FLC context may indicate that intensive programs diminish its predictive validity as suggested by Linck et al. (2012) when they state that measures like the DLAB may only be useful in distinguishing rates of learning at lower proficiency levels in a classroom environment.

Next, although both the listening DLPT section and the reading DLPT section show limited predictive ability of the measures, an examination of the models provides some interesting findings. Beginning with the listening models, DLAB scores have their largest impact on the FLC group and little to no impact on the other two groups. Again, this is in line with the claims of Linck et al. (2012). Additionally, in post hoc analyses, foreign language GPA has a minor impact on the two USNA groups, but has a highly significant impact on the DLI group. This may indicate that the instructional methods at DLI have a greater focus on development of listening comprehension skills than at USNA. Finally, based on the standardized β coefficients, the listening models also demonstrate that L1 verbal skills positively impact the domestic (classroom) programs, but not study abroad. This may indicate that explicit instruction assists in L1 transfer since it is a greater benefit to students with greater L1 verbal abilities.

For the reading DLPT section, finding discernable evidence is a little more difficult. For the study abroad group and the intensive instruction group in post hoc analyses, FL GPA is a significant predictor of the DLPT scores. FL GPA also has a noticeable impact on the model for the foreign language classroom group although it does not reach significance as a predictor of reading DLPT. This may indicate that the programs at both USNA and DLI positively impact learner reading proficiency, but that intensity and/or time on task are needed to distinguish the effects of differing learner aptitudes on reading proficiency development. Another point to note for the reading DLPT models is that once again L1 verbal abilities positively impact only the domestic (classroom) programs.

In summary, this chapter provides evidence to support claims that suggest an interactive role between the learner and context leading to differential learning outcomes based on individual differences. This is in line with the research done by Pimsleur (1966) in developing the Pimsleur Language Aptitude Battery (PLAB). It also supports the suggestion of Stanhope & Surface (2014) that "it is reasonable to expect individuals with specific abilities that align with training content to have a higher likelihood of success" (Stanhope & Surface, 2014, p. 152). This is demonstrated by L1 verbal abilities having a greater impact on learner proficiency growth in a setting that allows for L1 use for explicit instruction in foreign language learning. This study only included the overall DLAB scores, but if Individual Difference measures are able to differentially predict outcomes, it would follow that the individual components of the DLAB may also differentially predict success based on context. Finally, this chapter examined possible differential effects of context as defined in this study (FLC, INI, and SA) on the predictive ability of individual differences and provided evidence that overtly challenges the default assumption that aptitude and other individual difference measures ought to be context independent.

Chapter 6: Predicting Success for Category IV Language Majors

6.1 Study Overview

Lett & O'Mara (1990) comment that within language categories there is a large degree of variation in the DLAB scores of successful learners. This in combination with the claims of Child (1998) that "distance" from English plays a vital role in the success of a learners would appear to suggest that there may be some variation in the patterns of aptitude measures that predict language learning. The following analysis is intended to take a closer look at two category IV languages and individual difference measures that may predict success in the learning of those languages in order to evaluate if evidence exists to support the claims of Child (1998) and the findings of Lett & O'Mara (1990).

6.2 Method

These analyses use SPSS version 22 to perform linear and logistic regression to look at predictors of foreign language proficiency for two groups of individuals who studied category IV languages in an undergraduate program at a military institution. One group majored in Chinese and the other in Arabic. The analyses use scores upon graduation from the military institution and are conducted for three outcome measures: foreign language grade point average (FL_GPA), listening DLPT scores

(DLPT_List), and reading DLPT scores (DLPT_Read). A linear regression model was used for the foreign language grade point average outcome while a logistic regression model was used for each of the DLPT outcomes. For the logistic regression models, scores that were above average were assigned a value of "1," and below average scores were assigned a value of "0." SATM, SATV, and English Composition grades were the predictor variables for each of the models. Hierarchical entry into the models was used. Entry of predictors followed chronological progression. The predictors were hypothesized to have varying degrees of impact on the different outcomes being assessed. SATM was entered first as a general cognitive measure and was expected to have a significant impact on the predictive models, since general cognitive measures like SATM have shown some correlation with L2 proficiency measures in the past (Surface et al., 2004; Wagener, 2014) particularly for category IV languages. SATV was entered into the model following SATM scores. Unlike SATM, SATV was not expected to have a significant impact on L2 proficiency outcomes since the evidence that researchers have found to support correlations between native language verbal measures and foreign language proficiency in language learning have been at higher L2 proficiency levels (Vandergrift, 2006; Carson et al., 1990). The learners in this study are mainly low proficiency learners. Finally, English Composition grades (ENG_Comp) were entered into the model. As a measure of native language achievement in writing, English Composition grades were expected to have a significant impact on predictive models of foreign language success, particularly on the reading DLPT model due to similarities in modality. Although Carson et al. (1990) showed that reading and writing skills in the L1 correlated with reading and writing skills in the L2 for higher proficiency learners, the overlap in the type of metric and predictor-criterion alignment for the skill sets involved would suggest this correlation may also be true of lower proficiency learners as well.

6.3 Participants and Data

Scores were taken from 153 graduates of the United States Naval Academy between the years of 2010 and 2015. 77 of the graduates majored in Arabic, and 76 majored in Chinese. Graduates are between the ages of 21 and 26. They are U.S. citizens and have completed all requirements of the 4-year institution.

The data collected includes the language studied, cumulative grade point average (CQPR), English composition grades (ENG_Comp), foreign language grade point average for the four years of study (FL_GPA), SAT verbal and math scores, and DLPT Listening and Reading scores. The descriptive statistics for the participants are summarized in Table 22, below.

Table 22. Descriptive Statistics for Category IV Language Majors.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
CQPR	153	1.94	4.00	3.2095	.45494
Eng_Comp	153	2.00	4.00	3.3824	.54036
SATV	153	460	800	664.97	74.287
SATM	153	480	800	667.91	68.197
FL_GPA	153	2.00	4.00	3.6205	.41593
DLPT_List	82	0	30	12.24	8.757
DLPT_Read	82	0	30	14.07	8.228
Valid N (listwise)	82				

The range of cumulative grade point average is from a minimum of 0.0 to a maximum of 4.0, although a minimum cumulative grade point average of 2.0 is required for graduation unless a special waiver is granted. A minimum of two courses of English Composition is required, and the range of possible grades is from 0.0 to 4.0. The English Composition scores in these analyses are the average English Composition grades for all English Composition courses taken by each student. The SAT verbal and math scores are the scores that were reported to USNA by each student for his/her initial application for admission. Students majoring in a foreign language complete 10 to 14 language courses in their chosen major language. The foreign language grade point average reported here is the average grade for these major language courses for each student. The listening and reading DLPT scores used in these analyses are also in the respective language, and they are taken voluntarily by each student. The descriptive statistics for the measures by language group are reported in the following sections.

6.4 Analyses

Foreign Language GPA

Linear regression with hierarchical entry of predictor variables was used for this analysis. This analysis used the foreign language GPA of the participants as the outcome measure. The significance of each predictor was determined when it was added to the model by looking at the change in the R^2 term. The descriptive statistics for the predictor variables and the outcome measure by language major is shown in table 23, below.

Table 23. Descriptive statistics by language for predictors and FL_GPA outcome.

		Eng_Comp		SATV		SATM		FL_GPA	
	N	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Arabic	77	3.38	0.53	667.4	71	666.4	67	3.56	0.42
Chinese	76	3.38	0.55	662.5	78	669.5	69.8	3.68	0.42

There are no significant differences between the means for any of these variables by language group (P > 0.05).

Table 24. Foreign Language GPA outcome models for Arabic Majors.

Foreign Language GPA for Arabic Majors

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	SATM	0.000	0.016	0.900
2	SATM, SATV	0.026	2.002	0.142
3	SATM, SATV, Eng_Comp	0.162	6.086	0.001

The models for the Arabic majors are summarized in Table 24, above. Each consecutive model adds an additional predictor using the hierarchical order described earlier. A significant change in R^2 happens when Eng_Comp is added to the model (F = 6.086, P = 0.001). The model with all three predictors is also highly significant (F = 4.713, P = 0.005). The standardized β coefficients for the three predictors in Model 3 are -0.104, 0.054, and 0.391 for SATM, SATV, and Eng_Comp, respectively. A post hoc stepwise linear regression model yielded a highly significant model with Eng_Comp as the only predictor ($F = 13.713$, $P < 0.001$) in the model explaining 15.5% of the variance. The standardized β coefficient for Eng_Comp in the model was 0.393. Table 25 shows the correlations between the independent

Table 25. Independent Variable Correlations for Arabic Models.

Correlations

Control Variables			SATM	SATV	Eng_Comp
FL_GPA	SATM	Correlation	1.000	.545	.174
		Significance (2-tailed)	.	.000	.133
		df	0	74	74
	SATV	Correlation	.545	1.000	.311
		Significance (2-tailed)	.000	.	.006
		df	74	0	74
	Eng_Comp	Correlation	.174	.311	1.000
		Significance (2-tailed)	.133	.006	.
		df	74	74	0

variables in the preceding models. There is a moderate correlation between SATM and SATV ($r = 0.545$).

Table 26. Foreign Language GPA outcome models for Chinese Majors.

Foreign Language GPA for Chinese Majors

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	SATM	0.106	8.785	0.004
2	SATM, SATV	0.122	1.3667	0.1959
3	SATM, SATV, Eng_Comp	0.186	2.948	0.0385

The models for the Chinese majors are summarized in Table 26, above. The same procedure and hierarchical order were used. A highly significant change in R^2 occurs when SATM is added as the first predictor to the model. ($F = 8.785$, $P =$

0.004). Another significant change in R^2 happens when Eng_Comp is added to the model (F = 2.948, P = 0.039). All three models are highly significant (P < 0.01). The standardized β coefficients for the three predictors in Model 3 are 0.237, 0.066, and 0.268 for SATM, SATV, and Eng_Comp, respectively. A post hoc stepwise linear regression model yielded two highly significant models; one with Eng_Comp as the only predictor (F = 9.49, P = 0.003), explaining 11.4% of the variance, and the other with SATM and Eng_Comp as the predictors (F = 8.197, P = 0.001), explaining 18.3% of the variance. The standardized β coefficients for SATM and Eng_Comp in the second of these models were 0.269 and 0.284, respectively. Table 27, below, shows the correlations between the independent variables in the preceding Chinese models. There is a moderate correlation between SATM and SATV (r = 0.497). Since SATM is a highly significant predictor of foreign language GPA for learners of Chinese, the impact of SATV on foreign language GPA may be diminished due to the correlation. A post hoc linear regression analysis shows that a model with only SATV as a predictor would be significant (F = 6.415, P = 0.013).

Table 27. Independent Variable Correlations for Chinese Models

Correlations

Control Variables			SATM	SATV	Eng_Comp
FL_GPA	SATM	Correlation	1.000	.497	.100
		Significance (2-tailed)	.	.000	.394
		df	0	73	73
	SATV	Correlation	.497	1.000	.259
		Significance (2-tailed)	.000	.	.025
		df	73	0	73
	Eng_Comp	Correlation	.100	.259	1.000
		Significance (2-tailed)	.394	.025	.
		df	73	73	0

Finally, the same procedure and hierarchical order were used for both majors together. The results are displayed in Table 28, below. A marginally significant

Table 28. Foreign Language GPA outcome models for both Majors.

Foreign Language GPA for Both Majors				
Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	SATM	0.025	3.799	0.053
2	SATM, SATV	0.043	1.411	0.247
3	SATM, SATV, Eng_Comp	0.141	4.278	0.006

change in R^2 occurs when SATM is added as the first predictor to the model. (F = 3.799, P = 0.053). A highly significant change in R^2 happens when Eng_Comp is added to the model (F = 4.278, P = 0.006). Model 2 is significant (F = 3.382, P = 0.037). Model 3 is highly significant (P < 0.001). The standardized β coefficients for the three predictors in Model 3 are 0.070, 0.052, and 0.331 for SATM, SATV, and Eng_Comp, respectively. A post hoc stepwise linear regression model yielded a highly significant model with Eng_Comp as the only predictor (F = 22.567, P < 0.001), explaining 13.0% of the variance. The standardized β coefficient for Eng_Comp in the model was 0.361. Table 29 shows the independent variable correlations for these models. As with the models for the individual majors, there is a moderate correlation between SATM and SATV (r = 0.525). A post hoc linear analysis shows that a model containing just SATV as a predictor would be significant (F = 6.251, P = 0.013). This indicates a large overlap in the variance explained by SATM and SATV.

Table 29. Independent Variable Correlations for both Majors.

Correlations

Control Variables			SATM	SATV	Eng_Comp
FL_GPA	SATM	Correlation	1.000	.525	.131
		Significance (2-tailed)	.	.000	.108
		df	0	150	150
	SATV	Correlation	.525	1.000	.283
		Significance (2-tailed)	.000	.	.000
		df	150	0	150
	Eng_Comp	Correlation	.131	.283	1.000
		Significance (2-tailed)	.108	.000	.
		df	150	150	0

This analysis yielded the expected overall results for a foreign language classroom with SAT Math and English Composition grades having some success in predicting foreign language grade point average for lower proficiency learners. A more in depth analysis shows that SAT Math scores may be masking some of the predictive ability of SAT Verbal scores for category IV language majors. Interestingly, however, the predictive patterns for this set of independent variables differ between the two languages as evidenced by the standardized β coefficients when all three predictors are in the models. For Chinese learners, the standardized β coefficients for SATM and Eng_Comp are of similar magnitude while the coefficient for SATV is 25% of that magnitude. For Arabic learners, the standardized β coefficient for Eng_Comp far exceeds that of the other variables. The magnitude of the standardized β coefficient for SATV is similar for the two language groups, but the coefficient for SATM is actually negative for Arabic learners. Since the correlation matrices are of similar magnitude, the differences in the predictive patterns cannot be explained by differing patterns of scores on the independent variables by learners of the two languages. This may indicate that there is some evidence to support the claims of Child (1998) that languages within the same category have fundamental differences in how they are learned by native speakers of English. This is further investigated in Chapter 7.

Listening DLPT

Logistic regression with hierarchical entry of predictor variables was used for this analysis. Foreign language GPA (FL_GPA) was added as an independent variable in this set of models since it follows Eng_Comp chronologically, and logically it should predict foreign language proficiency after a course of study in that language. To find the outcome measure, an average listening DLPT score was calculated for the two participant groups. Then, each participant's listening DLPT score was compared against the average. If the participant's score was above the average, a value of "1" was assigned as that individual's score on the outcome measure. If the participant's score was below the average, a value of "0" was assigned. The significance of each predictor was determined when it was added to the model by looking at the change in the χ^2 term. Only about 50% of the students took the DLPT since testing is optional. Students who did not take the DLPT were excluded from this analysis. The descriptive statistics for the predictor variables and listening DLPT scores by language major are shown in table 30, below.

Table 30. Descriptive Statistics for Predictor Variables and Listening DLPT scores for both Majors.

		Eng_Comp		SATV		SATM		FL_GPA		DLPT_List	
	N	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Arabic	39	3.46	0.51	678.0	72.0	669.0	71.0	3.68	0.54	8.97	9.00
Chinese	42	3.50	0.49	680.5	75.3	684.4	62.4	3.86	0.24	15.21	7.46

An independent samples t-test was run for the predictors and listening DLPT scores. Of the variables, the between group means differed significantly for both FL_GPA (P = 0.002) and DLPT_List (P = 0.001). Since the current analysis is looking at within group predictive patterns for above average listening DLPT scores, these differences are noteworthy, but should not impact the overall findings. Of note, however, the FL_GPA scores for the Chinese majors are nearer the maximum score of 4.0. This could lead to a ceiling effect for that predictor in the Chinese group much sooner than in the Arabic group.

Table 31. Listening DLPT outcome models for Arabic Majors.

Listening DLPT Models for Arabic Majors

Model	Predictors	Model Cox & Snell R^2	Model $\chi 2$	$\Delta\chi 2$	P of $\Delta\chi 2$
1	SATM	0.001	0.026	0.026	0.871
2	SATM, SATV	0.009	0.366	0.339	0.560
3	SATM, SATV, Eng_Comp	0.013	0.523	0.158	0.691
4	SATM, SATV, Eng_Comp, FL_GPA	0.029	1.132	0.609	0.435

The models for the Arabic majors are summarized in Table 31, above. Each consecutive model adds an additional predictor using the hierarchical order described earlier. There are no significant predictors of above average listening DLPT scores in this set. The exponent (β) terms for the four predictors in Model 4 are 1.001, 0.995, 1.156, and 2.409 for SATM, SATV, Eng_Comp, and FL_GPA, respectively. The exponential function of the β coefficient, once again, is the odds ratio associated with a one unit increase in the predictor variable when the predictor variable is on an interval scale. Therefore, when exponent (β) is greater than one, the odds of the independent variable increase the likelihood of the student getting an above average DLPT score. If exponent (β) equals one, then the independent variable has no effect on the outcome, and less than one indicates a constraint on the student achieving an above average score. That said, FL_GPA appears to have the largest impact on listening comprehension scores, although that effect would only be seen after a 0.4 or greater increase in the GPA.

Table 32. Independent Variable Correlation for Arabic Majors.

Correlations

Control Variables			SATM	SATV	Eng_Comp	FL_GPA
DLPT_List	SATM	Correlation	1.000	.414	-.001	.063
		Significance (2-tailed)	.	.010	.995	.709
		df	0	36	36	36
	SATV	Correlation	.414	1.000	.447	.381
		Significance (2-tailed)	.010	.	.005	.018
		df	36	0	36	36
	Eng_Comp	Correlation	-.001	.447	1.000	.397
		Significance (2-tailed)	.995	.005	.	.014
		df	36	36	0	36
	FL_GPA	Correlation	.063	.381	.397	1.000
		Significance (2-tailed)	.709	.018	.014	.
		df	36	36	36	0

Table 32 shows moderate correlations between SATV and all of the other predictors. There is also a moderate correlation between FL_GPA and Eng_Comp. This may reduce the predictive power of the independent variables. Also, there is a difference in the pattern of correlations for the two language groups. This will be discussed after looking at the predictive models for the Chinese majors.

Table 33. Listening DLPT outcome models for Chinese Majors.

Listening DLPT Models for Chinese Majors

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	SATM	0.015	0.636	0.636	0.425
2	SATM, SATV	0.026	1.138	0.502	0.479
3	SATM, SATV, Eng_Comp	0.026	1.141	0.003	0.956
4	SATM, SATV, Eng_Comp, FL_GPA	0.064	2.833	1.692	0.193

The models for the Chinese majors are summarized in Table 33, above. Each consecutive model adds an additional predictor using the hierarchical order described earlier. There are no significant predictors of above average listening DLPT scores in this set. The exponent (β) terms for the four predictors in Model 4 are 0.996, 1.005, 0.540, and 23.712 for SATM, SATV, Eng_Comp, and FL_GPA, respectively. Once again, FL_GPA appears to have the largest impact on listening comprehension scores. In the case of the Chinese group, however, that effect would be seen after a 0.04 or greater increase in the GPA. This is a ten-fold change in the magnitude seen for the Arabic group. As mentioned earlier, since the average foreign language GPA for the Chinese group is 3.86, it appears that the effectiveness of this predictor reaches ceiling prior to becoming statistically significant for the current N size.

Table 34. Independent Variable Correlation for Chinese Majors.

Correlations

Control Variables			SATM	SATV	Eng_Comp	FL_GPA
DLPT_List	SATM	Correlation	1.000	.360	.155	-.096
		Significance (2-tailed)	.	.019	.326	.545
		df	0	40	40	40
	SATV	Correlation	.360	1.000	.320	-.054
		Significance (2-tailed)	.019	.	.039	.732
		df	40	0	40	40
	Eng_Comp	Correlation	.155	.320	1.000	.504
		Significance (2-tailed)	.326	.039	.	.001
		df	40	40	0	40
	FL_GPA	Correlation	-.096	-.054	.504	1.000
		Significance (2-tailed)	.545	.732	.001	.
		df	40	40	40	0

Table 34 shows the independent variable correlations for the predictors in the Chinese group. Unlike the Arabic group where moderate correlations are seen between SATV and all of the other predictors, FL_GPA is uncorrelated with SATV scores for the Chinese group. In both language groups, however, there is still a moderate correlation between FL_GPA and Eng_Comp. Also, in both language groups, FL_GPA is uncorrelated with SATM scores.

Below, the two groups are combined and analyzed using the same procedure as above. The resultant models are displayed in Table 35. FL_GPA is a marginally significant predictor of above average listening DLPT scores, but there are no significant predictors in this set. The exponent (β) terms for the four predictors in Model 4 are 0.999, 0.999, 0.777, and 5.449 for SATM, SATV, Eng_Comp, and FL_GPA, respectively. The trend would suggest that FL_GPA would be significant for a larger N-size. In any case, as evidenced by their average listening DLPT scores, these are low proficiency learners. That, in combination with the limited granularity of the measure, may make it difficult to discern differences in listening performance.

Table 35. Listening DLPT outcome models for Combined Groups.

Listening DLPT Models for Both Majors

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	SATM	0.001	0.121	0.121	0.728
2	SATM, SATV	0.002	0.123	0.003	0.956
3	SATM, SATV, Eng_Comp	0.002	0.148	0.024	0.877
4	SATM, SATV, Eng_Comp, FL_GPA	0.042	3.533	3.385	0.066

For the combined groups, the independent variable correlations are minimal (Table 36). There is a small correlation between SATV and SATM as well as a small correlation between SATV and Eng_Comp. Additionally, there is a small correlation between Eng_Comp and FL_GPA.

Table 36. Independent Variable Correlations for the Combined Groups.

Correlations

Control Variables			SATM	SATV	Eng_Comp	FL_GPA
DLPT_List	SATM	Correlation	1.000	.385	.095	.051
		Significance (2-tailed)	.	.000	.398	.654
		df	0	79	79	79
	SATV	Correlation	.385	1.000	.382	.206
		Significance (2-tailed)	.000	.	.000	.065
		df	79	0	79	79
	Eng_Comp	Correlation	.095	.382	1.000	.393
		Significance (2-tailed)	.398	.000	.	.000
		df	79	79	0	79
	FL_GPA	Correlation	.051	.206	.393	1.000
		Significance (2-tailed)	.654	.065	.000	.
		df	79	79	79	0

Overall, the DLPT listening scores analysis shows some possible evidence of predictor variables in this set differentially impacting outcomes for the Chinese versus the Arabic majors, but due to the N-size and lack of granularity of the DLPT measure, it is difficult to show solid evidence. FL_GPA has a much larger effect on the predictive models for Chinese than it does for Arabic. Also, the patterns of independent variable correlations between the groups vary mainly in that for the Arabic group there is a moderate correlation between SATV and FL_GPA where the Chinese group does not show that correlation.

Reading DLPT

Logistic regression with hierarchical entry of predictor variables was used once again for this analysis. As in the listening DLPT analysis, foreign language GPA (FL_GPA) was also added as an independent variable in this set of models. The same procedure was used to find the outcome measure with a score of “1” indicating that the participant was above average for reading proficiency as measured by the reading DLPT, and a score of “0” indicating average or below average. The significance of each predictor was determined when it was added to the model by looking at the change in the χ^2 term. Students who did not take the DLPT were excluded from this analysis. The descriptive statistics for the predictor variables and reading DLPT scores by language major are shown in table 37, below.

Table 37. Descriptive Statistics for Predictor Variables and Reading DLPT scores for both Majors.

		Eng_Comp		SATV		SATM		FL_GPA		DLPT_Read	
	N	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Arabic	39	3.46	0.51	678	72	669	71	3.68	0.54	12.8	8.9
Chinese	42	3.5	0.49	680.5	75.3	684.4	62.4	3.86	0.24	15.2	7.5

An independent samples t-test was run for the predictors and reading DLPT scores. Of the variables, the between group means differed significantly for FL_GPA ($P = 0.002$). Since the current analysis is looking at within group predictive patterns for above average reading DLPT scores, this difference should not impact the overall findings. Once again, the FL_GPA scores for the Chinese majors are nearer the maximum score of 4.0 which could lead to a ceiling effect for that predictor in the Chinese group much sooner than in the Arabic group. The mean reading DLPT scores are not significantly different.

The models for the Arabic majors are summarized in Table 38, below.

Table 38. Reading DLPT outcome models for Arabic Majors.

Reading DLPT Models for Arabic Majors

Model	Predictors	Model Cox & Snell R^2	Model $\chi2$	$\Delta\chi2$	P of $\Delta\chi2$
1	SATM	0.065	2.628	2.628	0.105
2	SATM, SATV	0.065	2.633	0.005	0.944
3	SATM, SATV, Eng_Comp	0.205	8.960	6.327	0.012
4	SATM, SATV, Eng_Comp, FL_GPA	0.284	13.028	4.069	0.044

As in the listening models, each consecutive model adds an additional predictor using the hierarchical order described earlier. English composition grades and foreign language GPA are significant predictors of above average reading DLPT scores for learners of Arabic. The exponent (β) terms for the four predictors in Model 4 are 1.016, 0.988, 6.461, and 16.993 for SATM, SATV, Eng_Comp, and FL_GPA, respectively.

Table 39. Independent Variable Correlations for Arabic Majors for Reading DLPT Models.

Correlations

Control Variables			SATM	SATV	Eng_Comp	FL_GPA
DLPT_Read	SATM	Correlation	1.000	.398	-.049	.023
		Significance (2-tailed)	.	.013	.769	.890
		df	0	36	36	36
	SATV	Correlation	.398	1.000	.432	.364
		Significance (2-tailed)	.013	.	.007	.025
		df	36	0	36	36
	Eng_Comp	Correlation	-.049	.432	1.000	.366
		Significance (2-tailed)	.769	.007	.	.024
		df	36	36	0	36
	FL_GPA	Correlation	.023	.364	.366	1.000
		Significance (2-tailed)	.890	.025	.024	.
		df	36	36	36	0

Table 39 shows moderate correlations between SATV and all of the other predictors. There is also a moderate correlation between FL_GPA and Eng_Comp. SATM is uncorrelated with Eng_Comp scores and FL_GPA.

The models for the Chinese majors are summarized in Table 40, below. FL_GPA is a significant predictor of above average reading DLPT scores. The exponent (β) terms for the four predictors in Model 4 are 0.999, 0.999, 0.398, and

185.607 for SATM, SATV, Eng_Comp, and FL_GPA, respectively. FL_GPA has the largest impact on reading DLPT scores, and the magnitude of the impact is 10 times larger in the Chinese model than in the Arabic model. Also, of note, English composition grades have no impact on reading DLPT outcomes for the Chinese group.

Table 40. Reading DLPT outcome models for Chinese Majors.

Reading DLPT Models for Chinese Majors

Model	Predictors	Model Cox & Snell R^2	Model $\chi 2$	$\Delta\chi 2$	P of $\Delta\chi 2$
1	SATM	0.014	0.606	0.606	0.436
2	SATM, SATV	0.024	1.025	0.419	0.517
3	SATM, SATV, Eng_Comp	0.024	1.025	0.000	1.000
4	SATM, SATV, Eng_Comp, FL_GPA	0.117	5.338	4.313	0.038

Table 41 shows the independent variable correlations for the predictors in the Chinese group. Unlike the Arabic group where moderate correlations are seen between SATV and all of the other predictors, FL_GPA is uncorrelated with SATV scores for the Chinese group. In both language groups, however, there is still a moderate correlation between FL_GPA and Eng_Comp.

Table 41. Independent Variable Correlation for Chinese Majors for Reading DLPT Models

Correlations

Control Variables			SATM	SATV	Eng_Comp	FL_GPA
DLPT_Read	SATM	Correlation	1.000	.342	.161	-.126
		Significance (2-tailed)	.	.027	.309	.427
		df	0	40	40	40
	SATV	Correlation	.342	1.000	.305	-.045
		Significance (2-tailed)	.027	.	.050	.776
		df	40	0	40	40
	Eng_Comp	Correlation	.161	.305	1.000	.484
		Significance (2-tailed)	.309	.050	.	.001
		df	40	40	0	40
	FL_GPA	Correlation	-.126	-.045	.484	1.000
		Significance (2-tailed)	.427	.776	.001	.
		df	40	40	40	0

Below, the two groups are combined and analyzed using the same procedure as above. The resultant models are displayed in Table 42. FL_GPA is a highly significant predictor of above average reading DLPT scores. The exponent (β) terms for the four predictors in Model 4 are 1.004, 0.996, 1.312, and 15.439 for SATM, SATV, Eng_Comp, and FL_GPA, respectively.

Table 42. Reading DLPT outcome models for Combined Groups.

Reading DLPT Models for Both Majors

Model	Predictors	Model Cox & Snell R^2	Model $\chi 2$	$\Delta\chi 2$	P of $\Delta\chi 2$
1	SATM	0.006	0.521	0.521	0.470
2	SATM, SATV	0.009	0.732	0.211	0.646
3	SATM, SATV, Eng_Comp	0.035	2.955	2.223	0.136
4	SATM, SATV, Eng_Comp, FL_GPA	0.120	10.467	7.512	0.006

Table 43 shows the correlations between the independent variables in this model. There is a moderate correlation between SATV and SATM as well as a moderate correlation between SATV and Eng_Comp. Also, there is a moderate correlation between Eng_Comp and FL_GPA.

Table 43. Independent Variable Correlations for the Combined Groups for DLPT Reading Models.

Correlations

Control Variables			SATM	SATV	Eng_Comp	FL_GPA
DLPT_Read	SATM	Correlation	1.000	.382	.095	.030
		Significance (2-tailed)	.	.000	.401	.789
		df	0	79	79	79
	SATV	Correlation	.382	1.000	.382	.208
		Significance (2-tailed)	.000	.	.000	.063
		df	79	0	79	79
	Eng_Comp	Correlation	.095	.382	1.000	.381
		Significance (2-tailed)	.401	.000	.	.000
		df	79	79	0	79
	FL_GPA	Correlation	.030	.208	.381	1.000
		Significance (2-tailed)	.789	.063	.000	.
		df	79	79	79	0

The independent variable correlations are very similar for the reading and listening models.

6.5 Summary of Results and Discussion

The expected results for a foreign language classroom are shown in Figure 2.

Figure 2. Expected Results for the Foreign Language Classroom

	Foreign Language Classroom		
	FL GPA	DLPT-R	DLPT-L
Predictors			
SAT Math	*	*	
SAT Verbal			
English Comp.	*		
FL GPA		*	*

Column labels are the outcome measures, and row labels are the predictor variables. An asterisk indicates an expected significant predictor of the outcome. As explained in Chapter 3, these expected findings are based on the results of the unpublished study, Wagener (2014), and the literature review.

Figure 3. Results for the Combined Foreign Language Classrooms

	Foreign Language Classroom		
	FL GPA	DLPT-R	DLPT-L
Predictors			
SAT Math	√		
SAT Verbal			
English Comp.	*		
FL GPA		*	√

Figure 3 displays the results of the current analysis. An asterisk represents a significant predictor ($P < 0.05$), and a checkmark indicates a marginally significant predictor ($P < 0.10$). The findings are very similar to what was expected. The trend for the marginally significant predictors suggests that they would become significant for a larger N-size. After taking that into account, the main difference between the expected and actual results is the lack of predictive power of SAT Math for the reading DLPT outcomes. This result does not appear to be due to N-size since there is no trend towards significance. In fact, the exponent (β) term shows that changes in the SAT Math score have no impact at all on the odds ratio for the outcome. The reason for the expectation that SAT Math would predict reading DLPT outcomes was mainly based on the findings of Surface et al. (2004) and Wagener (2014). In both cases, the participants were higher proficiency learners (ILR scale scores of 2 or greater on the reading DLPT). The fact that the participants in the current analysis were low proficiency learners may have affected the outcome.

When looking at the individual language groups in this analysis, however, a more interesting picture emerges. The limited literature on languages within language category suggests that there may be differences in predictive patterns among aptitude and achievement measures for language learning outcomes, but there little evidence to support this claim. One of the goals of the current research is to investigate predictive patterns within language category. Figure 4, below, displays the findings for the Chinese and Arabic groups.

Figure 4. Comparison of Results for Category IV Language Classroom predictor models

	Arabic Language Classroom		
	FL GPA	DLPT-R	DLPT-L
Predictors			
SAT Math			
SAT Verbal			
English Comp.	*	*	
FL GPA		*	

	Chinese Language Classroom		
	FL GPA	DLPT-R	DLPT-L
Predictors			
SAT Math	*		
SAT Verbal			
English Comp.	*		
FL GPA		*	

When splitting the two category IV languages and performing individual analyses, the aptitude and achievement predictors perform very differently. For the Chinese language, SAT Math is a highly significant predictor of foreign language grade point average. For the USNA program in this analysis, as reported by Chinese professors, 15-20% of the grade for the first two Chinese courses depends on the ability to memorize and read Chinese characters. As a student progresses to higher level classes, the requirement to read Chinese characters continues. This may suggest that SAT Math scores predict character memory performance since early on in Chinese learning, character-meaning associations may be interpreted as a non-verbal skill. If this is the case, it may explain the findings of Bamford & Mizokawa (1991) which showed a relationship between non-verbal problem solving ability and learning a foreign language. Although Cooper (1987) also mentions a correlation between foreign language learning and SAT Math scores, but his analysis was not limited to any particular language. This relationship will be further investigated in Chapter 7.

Another interesting difference in the performance of the predictors for the two languages is that English Composition scores predict both foreign language GPA and reading DLPT above average learners for the Arabic group, but they only predict foreign language GPA for the Chinese group. The relationship between English composition grades and foreign language GPA may be due to similarities in the metric and the abilities that they measure outside of foreign language performance, however, since they are both based on classroom grades. But, the fact that English composition scores predict above average learners of Arabic may indicate the transfer of L1 language skills to the L2 in the reading modality for that language group. Carson et al. (1990) had similar findings for the reading and writing modalities, but interestingly, that study involved Chinese and Japanese learners of English and also the relationship was dependent on proficiency in the L2. The possibility exists, however, that L1 transfer may also depend on the “distance” between the L1 and L2 as hypothesized by Child (1998). Taking the findings of the current analysis in combination with the findings of Carson et al. (1990), that may indicate that there is a greater distance between Chinese and English in the reading modality, and a higher proficiency level, therefore, would be required of the Chinese learners in order for L1 transfer to be available.

In any case, this analysis has provided some evidence to support the expected findings. It has also provided evidence to support a hypothesis that individual differences may differentially impact languages within the same language category. This, in turn, could be taken as support for the claims made by Child (1998) as well as the componential nature of foreign language learning aptitude. Once again, this will be further investigated in Chapter 7.

Chapter 7: Language Category Effects

7.1 Study Overview

This study will present a more comprehensive analysis of languages within and across language categories by looking at the effects of a particular set of predictor and outcome variables for each of eight different languages. The analyses will look at two languages in each of the four language categories. The current language categorization system of the Defense Language Institute Foreign Language Center (DLIFLC) was developed based on average time to train individuals in a particular language. Languages were grouped by time to train in order to facilitate scheduling at DLI. According to Lowe (1998), the system aided in planning training, but it clustered languages together whose features cause difficulties for native speakers of English, but whose nature radically differs in structure and thought patterns from language to language. Lett & O'Mara (1990) concede that even though one of the primary purposes of the DLAB is to determine probable success of learners in a particular language category, there are large variations in DLAB scores within language categories that do not necessarily predict differences in learner proficiencies. This analysis examines the current language categorization system using the predictive patterns of several aptitude measures in an attempt to identify characteristics that differentiate learner success within and across language categories.

7.2 Method

The analyses in this chapter use linear and logistic regression to examine the proficiency growth in reading and listening for DOD language specialists. Each language is analyzed to compare outcomes as well as predictive aptitude measures. Linear regression is used to evaluate outcomes at DLI graduation and three successive annual DLPT measures post-graduation. Logistic regression examines growth over the annual time intervals between testing. Growth over a time interval is coded as “1,” and no growth or loss is coded as “0”. This will allow for investigation of specific aptitude factors that predict proficiency growth over time. A follow on reverse polarity analysis is also done where attrition is coded as “1” and no attrition or growth is coded as “0”. This provides an opportunity to examine factors that prevent language proficiency loss since, more often than not in the military, language attrition is common.

This chapter begins by examining the groups as a whole to analyze the overall predictive validity of the aptitude measures across languages. Then, the group is broken into component language categories for further analysis of different predictive patterns across language category boundaries. Finally, each category is broken into individual languages to study the predictive patterns within language category. The analysis will provide evidence to help evaluate current language categorization, while looking at the best predictors of success for each of the various language categories.

7.3 Participants and Data

The participants include 1389 DOD language specialists (approximately 200 specialists in each of eight languages). The language specialists are enlisted military members from each of the four departments of the Armed Forces. As a part of the application process to join the military, members are required to take the Armed Services Vocational Aptitude Battery (ASVAB). Successful performance on the ASVAB allows members to select from a variety of career paths. If the member wishes to pursue a career as a language specialist, they are administered the Defense Language Aptitude Battery (DLAB). Recruits with high scores on the DLAB are then encouraged to accept positions as specialists in languages commensurate with the level of their scores. Once individuals are selected, training begins at the Defense Language Institute (DLI) where individuals take the prescribed intensive immersion program for their particular languages. The language courses range from 26 to 64 weeks depending on language category, as described earlier. Upon completion of training, participants take the Defense Language Proficiency Test (DLPT) to assess their proficiency in the L2. Graduates of DLI are expected to score ILR Level 2 in listening and reading and Level 1+ in speaking. Defense Language Proficiency Tests (DLPT) are scored using the Interagency Language Roundtable (ILR) Skill Level Descriptions (1985). Military language specialists are required to test annually.

The data was provided by the Defense Manpower and Data Center (DMDC) in coordination with the Defense Language Institute Foreign Language Center (DLIFLC) and includes DLAB and ASVAB scores as well as four consecutive annual

DLPT reading and listening scores beginning upon DLI graduation for each individual. Language specialists are actively encouraged to continue self-study of the newly learned L2 through pay incentives, specifically an additional monthly stipend based on their most recent DLPT scores. The ASVAB has served as an entrance exam for the military since 1976 (Segall and Moreno, 1999). The original purpose of the ASVAB was to predict occupational success in the military. The ASVAB is broken into four sections including arithmetic reasoning (AR), math knowledge (MK), word knowledge (WK) and paragraph comprehension (PC). DMDC provided scores for each of the four parts for each participant. The DLAB includes four sections: biographical data, spoken stress, deductive rule application and inductive pattern application (Lett et al., 2003). Only the overall DLAB score was provided for the current study.

7.4 Analyses

7.4.1 Combined Language Groups

Foreign Language GPA

Linear regression with hierarchical entry of predictor variables is used for this analysis. This analysis uses the grade point average of the participants at graduation from DLI as the outcome measure. The significance of each predictor is determined when it is added to the model by looking at the change in the R^2 term. The predictors are added from general cognitive to more verbal specific aptitudes, with foreign language aptitude being the final predictor added to the models. This order is intended to show the incremental predictive validity of specific aptitudes. The descriptive statistics for the predictor variables and the outcome measure by language is shown in table 44, below. Standard deviations are shown in parenthesis.

Table 44. Descriptive statistics by language for predictors and DLI GPA.

Language	N	AR	MK	PC	WK	DLAB	DLI GPA
French	197	57.7 (7.4)	60.5 (5.5)	57.0 (5.5)	56.7 (6.2)	104.2 (12.9)	3.27 (0.43)
Spanish	198	58.9 (5.8)	60.9 (5.3)	57.8 (5.2)	57.8 (4.7)	103.7 (10.7)	3.28 (0.43)
German	116	58.3 (6.7)	60.8 (6.4)	56.9 (5.1)	56.9 (6.9)	105.5 (13.5)	3.07 (0.48)
Indonesian	104	60.4 (6.2)	61.1 (5.2)	59.7 (5.5)	59.5 (7.1)	110.3 (11.6)	3.54 (0.34)
Russian	190	61.0 (5.4)	62.8 (4.4)	58.8 (4.0)	59.0 (4.5)	109.5 (11.5)	3.26 (0.44)
Tagalog	186	59.0 (7.1)	61.1 (5.7)	58.1 (5.6)	57.0 (6.3)	107.0 (10.9)	3.44 (0.34)
Arabic	197	61.1 (5.9)	62.9 (5.0)	59.4 (4.9)	59.7 (5.6)	118.2 (12.1)	3.17 (0.42)
Chinese	201	61.2 (6.0)	63.5 (4.0)	59.3 (4.9)	59.4 (5.5)	119.1 (13.3)	3.32 (0.39)

Table 45. Foreign Language GPA outcome models for DLI Graduates.

Foreign Language GPA for DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.008	11.317	0.001
2	AR, MK	0.010	2.802	0.061
3	AR, MK, PC	0.014	2.813	0.038
4	AR, MK, PC, WK	0.014	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.063	18.107	0.000

The models for the DLI graduates are summarized in Table 45, above. Each consecutive model adds an additional predictor using the hierarchical order described earlier. A significant change in R^2 happens when Arithmetic Reasoning (AR) is added to the model ($F = 11.317$, $P = 0.001$). A significant change in R^2 also happens when Paragraph Comprehension (PC) is added to the model ($F = 2.813$, $P = 0.038$). Finally, a significant change R^2 occurs when DLAB is added ($F = 18.107$, $P < 0.001$). The model with all five predictors is also highly significant ($F = 18.683$, $P < 0.001$). The standardized β coefficients for the five predictors in Model 5 are 0.009, -0.011, 0.051, -0.004, 0.238 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression model yielded a highly significant model with DLAB as the only predictor ($F = 89.479$, $P < 0.001$) in the model explaining 6.1% of the variance. The standardized β coefficient for DLAB in the post hoc model is 0.246.

Table 46. Independent Variable Correlation for DLI Graduates for DLI GPA Models.

Correlations

Control Variables			AR	MK	PC	WK	DLAB
DLI_GPA	AR	Correlation	1.000	.432	.540	.501	.238
		Significance (2-tailed)	.	.000	.000	.000	.000
		df	0	1386	1386	1386	1386
	MK	Correlation	.432	1.000	.190	.182	.313
		Significance (2-tailed)	.000	.	.000	.000	.000
		df	1386	0	1386	1386	1386
	PC	Correlation	.540	.190	1.000	.614	.173
		Significance (2-tailed)	.000	.000	.	.000	.000
		df	1386	1386	0	1386	1386
	WK	Correlation	.501	.182	.614	1.000	.163
		Significance (2-tailed)	.000	.000	.000	.	.000
		df	1386	1386	1386	0	1386
	DLAB	Correlation	.238	.313	.173	.163	1.000
		Significance (2-tailed)	.000	.000	.000	.000	.
		df	1386	1386	1386	1386	0

Table 46 shows the correlations between the independent variables for the DLI GPA model. There are moderate correlations between Arithmetic Reasoning (AR) the other three components of the ASVAB: Math Knowledge (MK), Paragraph Comprehension (PC), and Word Knowledge (WK). There is also a moderate correlation between Paragraph Comprehension and Word Knowledge.

Listening DLPT

Table 47. Mean Listening DLPT scores for DLI Graduates by Language Group.

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
French	197	24.4 (4.5)	23.3 (5.3)	23.7 (5.1	23.4 (5.0
Spanish	198	23.8 (4.7)	23.7 (5.1)	24.4 (4.7)	24.7 (4.6)
German	116	19.9 (5.1)	19.7 (5.2)	20.4 (4.5)	20.5 (5.0)
Indonesian	104	26.5 (3.7)	25.4 (4.1)	25.5 (3.3)	24.8 (2.6)
Russian	190	23.7 (4.7)	22.9 (4.3)	23.0 (4.6)	23.5 (3.7)
Tagalog	186	23.6 (4.2)	22.9 (4.3)	23.0 (4.6)	23.5 (3.7)
Arabic	197	22.1 (5.8)	20.3 (6.5)	20.3 (6.6)	20.2 (5.8)
Chinese	201	23.1 (4.9)	21.6 (5.7)	21.5 (6.2)	21.9 (5.8)

The same procedure as described in the previous section is used for the subsequent listening analyses for the DLI graduates to examine predictor performance on four annual listening Defense Language Proficiency Tests. Table 47 shows the mean score for each of the four listening tests by language group. As described previously, the DLPT scores are transformed by multiplying the ILR scale score by 10. Then, if the scale score has a ‘+” value assigned, 6 is added to the transformed score. For example, an ILR scale score of 1+ is transformed to equal 16. The standard deviations are shown in parenthesis. Once again, the predictors are entered into the models using the hierarchical order described earlier. The first listening DLPT was given at the time of graduation. The results of the analysis are displayed in Table 48.

Table 48. Listening DLPT 1 outcome models for DLI Graduates.

Listening DLPT 1 for DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.006	8.959	0.003
2	AR, MK	0.012	8.423	0.000
3	AR, MK, PC	0.015	2.112	0.097
4	AR, MK, PC, WK	0.015	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.032	6.081	0.000

A significant change in R^2 happens when Arithmetic Reasoning (AR) is added to the model ($F = 8.959$, $P = 0.003$). A significant change in R^2 also happens when Math Knowledge (MK) is added to the model ($F = 8.423$, $P < 0.001$). Finally, a significant change R^2 occurs when DLAB is added ($F = 6.081$, $P < 0.001$). The model with all five predictors is also highly significant ($F = 9.018$, $P < 0.001$). The standardized β coefficients for the five predictors in Model 5 are -0.001, 0.045, 0.060, -0.013, and

0.139 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression model yielded a highly significant model with DLAB and PC as the two predictors ($F = 21.238$, $P < 0.001$) in the model explaining 3.0% of the variance. The standardized β coefficients for PC and DLAB in the post hoc model are 0.058 and 0.152, respectively.

Table 49. Listening DLPT 2 outcome models for DLI Graduates.

Listening DLPT 2 for DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	2.418	0.120
2	AR, MK	0.004	2.785	0.062
3	AR, MK, PC	0.005	0.697	0.554
4	AR, MK, PC, WK	0.005	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.012	2.453	0.626

Approximately one year after graduating from DLI, participants took their second DLPT. The mean score for the participants' language group on the second listening DLPT was substituted for missing data. This affected 2.5% of the participant data. The models are displayed in Table 49. The only predictor that has a marginally significant effect on the models is MK ($F = 2.875$, $P = 0.062$). The standardized β coefficients for the five predictors in Model 5 are -0.001, 0.031, 0.028, -0.018, 0.090 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression model yielded a highly significant model with DLAB as the only predictor ($F = 14.716$, $P < 0.001$) in the model explaining 1.0% of the variance.

Table 50. Listening DLPT 3 outcome models for DLI Graduates.

Listening DLPT 3 for DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	2.354	0.125
2	AR, MK	0.003	1.391	0.249
3	AR, MK, PC	0.003	0.000	1.000
4	AR, MK, PC, WK	0.004	0.464	0.762
5	AR, MK, PC, WK, DLAB	0.004	0.000	1.000

Approximately one year later, participants took their third DLPT. The mean score for the participants' language group on the third listening DLPT was substituted for missing data. This affected 9.8% of the participant data. The models for the third listening DLPT are displayed in Table 50. None of the predictors are significant for this listening DLPT. The standardized β coefficients for the five predictors in Model 5 are 0.006, 0.034, -0.001, 0.026, 0.029 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression model yielded no significant models.

Finally, three years after graduation, participants took their fourth listening DLPT. The mean score for the participants' language group on this listening DLPT was substituted, once again, for missing data. This affected 23.5% of the participant data. The results of the predictive models for the fourth listening DLPT are shown in Table 51. A significant change in R^2 happens when Arithmetic Reasoning (AR) is added to the model ($F = 6.006$, $P = 0.014$). Also, a highly significant change in R^2 happens when Math Knowledge (MK) is added to the model ($F = 5.593$, $P = 0.004$). The standardized β coefficients for the five predictors in Model 5 are 0.034, 0.060, -0.010, 0.015, 0.015 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression model yielded a highly significant model with MK as the only predictor ($F = 8.935$, $P = 0.003$) in the model explaining 0.6% of the variance. The standardized β coefficient for MK in the post hoc model is 0.080.

Table 51. Listening DLPT 4 outcome models for DLI Graduates.

Listening DLPT 4 for DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.004	6.006	0.014
2	AR, MK	0.008	5.593	0.004
3	AR, MK, PC	0.008	0.000	1.000
4	AR, MK, PC, WK	0.008	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.008	0.000	1.000

Reading DLPT

Concurrent with the listening Defense Language Proficiency Tests, the DLI graduates also took the reading DLPT. This next section examines predictor performance on four annual reading Defense Language Proficiency Tests for the DLI graduates. Table 52 shows the mean score for each of the four reading tests by language group. The standard deviations are shown in parenthesis. As in the listening analyses, the predictors are entered into the models using the hierarchical order described previously. The first reading DLPT was given at the time of graduation, and the three subsequent tests were given at one year intervals. The results of the analysis for the first reading DLPT are displayed in Table 53.

Table 52. Mean Reading DLPT scores for DLI Graduates by Language Group.

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
French	197	24.8 (5.0)	23.3 (5.5)	24.5 (5.1)	24.4 (4.6)
Spanish	198	27.5 (3.3	27.1 (4.0)	27.0 (4.5)	27.7 (3.5)
German	116	27.4 (4.9)	25.8 (6.0)	27.0 (4.9)	26.4 (5.5)
Indonesian	104	26.0 (3.2)	24.7 (3.9)	24.9 (2.9)	25.2 (1.9)
Russian	190	24.8 (4.5)	24.4 (5.1)	24.5 (5.1)	24.9 (4.5)
Tagalog	186	23.9 (4.1)	23.2 (4.1)	23.2 (4.0)	23.4 (3.7)
Arabic	197	23.3 (5.3)	21.6 (5.9)	21.3 (6.0)	21.3 (5.8)
Chinese	201	25.7 (4.3)	23.2 (5.3)	22.8 (5.8)	22.8 (5.7)

Arithmetic Reasoning (AR) causes a highly significant change in R^2 when added to the model ($F = 9.324$, $P = 0.002$). A highly significant change in R^2 also happens when Math Knowledge (MK) is added to the model ($F = 14.110$, $P < 0.001$).

Finally, DLAB causes a highly significant change in R^2 added ($F = 4.292$, $P = 0.001$). The model with all five predictors is highly significant ($F = 9.140$, $P < 0.001$). The

Table 53. Reading DLPT 1 outcome models for DLI Graduates.

Reading DLPT 1 for DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.007	9.324	0.002
2	AR, MK	0.017	14.110	0.000
3	AR, MK, PC	0.019	1.414	0.237
4	AR, MK, PC, WK	0.020	0.471	0.757
5	AR, MK, PC, WK, DLAB	0.032	4.292	0.001

standardized β coefficients for the five predictors in Model 5 are -0.018, 0.084, 0.032, 0.031, and 0.118 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression model yielded a highly significant model with DLAB and MK as the two predictors ($F = 21.192$, $P < 0.001$) in the model explaining 3.0% of the variance. The standardized β coefficients for MK and DLAB in the post hoc model are 0.086 and 0.124, respectively.

As in the listening section, approximately one year after graduating from DLI, participants took their second DLPT. The mean score for the participants' language group on the second reading DLPT was substituted for missing data. This affected

Table 54. Reading DLPT 2 outcome models for DLI Graduates.

Reading DLPT 2 for DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	2.650	0.104
2	AR, MK	0.005	4.182	0.016
3	AR, MK, PC	0.006	0.698	0.553
4	AR, MK, PC, WK	0.006	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.009	1.048	0.370

1.9% of the participant data. The models are displayed in Table 54. Math Knowledge is the only predictor to have a significant impact when added to the model (F = 4.182, P = 0.016). The model with all five predictors is significant (F = 2.599, P = 0.024). The standardized β coefficients for the five predictors in Model 5 are -0.011, 0.049, 0.021, 0.014, 0.061 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression model yielded a highly significant model with DLAB as the only predictor (F = 8.945, P = 0.003) in the model explaining 0.6% of the variance. A post hoc linear regression model with MK as the only predictor yields a highly significant model (F = 6.888, P = 0.009) explaining 0.5% of the variance.

Approximately two years after graduation, the participants took their third reading DLPT. The results for the predictor models are shown in Table 55. Once again, the mean score for the participants' language group on the third listening

Table 55. Reading DLPT 3 outcome models for DLI Graduates.

Reading DLPT 3 for DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	1.612	0.204
2	AR, MK	0.004	4.178	0.016
3	AR, MK, PC	0.004	0.000	1.000
4	AR, MK, PC, WK	0.005	0.464	0.762
5	AR, MK, PC, WK, DLAB	0.005	0.000	1.000

DLPT was substituted for missing data. This affected 9.9% of the participant data. As in the previous year's model, Math Knowledge is the only predictor to have a significant impact ($F = 4.178$, $P = 0.016$). The model with all five predictors is not significant. The standardized β coefficients for the five predictors in Model 5 are 0.003, 0.053, -0.020, 0.024, 0.029 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression model yielded a significant model with MK as the only predictor ($F = 5.745$, $P = 0.017$) in the model explaining 0.4% of the variance.

Table 56. Reading DLPT 4 outcome models for DLI Graduates.

Reading DLPT 4 for DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	2.312	0.129
2	AR, MK	0.003	1.391	0.249
3	AR, MK, PC	0.003	0.000	1.000
4	AR, MK, PC, WK	0.003	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.005	0.696	0.627

Finally, three years after graduation, participants took their fourth reading DLPT. The mean score for the participants' language group on this reading DLPT was substituted for missing data. This affected 23.5% of the participant data. The results of the predictive models are shown in Table 56. There are no significant predictors or significant models in this analysis. The standardized β coefficients for the five predictors in Model 5 are 0.020, 0.057, 0.005, 0.010, -0.048 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression model yielded a significant model with MK as the only predictor ($F = 3.979$, $P = 0.046$) explaining 0.3% of the variance. The standardized β coefficient for MK in the post hoc model is 0.053.

Oral Proficiency Interview (OPI)

Table 57. Mean OPI scores for DLI Graduates by Language Group.

Language	N	OPI
French	197	18.8 (2.4)
Spanish	198	20.0 (2.6)
German	116	18.3 (2.6)
Indonesian	104	18.7 (2.4)
Russian	190	18.2 (3.2)
Tagalog	186	19.2 (2.2)
Arabic	197	17.4 (2.5)
Chinese	201	18.6 (2.4)

In addition to the listening and reading Defense Language Proficiency Tests, each DLI graduate also takes the Oral Proficiency Interview. More often than not, this is the only time in a military member's career that he/she will take the OPI. Therefore, it serves as a one-time benchmark of the foreign language speaking ability of each participant. This next section examines predictor performance of OPI scores for the DLI graduates. Table 57 shows the mean OPI score by language group. The standard deviations are shown in parenthesis.

The results of the analysis for the OPI are displayed in Table 58. There are no significant predictors or models in this analysis. The standardized β coefficients for the five predictors in Model 5 are 0.062, -0.025, -0.035, -0.054, 0.028 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression also failed to yield a significant model for predicting OPI scores.

Table 58. OPI outcome models for DLI Graduates.

Oral Proficiency Interview 1 for DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.000	0.184	0.668
2	AR, MK	0.000	0.277	0.758
3	AR, MK, PC	0.003	1.878	0.131
4	AR, MK, PC, WK	0.004	0.464	0.762
5	AR, MK, PC, WK, DLAB	0.005	0.348	0.884

7.4.2 Language Categories

A similar set of analyses are performed in this section, but the participant data are now sorted into groups by language category. Outcome measures, once again, include foreign language GPA, listening and reading DLPT scores, and OPI scores. Two additional analyses are also added. These include a year-to-year growth analysis and a year-to-year attrition analysis. These last two analyses use logistic regression.

Foreign Language GPA

The foreign language GPA analysis by language category uses linear regression with hierarchical entry of predictor variables, as used previously. This analysis uses the grade point average of the participants at graduation from DLI as the outcome measure. The significance of each predictor is determined when it is added to the model by looking at the change in the R^2 term. The predictors are added according to chronological progression of test scores, primarily; and general cognitive to more verbal specific aptitudes, secondarily. The descriptive statistics for the predictor variables and the outcome measure by language category are shown in table 59, below. Standard deviations are shown in parenthesis.

Table 59. Descriptive statistics by language category for predictors and DLI GPA.

Language Category	N	AR	MK	PC	WK	DLAB	DLI GPA
I	395	58.3 (6.7)	60.7 (5.4)	57.4 (5.3)	57.3 (5.5)	104.0 (11.9)	3.27 (0.44)
II	220	59.3 (6.5)	60.9 (5.8)	58.2 (5.5)	58.1 (7.1)	107.8 (12.8)	3.30 (0.48)
III	376	60.0 (6.4)	62.0 (5.2)	58.5 (4.9)	58.0 (5.5)	108.2 (11.3)	3.35 (0.40)
IV	398	61.2 (6.0)	63.2 (4.5)	59.3 (4.9)	59.6 (5.5)	118.6 (12.7)	3.25 (0.41)

The next figure (Figure 5) is divided by language category and shows the models for predicting DLI GPA. The models for category 1 languages show that

Figure 5. Predictive Models for Foreign Language Grade Point Average by Language Category.

Foreign Language GPA for CAT I DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.011	4.507	0.034
2	AR, MK	0.013	0.794	0.453
3	AR, MK, PC	0.014	0.199	0.897
4	AR, MK, PC, WK	0.017	0.398	0.810
5	AR, MK, PC, WK, DLAB	0.091	7.937	0.000

Foreign Language GPA for CAT II DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.025	5.549	0.019
2	AR, MK	0.026	0.223	0.800
3	AR, MK, PC	0.042	1.812	0.146
4	AR, MK, PC, WK	0.042	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.119	4.698	0.000

Foreign Language GPA for CAT III DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.000	0.122	0.727
2	AR, MK	0.003	1.122	0.327
3	AR, MK, PC	0.004	0.187	0.905
4	AR, MK, PC, WK	0.006	0.249	0.910
5	AR, MK, PC, WK, DLAB	0.080	7.460	0.000

Foreign Language GPA for CAT IV DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.022	9.024	0.003
2	AR, MK	0.028	2.438	0.089
3	AR, MK, PC	0.033	1.021	0.383
4	AR, MK, PC, WK	0.033	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.109	8.380	0.000

Arithmetic Reasoning (AR) is a significant predictor and DLAB is a highly significant predictor of DLI GPA. With all five predictors in the Category I model, the model is highly significant ($F = 7.747$, $P < 0.001$) and explains 9.1% of the variance. The standardized β coefficients are 0.035, -0.022, -0.016, 0.065, and 0.285 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 36.407$, $P < 0.001$) where DLAB is the only predictor in the model. The post hoc model explains 8.5% of the variance.

Similarly for Category II languages, AR and DLAB are significant predictors with DLAB being highly significant. The model with all five predictors is highly significant ($F = 5.769$, $P < 0.001$) and explains 11.9% of the variance. The standardized β coefficients are 0.037, -0.061, 0.099, 0.066, and 0.298 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 13.870$, $P < 0.001$) where PC and DLAB are retained in the model. Here, the standardized β coefficients are 0.145 and 0.280 for PC and DLAB, respectively. The post hoc model explains 11.3% of the variance.

For the Category III languages, DLAB is the only significant predictor of DLI GPA. A model with all five predictors is highly significant ($F = 6.439$, $P < 0.001$) and explains 8.0% of the variance. The standardized β coefficients are -0.066, 0.009, 0.058, -0.062, and 0.279 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model (F =

29.379, P < 0.001) that retains only DLAB as a predictor and explains 7.3% of the variance. The standardized β coefficient for DLAB in the post hoc model is 0.270.

Finally, for the Category IV languages, AR and DLAB are highly significant predictors of DLI GPA, and MK is marginally significant. The model with all five predictors accounts for 10.9% of the variance is highly significant (F = 9.633, P < 0.001). The standardized β coefficients are 0.045, 0.019, 0.082, -0.041, and 0.295 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model (F = 44.423, P < 0.001) where DLAB is the only predictor in the model. The post hoc model explains 10.1% of the variance.

When examining variable correlations in DLI GPA models, as seen in Table 60 below, there is a moderate correlation between AR and all of the other ASVAB predictors. Additionally, PC moderately correlates with WK. This may be resulting in some collinearity between the predictors in the various language category models. DLAB, on the other hand, has a low correlation with the other predictors.

Table 60. Independent Variable Correlations for DLI GPA Models.

Correlations

Control Variables			AR	MK	PC	WK	DLAB
Lang_Cat & DLI_GPA	AR	Correlation	1.000	.413	.528	.489	.187
		Significance (2-tailed)	.	.000	.000	.000	.000
		df	0	1385	1385	1385	1385
	MK	Correlation	.413	1.000	.168	.160	.264
		Significance (2-tailed)	.000	.	.000	.000	.000
		df	1385	0	1385	1385	1385
	PC	Correlation	.528	.168	1.000	.606	.127
		Significance (2-tailed)	.000	.000	.	.000	.000
		df	1385	1385	0	1385	1385
	WK	Correlation	.489	.160	.606	1.000	.117
		Significance (2-tailed)	.000	.000	.000	.	.000
		df	1385	1385	1385	0	1385
	DLAB	Correlation	.187	.264	.127	.117	1.000
		Significance (2-tailed)	.000	.000	.000	.000	.
		df	1385	1385	1385	1385	0

Listening DLPT for DLI Graduates

The Listening DLPT analysis by language category also uses linear regression with hierarchical entry of predictor variables. Figure 6 is divided by language category and shows the models for predicting Listening DLPT 1, which is the DLPT taken by participants at graduation from DLI. The models for Category I languages show that Arithmetic Reasoning (AR) is a highly significant predictor of Listening

Figure 6. Listening DLPT 1 Predictive Models for DLI Graduates by Language Category.

Listening DLPT 1 for CAT I DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.031	12.489	0.000
2	AR, MK	0.036	2.033	0.132
3	AR, MK, PC	0.039	0.612	0.608
4	AR, MK, PC, WK	0.042	0.408	0.803
5	AR, MK, PC, WK, DLAB	0.054	1.237	0.291

Listening DLPT 1 for CAT II DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.01	2.145	0.144
2	AR, MK	0.014	0.880	0.416
3	AR, MK, PC	0.021	0.777	0.508
4	AR, MK, PC, WK	0.023	0.147	0.964
5	AR, MK, PC, WK, DLAB	0.131	6.680	0.000

Listening DLPT 1 for CAT III DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0	0.155	0.694
2	AR, MK	0.008	1.750	0.175
3	AR, MK, PC	0.01	0.219	0.883
4	AR, MK, PC, WK	0.01	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.045	1.970	0.082

Listening DLPT 1 for CAT IV DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.01	3.892	0.049
2	AR, MK	0.025	6.077	0.003
3	AR, MK, PC	0.03	1.018	0.385
4	AR, MK, PC, WK	0.034	0.544	0.704
5	AR, MK, PC, WK, DLAB	0.086	5.590	0.000

DLPT 1 scores. With all five predictors in the Category I model, the model is highly significant (F = 4.447, P = 0.001) and explains 5.4% of the variance. The standardized β coefficients are 0.081, 0.061, 0.026, 0.066, and 0.116 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 9.762$, $P < 0.001$) where AR and DLAB are retained in the model. The post hoc model explains 4.7% of the variance, and the standardized β coefficients are 0.149 and 0.132 for AR and DLAB, respectively.

For Category II languages, on the other hand, DLAB is a highly significant predictor of the listening scores. The model with all five predictors is highly significant ($F = 6.439$, $P < 0.001$) and explains 13.1% of the variance. The standardized β coefficients are -0.024, -0.043, 0.013, 0.128, and 0.352 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 28.594$, $P < 0.001$) where DLAB is the only predictor in the model. Here, the standardized β coefficient for DLAB is 0.341. The post hoc model explains 11.6% of the variance.

For the Category III languages, none of the predictors have a significant impact on the model, and only DLAB is marginally significant. A model with all five predictors is highly significant ($F = 3.521$, $P = 0.004$) but explains only 4.5% of the variance. The standardized β coefficients for the predictors in the model are -0.057, 0.063, 0.071, -0.032, and 0.192 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 15.328$, $P < 0.001$) that retains only DLAB as a predictor and explains 3.9% of the variance. The standardized β coefficient for DLAB in the post hoc model is 0.198.

Finally, for the Category IV languages, AR, MK and DLAB all have a significant impact on the model, with the impact of MK and DLAB being highly significant. The model with all five predictors, accounting for 8.6% of the variance, is highly significant ($F = 7.378$, $P < 0.001$). The standardized β coefficients are -0.001, 0.081, 0.135, -0.124, and 0.244 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 29.508$, $P < 0.001$) where DLAB is the only predictor in the model. The post hoc model explains 6.9% of the variance.

When examining variable correlations in the DLPT 1 listening models, as seen in Table 61 below, there is a moderate correlation between AR and all of the other ASVAB predictors. Additionally, PC moderately correlates with WK. This may be resulting in some collinearity between the predictors in the various language category models. DLAB, on the other hand, has a low correlation with the other predictors.

Table 61. Independent Variable Correlations for DLPT 1 Listening Models.

Correlations

Control Variables			AR	MK	PC	WK	DLAB
Lang_Cat & DLPT_List1	AR	Correlation	1.000	.411	.528	.489	.188
		Significance (2-tailed)	.	.000	.000	.000	.000
		df	0	1385	1385	1385	1385
	MK	Correlation	.411	1.000	.165	.158	.257
		Significance (2-tailed)	.000	.	.000	.000	.000
		df	1385	0	1385	1385	1385
	PC	Correlation	.528	.165	1.000	.606	.130
		Significance (2-tailed)	.000	.000	.	.000	.000
		df	1385	1385	0	1385	1385
	WK	Correlation	.489	.158	.606	1.000	.120
		Significance (2-tailed)	.000	.000	.000	.	.000
		df	1385	1385	1385	0	1385
	DLAB	Correlation	.188	.257	.130	.120	1.000
		Significance (2-tailed)	.000	.000	.000	.000	.
		df	1385	1385	1385	1385	0

Figure 7, below, is divided by language category and shows the models for predicting Listening DLPT 2 scores. The models were produced using the same method as above. Arithmetic Reasoning (AR) is once again a significant predictor of Listening DLPT 2 scores for the Category I languages. With all five predictors in the Category I model, the model is significant (F = 2.364, P = 0.036) and explains 2.9% of the variance. The standardized β coefficients are 0.050, 0.019, 0.026, 0.048, and 0.107 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model (F = 7.115, P 0.008) where DLAB is the only predictor in the model. The post hoc model explains 1.8% of the variance, and the standardized β coefficient for DLAB is 0.133.

Figure 7. Listening DLPT 2 Predictive Models for DLI Graduates by Language Category.

Listening DLPT 2 for CAT I DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.014	5.567	0.019
2	AR, MK	0.015	0.398	0.672
3	AR, MK, PC	0.018	0.599	0.616
4	AR, MK, PC, WK	0.019	0.133	0.970
5	AR, MK, PC, WK, DLAB	0.029	1.004	0.415

Listening DLPT 2 for CAT II DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.004	0.826	0.365
2	AR, MK	0.008	0.875	0.418
3	AR, MK, PC	0.016	0.882	0.451
4	AR, MK, PC, WK	0.017	0.073	0.990
5	AR, MK, PC, WK, DLAB	0.107	5.417	0.000

Listening DLPT 2 for CAT III DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.218	0.641
2	AR, MK	0.008	2.632	0.073
3	AR, MK, PC	0.014	1.135	0.335
4	AR, MK, PC, WK	0.016	0.252	0.908
5	AR, MK, PC, WK, DLAB	0.039	2.220	0.052

Listening DLPT 2 for CAT IV DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.006	2.249	0.134
2	AR, MK	0.017	4.420	0.013
3	AR, MK, PC	0.018	0.201	0.896
4	AR, MK, PC, WK	0.018	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.060	4.390	0.001

For Category II languages, DLAB is a highly significant predictor of the listening scores. The model with all five predictors is also highly significant ($F = 5.142$, $P < 0.001$) and explains 10.7% of the variance. The standardized β coefficients are -0.058, -0.029, 0.043, 0.086, and 0.322 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 23.628$, $P < 0.001$) where DLAB is the only predictor in the model. Here, the standardized β coefficient for DLAB is 0.313. The post hoc model explains 9.8% of the variance.

For the Category III languages, the predictor with the largest impact when added to the model is MK, but even it is only marginally significant. A model with all five predictors is significant ($F = 3.521$, $P = 0.004$) but explains only 3.9% of the variance. The standardized β coefficients for the predictors in the model are -0.064, 0.076, 0.110, -0.050, and 0.157 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 10.597$, $P = 0.001$) that retains only DLAB as a predictor and explains 2.8% of the variance. The standardized β coefficient for DLAB in the post hoc model is 0.166.

Finally, for the Category IV languages, MK has a significant impact on the model, and DLAB has a highly significant impact. The model with all five predictors accounts for 6.0% of the variance and is highly significant ($F = 5.014$, $P < 0.001$). The standardized β coefficients are 0.037, 0.058, -0.037, -0.038, and 0.220 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 22.281$, $P < 0.001$) where

DLAB is the only predictor in the model. The post hoc model explains 5.3% of the variance.

Figure 8. Listening DLPT 3 Predictive Models for DLI Graduates by Language Category.

Listening DLPT 3 for CAT I DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.013	5.189	0.023
2	AR, MK	0.015	0.796	0.452
3	AR, MK, PC	0.023	1.605	0.188
4	AR, MK, PC, WK	0.027	0.536	0.709
5	AR, MK, PC, WK, DLAB	0.030	0.302	0.912

Listening DLPT 3 for CAT II DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.219	0.640
2	AR, MK	0.006	1.092	0.337
3	AR, MK, PC	0.017	1.214	0.306
4	AR, MK, PC, WK	0.019	0.147	0.964
5	AR, MK, PC, WK, DLAB	0.100	4.838	0.000

Listening DLPT 3 for CAT III DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	0.708	0.401
2	AR, MK	0.004	0.749	0.474
3	AR, MK, PC	0.005	0.187	0.905
4	AR, MK, PC, WK	0.007	0.250	0.910
5	AR, MK, PC, WK, DLAB	0.025	1.712	0.131

Listening DLPT 3 for CAT IV DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.011	4.343	0.038
2	AR, MK	0.023	4.852	0.008
3	AR, MK, PC	0.026	0.608	0.610
4	AR, MK, PC, WK	0.026	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.044	1.850	0.102

As shown in Figure 8, Arithmetic Reasoning (AR) is a significant predictor of Listening DLPT 3 scores for the Category I languages. With all five predictors in the Category I model, the model is significant (F = 2.400, P = 0.037) and explains 3.0% of the variance. The standardized β coefficients are 0.005, 0.050, 0.062, 0.088, and 0.052 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model (F = 7.664, P = 0.006) where WK is the only predictor in the model. The post hoc model explains 1.9% of the variance, and the standardized β coefficient for WK is 0.138.

Once again for Category II languages, DLAB is a highly significant predictor of the listening scores. The model with all five predictors is also highly significant ($F = 4.773$, $P < 0.001$) and explains 10.0% of the variance. The standardized β coefficients are -0.105, -0.025, 0.047, 0.119, and 0.305 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 19.933$, $P < 0.001$) with DLAB as the only predictor in the model. The standardized β coefficient for DLAB in the post hoc model is 0.289, and the model explains 8.4% of the variance.

The predictors in the models for the Category III language are not significant. The model with all five predictors is only marginally significant ($F = 1.931$, $P = 0.088$) and explains only 2.5% of the variance. The standardized β coefficients for the predictors in the model are -0.022, 0.033, -0.002, 0.064, and 0.138 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis, however, does yield a highly significant predictive model ($F = 8.120$, $P = 0.005$) with DLAB as the only predictor. The model explains 2.1% of the variance. The standardized β coefficient for DLAB in the post hoc model is 0.146.

For the Category IV languages, AR and MK have a significant impact on the model, with MK being highly significant impact. The model with all five predictors accounts for 4.4% of the variance and is highly significant ($F = 3.642$, $P = 0.003$). The standardized β coefficients are 0.076, 0.082, -0.058, -0.019, and 0.147 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 12.564$, $P < 0.001$) where DLAB is the only predictor in the model. The post hoc model explains 3.1% of the variance, and the standardized β coefficient for DLAB in the model is 0.175.

Figure 9. Listening DLPT 4 Predictive Models for DLI Graduates by Language Category.

Listening DLPT 4 for CAT I DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.015	5.800	0.016
2	AR, MK	0.019	1.598	0.204
3	AR, MK, PC	0.020	0.200	0.896
4	AR, MK, PC, WK	0.022	0.267	0.899
5	AR, MK, PC, WK, DLAB	0.026	0.400	0.849

Listening DLPT 4 for CAT II DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.122	0.727
2	AR, MK	0.027	5.799	0.004
3	AR, MK, PC	0.040	1.469	0.224
4	AR, MK, PC, WK	0.042	0.150	0.963
5	AR, MK, PC, WK, DLAB	0.121	4.831	0.000

Listening DLPT 4 for CAT III DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.005	1.998	0.158
2	AR, MK	0.013	3.023	0.050
3	AR, MK, PC	0.015	0.379	0.768
4	AR, MK, PC, WK	0.015	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.030	1.434	0.211

Listening DLPT 4 for CAT IV DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.025	10.104	0.002
2	AR, MK	0.032	2.856	0.059
3	AR, MK, PC	0.037	1.025	0.381
4	AR, MK, PC, WK	0.038	0.137	0.969
5	AR, MK, PC, WK, DLAB	0.051	1.346	0.244

For the fourth listening DLPT (taken 3 years after DLI graduation), Figure 9 demonstrates that Arithmetic Reasoning (AR) remains a significant predictor of Listening DLPT 4 scores for the Category I languages, but the model with all five predictors is only marginally significant ($F = 2.037$, $P = 0.073$), explaining only 2.6% of the variance. The standardized β coefficients for the model are 0.052, 0.061, 0.007, 0.055, and 0.066 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a significant predictive model (F = 5.800, P = 0.016) with AR as the only predictor in the model. The post hoc model explains 1.5% of the variance, and the standardized β coefficient for AR is 0.121.

For Category II languages, MK and DLAB are both highly significant predictors of the listening scores. The model with all five predictors is also highly significant (F = 5.867, P < 0.001) and explains 12.1% of the variance. The standardized β coefficients are -0.168, 0.081, 0.057, 0.120, and 0.300 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model (F = 22.906, P < 0.001) with DLAB as the only predictor in the model. The standardized β coefficient for DLAB in the post hoc model is 0.308, and the model explains 9.5% of the variance.

MK has a significant impact on the model for the Category III languages. The model with all five predictors is significant (F = 2.255, P = 0.048), but explains only 3.0% of the variance. The standardized β coefficients for the predictors in the model are -0.006, 0.079, 0.051, 0.000, and 0.125 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model (F = 7.908, P = 0.005) with DLAB as the only predictor. The model explains 2.1% of the variance. The standardized β coefficient for DLAB in the post hoc model is 0.144.

For the Category IV languages, AR has a highly significant impact on the model. The model with all five predictors accounts for 5.1% of the variance and is highly significant (F = 4.206, P = 0.001). The standardized β coefficients are 0.145, 0.057, -0.111, 0.036, and 0.122 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model (F = 8.479, P < 0.001) with AR and DLAB as the only two predictors retained in the model. The post hoc model explains 4.1% of the variance, and the standardized β coefficients for AR and DLAB are 0.124 and 0.132, respectively.

Reading DLPT for DLI Graduates

The Reading DLPT analysis by language category uses the same linear regression method as above with hierarchical entry of predictor variables. Figure 10 is divided by language category and shows the models for predicting Reading DLPT 1.

Figure 10. Reading DLPT 1 Predictive Models for DLI Graduates by Language Category.

Reading DLPT 1 for CAT I DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.032	13.049	0.000
2	AR, MK	0.041	3.679	0.026
3	AR, MK, PC	0.052	2.274	0.080
4	AR, MK, PC, WK	0.064	1.671	0.156
5	AR, MK, PC, WK, DLAB	0.097	3.563	0.004

Reading DLPT 1 for CAT II DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.000	0.001	0.969
2	AR, MK	0.027	6.022	0.003
3	AR, MK, PC	0.027	0.000	1.000
4	AR, MK, PC, WK	0.028	0.074	0.990
5	AR, MK, PC, WK, DLAB	0.085	3.348	0.006

Reading DLPT 1 for CAT III DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.008	3.061	0.081
2	AR, MK	0.022	5.339	0.005
3	AR, MK, PC	0.025	0.574	0.632
4	AR, MK, PC, WK	0.032	0.897	0.466
5	AR, MK, PC, WK, DLAB	0.085	5.372	0.000

Reading DLPT 1 for CAT IV DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.017	6.826	0.009
2	AR, MK	0.048	12.862	0.000
3	AR, MK, PC	0.052	0.833	0.476
4	AR, MK, PC, WK	0.057	0.696	0.595
5	AR, MK, PC, WK, DLAB	0.079	2.347	0.041

The models for Category I languages show that Arithmetic Reasoning (AR) and DLAB are highly significant predictors of DLPT 1 reading scores, and MK also has a significant impact on the scores. The model with all five predictors is highly significant ($F = 8.383$, $P < 0.001$), explaining 9.7% of the variance. The standardized β coefficients for the model are 0.019, 0.079, 0.042, 0.136, and 0.193 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 19.009$, $P < 0.001$) with WK and DLAB as the predictors in the model. The post hoc model explains 8.8% of the variance, and the standardized β coefficients for WK and DLAB are 0.173 and 0.221.

For Category II languages, MK and DLAB have highly significant effects on the listening models. The model with all five predictors is also highly significant ($F = 3.956$, $P = 0.002$) and explains 8.5% of the variance. The standardized β coefficients are -0.109, 0.109, -0.055, 0.080, and 0.255 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 16.191$, $P < 0.001$) with DLAB as the only predictor in the model. The standardized β coefficient for DLAB in the post hoc model is 0.263, and the model explains 6.9% of the variance.

MK and DLAB are highly significant predictors of the DLPT 1 reading scores for the Category III languages. The model with all five predictors is also highly significant ($F = 6.853$, $P < 0.001$) and explains 8.5% of the variance. The standardized β coefficients for the predictors in the model are -0.048, 0.099, 0.018, 0.107, and 0.235 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F =$

15.554, P < 0.001) with WK and DLAB as the predictors. The model explains 7.7% of the variance. The standardized β coefficients for WK and DLAB in the post hoc model are 0.107 and 0.250, respectively.

For the Category IV languages, AR and MK are highly significant predictors in the model, and DLAB is a significant predictor. The model with all five predictors accounts for 7.9% of the variance and is highly significant (F = 6.706, P < 0.001). The standardized β coefficients are 0.027, 0.157, 0.136, -0.132, and 0.157 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model (F = 14.117, P < 0.001) with MK and DLAB as the only two predictors retained in the model. The post hoc model explains 6.7% of the variance, and the standardized β coefficients for MK and DLAB are 0.166 and 0.155, respectively.

Figure 11, below, is divided by language category and shows the models for predicting Reading DLPT 2 scores. Arithmetic Reasoning (AR) is once again a significant predictor of Reading DLPT 2 scores for the Category I languages. With all five predictors in the Category I model, the model is highly significant (F = 3.664, P = 0.003) and explains 4.5% of the variance. The standardized β coefficients are -0.014, 0.077, 0.042, 0.112, and 0.106 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model (F = 7.930, P 0.008) where WK and DLAB are the predictors. The post hoc model explains 3.9% of the variance, and the standardized β coefficients for WK and DLAB are 0.134 and 0.129, respectively.

Figure 11. Reading DLPT 2 Predictive Models for DLI Graduates by Language Category.

Reading DLPT 2 for CAT I DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.014	5.430	0.020
2	AR, MK	0.02	2.400	0.092
3	AR, MK, PC	0.027	1.410	0.239
4	AR, MK, PC, WK	0.035	1.080	0.366
5	AR, MK, PC, WK, DLAB	0.045	1.021	0.405

Reading DLPT 2 for CAT II DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0	0.001	0.976
2	AR, MK	0.006	1.310	0.358
3	AR, MK, PC	0.007	0.109	0.955
4	AR, MK, PC, WK	0.008	0.073	0.990
5	AR, MK, PC, WK, DLAB	0.049	2.317	0.045

Reading DLPT 2 for CAT III DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.003	1.028	0.311
2	AR, MK	0.013	3.779	0.024
3	AR, MK, PC	0.019	1.141	0.332
4	AR, MK, PC, WK	0.025	0.763	0.550
5	AR, MK, PC, WK, DLAB	0.036	1.058	0.383

Reading DLPT 2 for CAT IV DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.016	6.333	0.012
2	AR, MK	0.029	5.288	0.005
3	AR, MK, PC	0.029	0.000	1.000
4	AR, MK, PC, WK	0.034	0.680	0.565
5	AR, MK, PC, WK, DLAB	0.093	6.391	0.000

For Category II languages, DLAB is a significant predictor of the reading scores. The model with all five predictors is only marginally significant (F = 2.191, P = 0.056) and explains 4.9% of the variance. The standardized β coefficients are -0.076, 0.023, -0.032, 0.074, and 0.218 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model (F = 9.831, P = 0.002) where DLAB is the only predictor in the model. Here, the standardized β coefficient for DLAB is 0.208. The post hoc model explains 4.3% of the variance.

For the Category III languages, MK has a significant impact on the model. A model with all five predictors is significant ($F = 2.734$, $P = 0.019$) but explains only 3.6% of the variance. The standardized β coefficients for the predictors in the model are -0.078, 0.107, 0.049, 0.094, and 0.107 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 5.987$, $P = 0.015$) that retains only DLAB as a predictor and explains 1.6% of the variance. The standardized β coefficient for DLAB in the post hoc model is 0.126.

Finally, for the Category IV languages, AR, MK and DLAB have a significant impact on the model, with the impact of MK and DLAB being highly significant. The model with all five predictors accounts for 9.3% of the variance and is highly significant ($F = 8.040$, $P < 0.001$). The standardized β coefficients are 0.079, 0.064, 0.054, -0.139, and 0.260 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 32.566$, $P < 0.001$) where DLAB is the only predictor in the model. The post hoc model explains 7.6% of the variance, and the standardized β coefficient for DLAB is 0.276.

The results for the third reading DLPT are shown in Figure 12. Arithmetic Reasoning (AR) and Math Knowledge are highly significant predictors of Reading DLPT 3 scores for the Category I languages. DLAB is also a significant predictor of the reading scores. With all five predictors in the Category I model, the model is highly significant ($F = 4.694$, $P < 0.001$) and explains 5.7% of the variance. The standardized β coefficients are 0.021, 0.094, 0.017, 0.068, and 0.152 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis

Figure 12. Reading DLPT 3 Predictive Models for DLI Graduates by Language Category.

Reading DLPT 3 for CAT I DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.018	7.122	0.008
2	AR, MK	0.03	4.850	0.008
3	AR, MK, PC	0.033	0.608	0.610
4	AR, MK, PC, WK	0.036	0.406	0.804
5	AR, MK, PC, WK, DLAB	0.057	2.171	0.057

Reading DLPT 3 for CAT II DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.006	1.313	0.253
2	AR, MK	0.022	3.550	0.030
3	AR, MK, PC	0.024	0.222	0.881
4	AR, MK, PC, WK	0.029	0.371	0.829
5	AR, MK, PC, WK, DLAB	0.057	1.596	0.162

Reading DLPT 3 for CAT III DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.012	4.488	0.035
2	AR, MK	0.016	1.516	0.221
3	AR, MK, PC	0.022	1.144	0.331
4	AR, MK, PC, WK	0.029	0.894	0.468
5	AR, MK, PC, WK, DLAB	0.046	1.653	0.145

Reading DLPT 3 for CAT IV DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.011	4.331	0.038
2	AR, MK	0.027	6.495	0.002
3	AR, MK, PC	0.029	0.407	0.748
4	AR, MK, PC, WK	0.029	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.067	4.002	0.002

retains MK and DLAB as predictors. The post hoc model is highly significant ($F = 10.071$, $P < 0.001$) and explains 4.9% of the variance. The standardized β coefficients for MK and DLAB are 0.109 and 0.164, respectively.

For Category II languages, MK is a significant predictor of the reading DLPT 3 scores. The model with all five predictors is also significant ($F = 2.569$, $P = 0.028$) and explains 5.7% of the variance. The standardized β coefficients are -0.150, 0.093,

-0.122, 0.121, and 0.179 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a significant predictive model ($F = 6.007$, $P = 0.015$) with DLAB as the only predictor in the model. The standardized β coefficient for DLAB in the post hoc model is 0.164, and the model explains 2.7% of the variance.

The only predictor to have a significant impact on the model for the Category III languages is AR. The model with all five predictors is highly significant ($F = 3.576$, $P = 0.004$) and explains 4.6% of the variance. The standardized β coefficients for the predictors in the model are -0.010, 0.063, 0.047, 0.103, and 0.133 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 7.979$, $P < 0.001$) with WK and DLAB as the retained predictors. The model explains 4.1% of the variance. The standardized β coefficients for WK and DLAB in the post hoc model are 0.132 and 0.146, respectively.

For the Category IV languages, AR has a significant impact and MK and DLAB have a highly significant impact on the model. The model with all five predictors accounts for 6.7% of the variance and is highly significant ($F = 5.660$, $P < 0.001$). The standardized β coefficients are 0.060, 0.086, -0.033, -0.048, and 0.211 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 22.896$, $P < 0.001$) where DLAB is the only predictor in the model. The post hoc model explains 5.5% of the variance, and the standardized β coefficient for DLAB in the model is 0.234.

Figure 13. Reading DLPT 4 Predictive Models for DLI Graduates by Language Category.

Reading DLPT 4 for CAT I DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.022	8.869	0.003
2	AR, MK	0.029	2.826	0.061
3	AR, MK, PC	0.036	1.423	0.233
4	AR, MK, PC, WK	0.038	0.271	0.897
5	AR, MK, PC, WK, DLAB	0.043	0.509	0.770

Reading DLPT 4 for CAT II DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.007	1.456	0.229
2	AR, MK	0.03	5.145	0.007
3	AR, MK, PC	0.03	0.000	1.000
4	AR, MK, PC, WK	0.033	0.223	0.925
5	AR, MK, PC, WK, DLAB	0.047	0.790	0.558

Reading DLPT 4 for CAT III DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.006	2.172	0.141
2	AR, MK	0.023	6.490	0.002
3	AR, MK, PC	0.036	2.515	0.058
4	AR, MK, PC, WK	0.038	0.258	0.905
5	AR, MK, PC, WK, DLAB	0.042	0.387	0.858

Reading DLPT 4 for CAT IV DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.028	11.285	0.001
2	AR, MK	0.031	1.222	0.296
3	AR, MK, PC	0.036	1.024	0.382
4	AR, MK, PC, WK	0.036	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.051	1.553	0.173

For the fourth reading DLPT, Figure 13 demonstrates that Arithmetic Reasoning (AR) is a highly significant predictor of Reading DLPT 4 scores for the Category I languages. The model with all five predictors is also highly significant ($F = 3.488$, $P = 0.004$), explaining only 4.3% of the variance. The standardized β coefficients for the model are 0.032, 0.087, 0.064, 0.062, and 0.072 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 8.869$, $P = 0.003$) with AR as the only predictor in the model. The post hoc model explains 2.2% of the variance, and the standardized β coefficient for AR is 0.149.

For Category II languages, MK is a highly significant predictor of the reading scores, but the model with all five predictors is only marginally significant ($F = 2.091$, $P = 0.068$). The model explains 4.7% of the variance. The standardized β coefficients are -0.197, 0.131, -0.025, 0.084, and 0.127 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis fails to yield a significant predictive model, although entering both AR and MK into the model does produce a significant model ($F = 3.348$, $P = 0.037$) due to opposite polarity of the coefficients. This model explains 3.0% of the variance and the standardized β coefficients are -0.162 and 0.173 for AR and MK, respectively.

MK has a highly significant impact on the model for the Category III languages. The model with all five predictors is highly significant ($F = 3.244$, $P = 0.007$), and explains 4.2% of the variance. The standardized β coefficients for the predictors in the model are -0.077, 0.148, 0.105, 0.061, and 0.064 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 6.535$, $P = 0.002$) with MK and PC as the two predictors. The model explains 3.4% of the variance. The standardized β coefficients for MK and PC are 0.137 and 0.107, respectively.

For the Category IV languages, AR has a highly significant impact on the model. The model with all five predictors accounts for 5.1% of the variance and is highly significant ($F = 4.187$, $P = 0.001$). The standardized β coefficients are 0.179, 0.026, -0.086, -0.018, and 0.129 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 8.883$, $P < 0.001$) with AR and DLAB as the only two predictors retained in the model. The post hoc model explains 4.3% of the variance, and the standardized β coefficients for AR and DLAB are 0.134 and 0.128, respectively.

Oral Proficiency Interview for DLI Graduates

This section examines predictive models for OPI performance for the DLI graduates sorted by language category. The results of the analysis for the OPI are displayed in Figure 14. The models for Category I languages show that Arithmetic Reasoning (AR) is the only predictor that reaches even marginal significance. The model with all five predictors is not significant and only explains 1.8% of the variance. Meanwhile, the models for Category II languages do not have any significant predictors, nor are any of the models significant.

Figure 14. OPI Predictive Models for DLI Graduates by Language Category.

OPI for CAT I DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.008	3.072	0.080
2	AR, MK	0.009	0.396	0.674
3	AR, MK, PC	0.012	0.595	0.619
4	AR, MK, PC, WK	0.018	0.796	0.528
5	AR, MK, PC, WK, DLAB	0.018	0.000	1.000

OPI for CAT II DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.000	0.000	0.989
2	AR, MK	0.005	1.090	0.338
3	AR, MK, PC	0.013	0.879	0.453
4	AR, MK, PC, WK	0.023	0.737	0.568
5	AR, MK, PC, WK, DLAB	0.032	0.500	0.776

OPI for CAT III DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.383	0.536
2	AR, MK	0.001	0.000	1.000
3	AR, MK, PC	0.003	0.374	0.772
4	AR, MK, PC, WK	0.010	0.877	0.478
5	AR, MK, PC, WK, DLAB	0.041	2.998	0.012

OPI for CAT IV DLI Graduates

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	0.803	0.371
2	AR, MK	0.003	0.396	0.673
3	AR, MK, PC	0.003	0.000	1.000
4	AR, MK, PC, WK	0.004	0.132	0.971
5	AR, MK, PC, WK, DLAB	0.033	2.946	0.013

For Category III languages, on the other hand, DLAB is a significant predictor of OPI scores for participants. The model with all five predictors is highly significant ($F = 3.154$, $P = 0.008$) and explains 4.1% of the variance. The standardized β coefficients for the predictors in the model are 0.070, -0.037, -0.009, -0.102, and 0.180 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 12.127$, $P = 0.001$) with DLAB as only predictor in the model. The model explains 3.1% of the variance. The standardized β coefficient for DLAB in the post hoc model is 0.177.

For the Category IV languages, DLAB is also a significant predictor. The model with all five predictors accounts for 3.3% of the variance and is significant ($F = 2.714$, $P = 0.020$). The standardized β coefficients are 0.011, -0.008, 0.036, -0.054, and 0.185 for AR, MK, PC, WK, and DLAB, respectively. A post hoc stepwise linear regression analysis yields a highly significant predictive model ($F = 13.069$, $P < 0.001$) with DLAB as the only predictor retained in the model. The post hoc model explains 3.2% of the variance, and the standardized β coefficient for DLAB is 0.179.

Listening Proficiency Growth for DLI Graduates

This analysis examines predictors of listening proficiency growth by language category using logistic regression with hierarchical entry of predictor variables. The outcome measure is assigned a value of “1” for a participant if the ILR scale score for that participant on the DLPT increases over the one-year period being analyzed. If the score does not increase a value of “0” is assigned. The significance of each predictor is determined when it was added to the model by looking at the change in the χ^2 term. Only participants with four consecutive annual DLPT scores are included in the analysis. The descriptive statistics for the predictor variables by language are shown in table 62, below.

Table 62. Descriptive statistics by language for predictors of language proficiency growth.

		AR		MK		PC		WK		DLAB	
	N	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
French	133	57.36	7.68	60.76	5.11	56.61	5.91	56.41	6.57	103.0	12.0
Spanish	170	58.76	6.03	60.78	5.33	57.64	5.29	57.79	4.79	103.4	10.5
German	82	58.68	6.21	61.37	4.78	57.30	3.90	57.43	5.64	106.6	14.9
Indonesian	41	60.68	6.18	61.49	6.27	59.63	6.47	58.46	7.63	112.1	11.9
Russian	159	61.31	5.19	62.77	4.45	59.04	3.79	59.14	4.21	109.6	11.5
Tagalog	155	58.81	7.32	61.38	5.68	57.85	5.92	56.95	6.69	106.6	10.6
Arabic	149	61.22	5.58	62.9	4.72	59.52	4.46	59.56	5.71	118.4	11.9
Chinese	173	61.57	5.65	63.34	4.13	59.43	4.98	59.61	5.49	119.5	13.2

The first growth period analyzed is the one-year period immediately following graduation from the Defense Language Institute. As seen earlier in Table 47 (Mean Listening DLPT scores for DLI Graduates by Language Group), the average listening DLPT scores for all language groups decreased during that time period. Figure 15 displays the predictor models for listening growth by language category for that time period. The only significant predictors in any of the models are Paragraph Comprehension (PC) and Word Knowledge (WK) in the category IV language models. Interestingly, PC is a negative predictor of growth since the exponent (β) for PC is 0.884, meaning the higher the PC score the less likely the participant is to grow in listening proficiency. WK, on the other hand, is a true predictor of listening proficiency growth. None of the listening growth models for this period are significant, and only the category IV language model is marginally significant (χ^2 = 9.795, P = 0.081).

Figure 15. Listening Growth Period 1 Models for DLI Graduates by Language Category.

Listening Growth 1 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.007	2.048	2.048	0.152
2	AR, MK	0.009	2.711	0.663	0.416
3	AR, MK, PC	0.009	2.757	0.046	0.830
4	AR, MK, PC, WK	0.009	2.825	0.068	0.794
5	AR, MK, PC, WK, DLAB	0.010	2.973	0.148	0.701

Listening Growth 1 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.001	0.098	0.098	0.755
2	AR, MK	0.002	0.283	0.185	0.667
3	AR, MK, PC	0.011	1.359	1.076	0.300
4	AR, MK, PC, WK	0.012	1.539	0.180	0.671
5	AR, MK, PC, WK, DLAB	0.021	2.559	1.020	0.313

Listening Growth 1 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.003	0.788	0.788	0.375
2	AR, MK	0.003	1.046	0.258	0.612
3	AR, MK, PC	0.004	1.210	0.164	0.686
4	AR, MK, PC, WK	0.008	2.538	1.328	0.249
5	AR, MK, PC, WK, DLAB	0.011	3.533	0.995	0.319

Listening Growth 1 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.002	0.547	0.547	0.460
2	AR, MK	0.002	0.564	0.017	0.896
3	AR, MK, PC	0.014	4.544	3.980	0.046
4	AR, MK, PC, WK	0.030	9.762	5.218	0.022
5	AR, MK, PC, WK, DLAB	0.030	9.795	0.033	0.856

The next set of models shown in Figure 16 is for the second annual growth period. During this period, groups in language categories I, II, and III showed growth in their average listening DLPT scores, albeit very limited. The language category IV group had a slight decline in its average scores. The only significant predictor of growth for this time period is the DLAB score for the language category I model, but it is a negative predictor for growth. It is highly significant, and its addition to the model makes the listening growth model for category I languages become significant

(χ^2 = 12.644, P = 0.027). Unfortunately, increased DLAB scores mean that participants are less likely to grow in listening proficiency during this time frame. A possible explanation for this outcome can be found in the discussion section of this chapter. None of the other language category growth models are significant.

Figure 16. Listening Growth Period 2 Models for DLI Graduates by Language Category.

Listening Growth 2 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.003	0.784	0.784	0.376
2	AR, MK	0.008	2.475	1.691	0.194
3	AR, MK, PC	0.010	3.182	0.707	0.400
4	AR, MK, PC, WK	0.011	3.202	0.020	0.888
5	AR, MK, PC, WK, DLAB	0.041	12.644	9.442	0.002

Listening Growth 2 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.001	0.083	0.083	0.773
2	AR, MK	0.002	0.214	0.131	0.717
3	AR, MK, PC	0.006	0.718	0.504	0.478
4	AR, MK, PC, WK	0.011	1.359	0.641	0.423
5	AR, MK, PC, WK, DLAB	0.011	1.407	0.048	0.827

Listening Growth 2 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.007	2.100	2.100	0.147
2	AR, MK	0.007	2.156	0.056	0.813
3	AR, MK, PC	0.014	4.411	2.255	0.133
4	AR, MK, PC, WK	0.022	7.094	2.683	0.101
5	AR, MK, PC, WK, DLAB	0.023	7.17	0.076	0.783

Listening Growth 2 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.003	1.027	1.027	0.311
2	AR, MK	0.007	2.186	1.159	0.282
3	AR, MK, PC	0.007	2.193	0.007	0.933
4	AR, MK, PC, WK	0.007	2.227	0.034	0.854
5	AR, MK, PC, WK, DLAB	0.008	2.428	0.201	0.654

For the third annual growth period, the average listening DLPT scores for all language groups remained fairly constant. Figure 17 shows the growth models for this period. None of the predictors are significant, and only Paragraph Comprehension

(PC) in the language category IV growth model is marginally significant. None of the growth models for any of the language categories is significant.

Figure 17. Listening Growth Period 3 Models for DLI Graduates by Language Category.

Listening Growth 3 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.037	0.037	0.847
2	AR, MK	0.001	0.292	0.255	0.614
3	AR, MK, PC	0.001	0.295	0.003	0.956
4	AR, MK, PC, WK	0.004	1.238	0.943	0.332
5	AR, MK, PC, WK, DLAB	0.004	1.275	0.037	0.848

Listening Growth 3 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.015	1.824	1.824	0.177
2	AR, MK	0.027	3.391	1.567	0.211
3	AR, MK, PC	0.030	3.737	0.346	0.556
4	AR, MK, PC, WK	0.031	3.849	0.112	0.738
5	AR, MK, PC, WK, DLAB	0.042	5.281	1.432	0.231

Listening Growth 3 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.001	0.429	0.429	0.512
2	AR, MK	0.008	2.587	2.158	0.142
3	AR, MK, PC	0.008	2.608	0.021	0.885
4	AR, MK, PC, WK	0.010	3.249	0.641	0.423
5	AR, MK, PC, WK, DLAB	0.01	3.289	0.040	0.842

Listening Growth 3 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.018	0.018	0.894
2	AR, MK	0.000	0.122	0.104	0.747
3	AR, MK, PC	0.010	3.381	3.259	0.071
4	AR, MK, PC, WK	0.017	5.429	2.048	0.152
5	AR, MK, PC, WK, DLAB	0.018	5.811	0.382	0.537

Reading Proficiency Growth for DLI Graduates

This analysis examines reading proficiency growth by language category using the same method as above. It also uses logistic regression with hierarchical entry of predictor variables. Once again, the significance of each predictor is determined when it was added to the model by looking at the change in the χ^2 term.

The first growth period during the year immediately following graduation from the Defense Language Institute shows a decline in the average reading DLPT scores across all language groups (see Table 52). Figure 18 displays the predictive models for reading growth for that time period. The significant predictors include DLAB scores for category I and III languages and Math Knowledge (MK) for category II languages. MK is also marginally significant in the category IV model, and WK is marginally significant in the category I model. All of the significant predictors, however, are negative predictors. This will be addressed in the discussion section of this chapter. The only model that is significant is the category I language model ($\chi^2 = 11.105$, P = 0.049).

Figure 18. Reading Growth Period 1 Models for DLI Graduates by Language Category.

Reading Growth 1 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.008	2.475	2.475	0.116
2	AR, MK	0.009	2.614	0.139	0.709
3	AR, MK, PC	0.012	3.521	0.907	0.341
4	AR, MK, PC, WK	0.021	6.532	3.011	0.083
5	AR, MK, PC, WK, DLAB	0.036	11.105	4.573	0.033

Reading Growth 1 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.007	0.861	0.861	0.354
2	AR, MK	0.049	6.120	5.259	0.022
3	AR, MK, PC	0.049	6.219	0.099	0.753
4	AR, MK, PC, WK	0.051	6.395	0.176	0.675
5	AR, MK, PC, WK, DLAB	0.059	7.505	1.110	0.292

Reading Growth 1 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.157	0.157	0.692
2	AR, MK	0.001	0.240	0.083	0.773
3	AR, MK, PC	0.006	1.816	1.576	0.209
4	AR, MK, PC, WK	0.006	1.870	0.054	0.816
5	AR, MK, PC, WK, DLAB	0.023	7.439	5.569	0.018

Reading Growth 1 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.001	0.300	0.300	0.584
2	AR, MK	0.010	3.319	3.019	0.082
3	AR, MK, PC	0.013	4.277	0.958	0.328
4	AR, MK, PC, WK	0.016	5.205	0.928	0.335
5	AR, MK, PC, WK, DLAB	0.020	6.344	1.139	0.286

Figure 19 displays the next set of models for the second annual growth period. During this period, groups in language categories I, II, and III showed limited growth in their average reading DLPT scores. The language category IV group showed no growth. None of the predictors in this set of models is significant. Only Arithmetic Reasoning (AR) is a marginally significant predictor of category III language reading proficiency growth. None of the reading growth models is significant.

Figure 19. Reading Growth Period 2 Models for DLI Graduates by Language Category.

Reading Growth 2 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model $\chi 2$	$\Delta\chi 2$	P of $\Delta\chi 2$
1	AR	0.002	0.622	0.622	0.430
2	AR, MK	0.002	0.641	0.019	0.890
3	AR, MK, PC	0.002	0.753	0.112	0.738
4	AR, MK, PC, WK	0.003	0.933	0.180	0.671
5	AR, MK, PC, WK, DLAB	0.007	1.996	1.063	0.303

Reading Growth 2 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model $\chi 2$	$\Delta\chi 2$	P of $\Delta\chi 2$
1	AR	0.000	0.010	0.010	0.921
2	AR, MK	0.004	0.555	0.545	0.460
3	AR, MK, PC	0.010	1.258	0.703	0.402
4	AR, MK, PC, WK	0.013	1.584	0.326	0.568
5	AR, MK, PC, WK, DLAB	0.032	4.006	2.422	0.120

Reading Growth 2 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model $\chi 2$	$\Delta\chi 2$	P of $\Delta\chi 2$
1	AR	0.011	3.440	3.440	0.064
2	AR, MK	0.011	3.458	0.018	0.893
3	AR, MK, PC	0.014	4.526	1.068	0.301
4	AR, MK, PC, WK	0.017	5.316	0.790	0.374
5	AR, MK, PC, WK, DLAB	0.019	5.896	0.580	0.446

Reading Growth 2 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model $\chi 2$	$\Delta\chi 2$	P of $\Delta\chi 2$
1	AR	0.002	0.558	0.558	0.455
2	AR, MK	0.004	1.308	0.750	0.387
3	AR, MK, PC	0.004	1.308	0.000	1.000
4	AR, MK, PC, WK	0.010	3.194	1.886	0.170
5	AR, MK, PC, WK, DLAB	0.012	3.793	0.599	0.439

For the third annual growth period, the average reading DLPT scores for all language groups remained nearly constant. Figure 20 shows the growth models for this period. DLAB in the category I model is the only significant predictor in this set. DLAB is also marginally significant in predicting reading growth during this period in the category IV model. AR is marginally significant in the category II model. All of the significant and marginally significant predictors are negative predictors of reading growth during this period. Only the language category II model significantly predicts reading growth during the third year after graduation from DLI.

Figure 20. Reading Growth Period 3 Models for DLI Graduates by Language Category.

Reading Growth 3 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.060	0.060	0.806
2	AR, MK	0.000	0.104	0.044	0.834
3	AR, MK, PC	0.003	0.810	0.706	0.401
4	AR, MK, PC, WK	0.010	2.905	2.095	0.148
5	AR, MK, PC, WK, DLAB	0.024	7.423	4.518	0.034

Reading Growth 3 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.028	3.535	3.535	0.060
2	AR, MK	0.032	4.014	0.479	0.489
3	AR, MK, PC	0.065	8.298	4.284	0.039
4	AR, MK, PC, WK	0.084	10.813	2.515	0.113
5	AR, MK, PC, WK, DLAB	0.086	11.112	0.299	0.585

Reading Growth 3 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.002	0.555	0.555	0.456
2	AR, MK	0.006	1.938	1.383	0.240
3	AR, MK, PC	0.007	2.085	0.147	0.701
4	AR, MK, PC, WK	0.007	2.152	0.067	0.796
5	AR, MK, PC, WK, DLAB	0.007	2.218	0.066	0.797

Reading Growth 3 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.006	1.935	1.935	0.164
2	AR, MK	0.007	2.253	0.318	0.573
3	AR, MK, PC	0.007	2.396	0.143	0.705
4	AR, MK, PC, WK	0.008	2.644	0.248	0.619
5	AR, MK, PC, WK, DLAB	0.020	6.421	3.777	0.052

Listening Proficiency Attrition for DLI Graduates

Studies such as Surface et al. (2004) were concerned with the ability of service members to retain language skills over time, post-DLI graduation, which indicates that attrition of language skill may be more common than continued growth. And, as noted above, the listening DLPT scores for all language groups decreased over the one-year period after DLI graduation. For this reason, an analysis of language attrition is warranted. Therefore, this analysis examines predictors of listening proficiency attrition by language category using logistic regression with hierarchical entry of predictor variables. The outcome measure is assigned a value of "1" for a participant if the ILR scale score for that participant on the DLPT decreases over the one-year period being analyzed. If the score does not decrease a value of "0" is assigned. The significance of each predictor is determined when it was added to the model by looking at the change in the χ^2 term. Once again, only participants with four consecutive annual DLPT scores are included in the analysis.

In an analysis like this, independent variables that predict a "0" outcome are of more interest since those are the factors that impede attrition. Figure 21 shows the models for each language category that predict attrition. This analysis will pay close attention to the exponent (β) for the significant predictors. When the exponent (β) is less than 1, it indicates that the greater the score on the independent variable the more it impedes an outcome of "1" for the dependent variable. The only significant predictor for this first set of models is Paragraph Comprehension (PC) for the category IV languages. Math Knowledge (MK) is marginally significant for both the category II and category IV languages.

Figure 21. Listening Attrition Period 1 Models for DLI Graduates by Language Category.

Listening Attrition 1 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.075	0.075	0.784
2	AR, MK	0.000	0.087	0.012	0.913
3	AR, MK, PC	0.003	1.022	0.935	0.334
4	AR, MK, PC, WK	0.004	1.220	0.198	0.656
5	AR, MK, PC, WK, DLAB	0.008	2.458	1.238	0.266

Listening Attrition 1 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.008	1.043	1.043	0.307
2	AR, MK	0.030	3.797	2.754	0.097
3	AR, MK, PC	0.041	5.096	1.299	0.254
4	AR, MK, PC, WK	0.053	6.710	1.614	0.204
5	AR, MK, PC, WK, DLAB	0.072	9.151	2.441	0.118

Listening Attrition 1 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.000	0.000	0.983
2	AR, MK	0.000	0.016	0.016	0.899
3	AR, MK, PC	0.000	0.030	0.014	0.906
4	AR, MK, PC, WK	0.000	0.061	0.031	0.806
5	AR, MK, PC, WK, DLAB	0.000	0.061	0.000	1.000

Listening Attrition 1 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.003	0.964	0.964	0.326
2	AR, MK	0.013	4.153	3.189	0.074
3	AR, MK, PC	0.029	9.366	5.213	0.022
4	AR, MK, PC, WK	0.030	9.659	0.293	0.588
5	AR, MK, PC, WK, DLAB	0.030	9.841	0.182	0.670

For the category II language model, the exponent (β) for MK is 0.931 meaning that higher MK scores predict lower attrition rates. The number of cases for this model is only 123, so MK as a predictor only reaches marginal significance. For the category IV language model, MK is once again marginally significant with an exponent (β) of 0.948. PC, on the other hand, does reach significance, but the exponent (β) for PC is 1.074 indicating that higher scores on the Paragraph Comprehension portion of the ASVAB predicts higher rates of attrition. The number of cases for the category IV language model is 322.

Figure 22 shows the models for the second-year period after graduating from DLI. The category III language model demonstrates that higher scores on Word Knowledge (in L1 English) significantly predicts lower attrition rates of L2 listening proficiency since the exponent (β) for WK is 0.938. For the category IV language model, Arithmetic Reasoning is a significant predictor of lower L2 listening proficiency attrition rates, and Math Knowledge reaches marginal significance at predicting lower attrition rates.

Figure 22. Listening Attrition Period 2 Models for DLI Graduates by Language Category.

Listening Attrition 2 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.006	0.006	0.940
2	AR, MK	0.002	0.524	0.518	0.472
3	AR, MK, PC	0.003	0.858	0.334	0.563
4	AR, MK, PC, WK	0.004	1.189	0.331	0.565
5	AR, MK, PC, WK, DLAB	0.004	1.239	0.050	0.823

Listening Attrition 2 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.002	0.292	0.292	0.589
2	AR, MK	0.008	0.937	0.645	0.422
3	AR, MK, PC	0.008	0.968	0.031	0.860
4	AR, MK, PC, WK	0.008	0.995	0.027	0.870
5	AR, MK, PC, WK, DLAB	0.010	1.195	0.200	0.655

Listening Attrition 2 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.050	0.050	0.823
2	AR, MK	0.000	0.051	0.001	0.975
3	AR, MK, PC	0.000	0.146	0.095	0.758
4	AR, MK, PC, WK	0.014	4.324	4.178	0.041
5	AR, MK, PC, WK, DLAB	0.014	4.327	0.003	0.956

Listening Attrition 2 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.012	4.099	4.099	0.047
2	AR, MK	0.021	6.918	2.819	0.093
3	AR, MK, PC	0.028	9.205	2.287	0.131
4	AR, MK, PC, WK	0.030	9.767	0.562	0.454
5	AR, MK, PC, WK, DLAB	0.032	10.408	0.641	0.423

For the third-year period after graduating from DLI, only Arithmetic Reasoning reaches marginal significance at predicting lower attrition rates for the category IV language model (see Figure 23, below). The exponent (β) for AR in the model is 0.965. No other models or predictors are significant in this set.

Figure 23. Listening Attrition Period 3 Models for DLI Graduates by Language Category.

Listening Attrition 3 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.017	0.017	0.896
2	AR, MK	0.003	0.959	0.942	0.332
3	AR, MK, PC	0.003	1.011	0.052	0.820
4	AR, MK, PC, WK	0.007	2.261	1.250	0.264
5	AR, MK, PC, WK, DLAB	0.012	3.705	1.444	0.230

Listening Attrition 3 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.001	0.114	0.114	0.736
2	AR, MK	0.007	0.912	0.798	0.372
3	AR, MK, PC	0.008	0.970	0.058	0.810
4	AR, MK, PC, WK	0.008	1.027	0.057	0.811
5	AR, MK, PC, WK, DLAB	0.008	1.045	0.018	0.893

Listening Attrition 3 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.001	0.166	0.166	0.684
2	AR, MK	0.001	0.325	0.159	0.690
3	AR, MK, PC	0.006	1.793	1.468	0.226
4	AR, MK, PC, WK	0.006	1.945	0.152	0.697
5	AR, MK, PC, WK, DLAB	0.007	2.231	0.286	0.593

Listening Attrition 3 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.008	2.738	2.738	0.098
2	AR, MK	0.010	3.091	0.353	0.552
3	AR, MK, PC	0.010	3.132	0.041	0.840
4	AR, MK, PC, WK	0.014	4.481	1.349	0.246
5	AR, MK, PC, WK, DLAB	0.015	4.705	0.224	0.636

Reading Proficiency Attrition for DLI Graduates

This analysis examines predictors of reading proficiency attrition by language category using logistic regression with hierarchical entry of predictor variables, as above. Once again, the outcome measure is assigned a value of "1" for a participant if the ILR scale score for that participant on the DLPT decreases over the one-year period being analyzed. If the score does not decrease a value of "0" is assigned. The significance of each predictor is determined when it was added to the model by looking at the change in the χ^2 term. For the attrition analysis, independent variables that predict a "0" outcome are the factors that impede attrition.

Figure 24. Reading Attrition Period 1 Models for DLI Graduates by Language Category.

Reading Attrition 1 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.002	0.566	0.566	0.452
2	AR, MK	0.002	0.578	0.012	0.913
3	AR, MK, PC	0.002	0.581	0.003	0.956
4	AR, MK, PC, WK	0.002	0.603	0.022	0.882
5	AR, MK, PC, WK, DLAB	0.007	2.136	1.533	0.216

Reading Attrition 1 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.058	0.058	0.809
2	AR, MK	0.008	1.007	0.949	0.330
3	AR, MK, PC	0.017	2.166	1.159	0.282
4	AR, MK, PC, WK	0.040	5.036	2.870	0.090
5	AR, MK, PC, WK, DLAB	0.062	7.828	2.792	0.095

Reading Attrition 1 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.001	0.290	0.290	0.590
2	AR, MK	0.001	0.298	0.008	0.929
3	AR, MK, PC	0.005	1.498	1.200	0.273
4	AR, MK, PC, WK	0.005	1.500	0.002	0.964
5	AR, MK, PC, WK, DLAB	0.009	2.757	1.257	0.262

Reading Attrition 1 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.071	0.071	0.790
2	AR, MK	0.001	0.282	0.211	0.646
3	AR, MK, PC	0.007	2.119	1.837	0.175
4	AR, MK, PC, WK	0.014	4.409	2.290	0.130
5	AR, MK, PC, WK, DLAB	0.020	6.424	2.015	0.156

Figure 24 shows the models for each language category that predict attrition during the first year after participant graduation from the Defense Language Institute.

For the category II language model, the exponent (β) for WK and DLAB are 0.919 and 0.973 making them the only marginally significant predictors of lower L2 reading proficiency attrition rates for this set of models. None of the models in this set are significant.

Figure 25. Reading Attrition Period 2 Models for DLI Graduates by Language Category.

Reading Attrition 2 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.139	0.139	0.710
2	AR, MK	0.002	0.576	0.437	0.509
3	AR, MK, PC	0.003	0.810	0.234	0.629
4	AR, MK, PC, WK	0.022	6.595	5.785	0.016
5	AR, MK, PC, WK, DLAB	0.031	9.599	3.004	0.083

Reading Attrition 2 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.010	1.248	1.248	0.264
2	AR, MK	0.027	3.394	2.146	0.143
3	AR, MK, PC	0.035	4.381	0.987	0.321
4	AR, MK, PC, WK	0.035	4.401	0.020	0.888
5	AR, MK, PC, WK, DLAB	0.038	4.727	0.326	0.568

Reading Attrition 2 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.001	0.297	0.297	0.586
2	AR, MK	0.001	0.365	0.068	0.794
3	AR, MK, PC	0.005	1.578	1.213	0.271
4	AR, MK, PC, WK	0.006	1.984	0.406	0.524
5	AR, MK, PC, WK, DLAB	0.010	3.279	1.295	0.255

Reading Attrition 2 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.092	0.092	0.762
2	AR, MK	0.014	4.676	4.584	0.032
3	AR, MK, PC	0.015	4.883	0.207	0.649
4	AR, MK, PC, WK	0.015	4.992	0.109	0.741
5	AR, MK, PC, WK, DLAB	0.016	5.068	0.076	0.783

Figure 25 shows the models for the second-year period after graduating from DLI. Word Knowledge (in L1 English) scores in the category I language model significantly predict higher attrition rates of L2 reading proficiency since the exponent (β) for WK is 1.092. DLAB, on the other hand, marginally predicts lower attrition rates for L2 reading proficiency since its exponent (β) is 0.974. For the category IV language model, Math Knowledge is a significant predictor of lower L2 reading proficiency attrition rates, and its exponent (β) is 0.935.

Figure 26. Reading Attrition Period 3 Models for DLI Graduates by Language Category.

Reading Attrition 3 for Category I Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.000	0.125	0.125	0.724
2	AR, MK	0.001	0.315	0.190	0.663
3	AR, MK, PC	0.002	0.723	0.408	0.523
4	AR, MK, PC, WK	0.007	2.175	1.452	0.228
5	AR, MK, PC, WK, DLAB	0.008	2.409	0.234	0.629

Reading Attrition 3 for Category II Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.011	1.354	1.354	0.245
2	AR, MK	0.012	1.522	0.168	0.682
3	AR, MK, PC	0.013	1.592	0.070	0.791
4	AR, MK, PC, WK	0.020	2.423	0.831	0.362
5	AR, MK, PC, WK, DLAB	0.023	2.828	0.405	0.525

Reading Attrition 3 for Category III Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.014	4.315	4.315	0.038
2	AR, MK	0.015	4.702	0.387	0.534
3	AR, MK, PC	0.015	4.802	0.100	0.752
4	AR, MK, PC, WK	0.015	4.846	0.044	0.834
5	AR, MK, PC, WK, DLAB	0.029	9.147	4.301	0.038

Reading Attrition 3 for Category IV Languages

Model	Predictors	Model Cox & Snell R^2	Model χ^2	$\Delta\chi^2$	P of $\Delta\chi^2$
1	AR	0.011	3.580	3.580	0.058
2	AR, MK	0.011	3.583	0.003	0.956
3	AR, MK, PC	0.021	6.875	3.292	0.070
4	AR, MK, PC, WK	0.022	7.167	0.292	0.589
5	AR, MK, PC, WK, DLAB	0.023	7.344	0.177	0.674

For the third-year period after graduating from DLI, the significant predictors of attrition are Arithmetic Reasoning and DLAB for the category III language model (see Figure 26, above). The exponent (β) for AR in the model is 1.054, and the exponent (β) of DLAB is 1.027. So, both predict higher attrition rates for L2 reading proficiency for those participants that score higher on the ASVAB Arithmetic

Reasoning section and the DLAB. In the category IV language model, both AR and PC reach marginal significance. The exponent (β) of AR is 0.959 when added to the model meaning that it predicts a lower rate of L2 reading proficiency attrition. The exponent (β) of PC, however, is 1.060 and predicts higher attrition rates.

7.4.3 Individual Languages

In this section the participant data is sorted into groups by individual language. The two languages within each language category are analyzed side by side to evaluate intra-category patterns for predictor variable sets. All languages are analyzed using a similar procedure as used in the previous section. Outcome measures, once again, include foreign language GPA, listening and reading DLPT scores, and OPI scores. The analyses use linear regression with hierarchical entry of predictor variables.

Language Category I: French and Spanish

French and Spanish models are examined to evaluate consistency of predictor patterns for the various outcomes as representative languages for language category I. The significance of each predictor is determined when it is added to the model by looking at the change in the R^2 term. The predictors are added according to chronological progression of test scores, primarily; and general cognitive to more verbal specific aptitudes, secondarily. The descriptive statistics for the predictor variables and the DLI GPA outcome measure by language are shown in table 63, below. Standard deviations are shown in parenthesis. There are no significant differences between the means of the predictor variables or the outcome measure for the French and Spanish groups.

Table 63. Descriptive statistics for predictors of language proficiency for French and Spanish groups.

Language	N	AR	MK	PC	WK	DLAB	DLI GPA
French	197	57.7 (7.4)	60.5 (5.5)	57.0 (5.5)	56.7 (6.2)	104.2 (12.9)	3.27 (0.43)
Spanish	198	58.9 (5.8)	60.9 (5.3)	57.8 (5.2)	57.8 (4.7)	103.7 (10.7)	3.28 (0.43)

Figure 27, however, does show differences in the predictive models for foreign language GPA between the two language groups. For the French group, DLAB is a highly significant predictor of GPA, and AR is marginally significant. There are no significant predictors in the model for the Spanish learners. With all five predictors in the French model, the standardized β coefficients are 0.010, -0.018, -0.036, 0.079, and 0.421 for AR, MK, PC, WK, and DLAB, respectively. For the Spanish model the standardized β coefficients are 0.043, -0.035, -0.007, 0.063, and 0.131 for AR, MK, PC, WK, and DLAB, respectively. The Spanish model is not significant, but the overall magnitude of the β coefficients is similar between the two language models. The French model, however, weights DLAB scores four times higher than the Spanish model, and the Spanish model weights AR scores four times higher than the French model.

Figure 27. Predictive Models for Foreign Language Grade Point Average by Language for French and Spanish Learners.

Foreign Language GPA for French Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.019	3.755	0.054
2	AR, MK	0.026	1.394	0.251
3	AR, MK, PC	0.029	0.300	0.825
4	AR, MK, PC, WK	0.032	0.199	0.939
5	AR, MK, PC, WK, DLAB	0.182	8.802	0.000

Foreign Language GPA for Spanish Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.004	0.855	0.356
2	AR, MK	0.004	0.000	1.000
3	AR, MK, PC	0.005	0.098	0.961
4	AR, MK, PC, WK	0.009	0.261	0.903
5	AR, MK, PC, WK, DLAB	0.025	0.792	0.557

Listening DLPT scores over the three year period after graduating from DLI are very similar for the two language groups, as well (see Table 64), but trend in opposite directions. The French group trends towards listening proficiency attrition while the Spanish group trends towards growth. In fact, by the fourth listening DLPT, there is a highly significant difference between mean test scores for the groups on a two-tailed independent samples t-test ($t = -2.788$, $p = 0.006$).

Table 64. Descriptive statistics for listening DLPT scores for French and Spanish learners.

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
French	197	24.4 (4.5)	23.3 (5.3)	23.7 (5.1	23.4 (5.0
Spanish	198	23.8 (4.7)	23.7 (5.1)	24.4 (4.7)	24.7 (4.6)

In examining the models over the four listening DLPT testing cycles, once again, there are differences between the models for the two language groups (see Figure 28). There are no significant predictors in any of the Spanish models.

Figure 28. Predictive Models for Listening DLPT scores by Language for French and Spanish Learners.

Listening DLPT 1 for French Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.083	17.605	0.000
2	AR, MK	0.095	2.572	0.079
3	AR, MK, PC	0.106	1.194	0.313
4	AR, MK, PC, WK	0.115	0.654	0.625
5	AR, MK, PC, WK, DLAB	0.134	1.053	0.388

Listening DLPT 1 for Spanish Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.004	0.694	0.406
2	AR, MK	0.006	0.392	0.676
3	AR, MK, PC	0.007	0.098	0.961
4	AR, MK, PC, WK	0.007	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.009	0.097	0.993

Listening DLPT 2 for French Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.054	11.162	0.001
2	AR, MK	0.067	2.703	0.070
3	AR, MK, PC	0.079	1.264	0.288
4	AR, MK, PC, WK	0.080	0.070	0.991
5	AR, MK, PC, WK, DLAB	0.083	0.157	0.960

Listening DLPT 2 for Spanish Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.285	0.594
2	AR, MK	0.002	0.195	0.823
3	AR, MK, PC	0.002	0.000	1.000
4	AR, MK, PC, WK	0.004	0.130	0.971
5	AR, MK, PC, WK, DLAB	0.017	0.638	0.671

Listening DLPT 3 for French Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.025	4.986	0.027
2	AR, MK	0.029	0.799	0.451
3	AR, MK, PC	0.052	2.353	0.074
4	AR, MK, PC, WK	0.055	0.204	0.936
5	AR, MK, PC, WK, DLAB	0.058	0.153	0.979

Listening DLPT 3 for Spanish Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	0.319	0.573
2	AR, MK	0.003	0.196	0.822
3	AR, MK, PC	0.003	0.000	1.000
4	AR, MK, PC, WK	0.009	0.392	0.814
5	AR, MK, PC, WK, DLAB	0.010	0.049	0.999

Listening DLPT 4 for French Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.024	4.716	0.031
2	AR, MK	0.038	2.823	0.062
3	AR, MK, PC	0.039	0.101	0.959
4	AR, MK, PC, WK	0.039	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.041	0.100	0.992

Listening DLPT 4 for Spanish Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	0.429	0.513
2	AR, MK	0.002	0.000	1.000
3	AR, MK, PC	0.004	0.196	0.899
4	AR, MK, PC, WK	0.009	0.326	0.860
5	AR, MK, PC, WK, DLAB	0.014	0.245	0.942

Arithmetic Reasoning is a highly significant predictor of French listening DLPT 1 and DLPT 2 scores. It continues to be a significant predictor of listening DLPT 3 and DLPT 4 scores for the French learners. Additionally, Math Knowledge is a marginally significant predictor of listening DLPT 1, 2, and 4 scores for the French learners. Overall, the variance accounted for by the French models decreases over time. The variance accounted for by the Spanish models remains around 1% for all four testing cycles.

Table 65 demonstrates the differing patterns for the predictor models for the two languages. In the table, an asterisk indicates a significant predictor, and a check mark indicates a marginally significant predictor. Upon examining the table, it is clear that the predictive patterns for the two languages are substantially different. Over time, the pattern for each of the languages does vary slightly, but they remain fairly stable when compared across the two languages.

Table 65. Listening DLPT Model Standardized β Coefficients for French and Spanish learners.

Standardized β Coefficients for Listening DLPT Models

	Language	AR	MK	PC	WK	DLAB
DLPT 1	French	0.120 *	0.103 √	0.053	0.119	0.153
	Spanish	0.014	0.048	0.026	0.008	0.047
DLPT 2	French	0.089 *	0.127 √	0.104	0.045	0.060
	Spanish	-0.033	-0.057	-0.012	0.032	0.121
DLPT 3	French	-0.004 *	0.081	0.146 √	0.074	0.055
	Spanish	-0.011	0.036	-0.018	0.091	0.032
DLPT 4	French	0.074 *	0.120 √	0.019	0.020	0.054
	Spanish	-0.005	0.003	0.009	0.080	0.076

Reading DLPT scores over the three year period after graduating from DLI differ significantly for the two language groups, but the scores are consistent over the three-year period for both groups (see Table 66). Reading proficiency growth seems to have leveled off in both cases.

Table 66. Descriptive Statistics for Reading DLPT Scores for French and Spanish Learners.

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
French	197	24.8 (5.0)	23.3 (5.5)	24.5 (5.1)	24.4 (4.6)
Spanish	198	27.5 (3.3	27.1 (4.0)	27.0 (4.5)	27.7 (3.5)

The models for the four reading DLPT testing cycles also demonstrate substantial differences between the two language groups (see Figure 29). For the French group, Arithmetic Reasoning and DLAB are highly significant predictors of French reading DLPT 1 scores. For the reading DLPT 2 score, AR remains marginally significant, and MK is significant. For DLPT 3, MK and DLAB are significant. Finally, AR and PC are significant predictors of the reading DLPT 4 scores for the French learners. In the Spanish models, MK is the only predictor that reaches marginal significance for DLPT 1. No other predictor reaches significance in any of the other Spanish reading DLPT models. The variance accounted for by the models over time follows a similar pattern for both the French and Spanish language groups with a steep drop after the first testing cycle, and then a level off.

Figure 29. Predictive Models for Reading DLPT scores by Language for French and Spanish Learners.

Reading DLPT 1 for French Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.035	7.146	0.008
2	AR, MK	0.046	2.237	0.110
3	AR, MK, PC	0.060	1.445	0.231
4	AR, MK, PC, WK	0.066	0.413	0.799
5	AR, MK, PC, WK, DLAB	0.126	3.295	0.007

Reading DLPT 1 for Spanish Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.010	1.989	0.160
2	AR, MK	0.025	3.000	0.088
3	AR, MK, PC	0.033	0.807	0.491
4	AR, MK, PC, WK	0.049	1.088	0.364
5	AR, MK, PC, WK, DLAB	0.065	0.826	0.510

Reading DLPT 2 for French Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.014	2.786	0.097
2	AR, MK	0.030	3.200	0.043
3	AR, MK, PC	0.048	1.834	0.142
4	AR, MK, PC, WK	0.053	0.340	0.851
5	AR, MK, PC, WK, DLAB	0.067	0.720	0.609

Reading DLPT 2 for Spanish Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.205	0.651
2	AR, MK	0.002	0.195	0.823
3	AR, MK, PC	0.003	0.098	0.961
4	AR, MK, PC, WK	0.010	0.457	0.767
5	AR, MK, PC, WK, DLAB	0.023	0.642	0.668

Reading DLPT 3 for French Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.008	1.563	0.213
2	AR, MK	0.030	4.400	0.014
3	AR, MK, PC	0.036	0.604	0.613
4	AR, MK, PC, WK	0.037	0.067	0.992
5	AR, MK, PC, WK, DLAB	0.084	2.463	0.034

Reading DLPT 3 for Spanish Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.021	4.297	0.039
2	AR, MK	0.026	1.001	0.369
3	AR, MK, PC	0.027	0.100	0.960
4	AR, MK, PC, WK	0.032	0.334	0.855
5	AR, MK, PC, WK, DLAB	0.040	0.402	0.847

Reading DLPT 4 for French Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.029	5.918	0.016
2	AR, MK	0.040	2.223	0.111
3	AR, MK, PC	0.066	2.700	0.047
4	AR, MK, PC, WK	0.066	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.069	0.155	0.978

Reading DLPT 4 for Spanish Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	0.311	0.578
2	AR, MK	0.010	1.576	0.209
3	AR, MK, PC	0.010	0.000	1.000
4	AR, MK, PC, WK	0.014	0.262	0.902
5	AR, MK, PC, WK, DLAB	0.025	0.544	0.743

Table 67 demonstrates the differing patterns for the predictor models for the two languages. In the table, an asterisk indicates a predictor that had a significant impact when it was added to the model, and a check mark indicates a marginally significant predictor. Upon examining the table, it is clear that the predictive patterns for the two languages are substantially different. The pattern for each of the individual languages changes with time, but remains more or less stable when compared across the two languages. AR, MK, and DLAB are the best predictors of reading proficiency for the French language over time, and the model continues to account for a significant amount of the variance. For the Spanish language, the predict pattern is less stable, and does not significantly account for variance in reading proficiency.

Table 67. Reading DLPT Model Standardized β Coefficients for French and Spanish learners.

Standardized β Coefficients for Reading DLPT Models

	Language	AR	MK	PC	WK	DLAB
DLPT 1	French	0.010 *	0.061	0.053	0.109	0.266 *
	Spanish	-0.056	0.122	0.033	0.146	0.131
DLPT 2	French	-0.066 √	0.129 *	0.104	0.090	0.128
	Spanish	-0.024	0.023	-0.018	0.094	0.119
DLPT 3	French	-0.056	0.107 *	0.039	0.040	0.235 *
	Spanish	0.085 *	0.063	-0.029	0.084	0.094
DLPT 4	French	0.006 *	0.107	0.179 *	0.010	0.067
	Spanish	-0.011	0.079	-0.055	0.073	0.107

The models for the Oral Proficiency Interview for the two languages are shown in Figure 30. There are no significant predictors in either of the models. The model for predicting speaking proficiency levels for the French group, however, does reach marginal significance (F = 1.908, P = 0.095), and accounts for 4.8% of the variance. The model for the Spanish group is not significant and only accounts for 1.9% of the variance. With all five predictors in the French model, the standardized β coefficients are 0.192, -0.096, -0.020, -0.178, and 0.130 for AR, MK, PC, WK, and DLAB, respectively. For the Spanish model the standardized β coefficients are 0.100, -0.017, -0.047, -0.040, and -0.110 for AR, MK, PC, WK, and DLAB, respectively. With the exception of the DLAB coefficient, the overall profile of the predictors is similar between the languages, but the magnitudes of the β coefficients for the French group are generally greater.

Figure 30. Predictive Models for OPI scores by Language for French and Spanish Learners.

OPI for French Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.009	1.697	0.194
2	AR, MK	0.010	0.196	0.822
3	AR, MK, PC	0.013	0.295	0.829
4	AR, MK, PC, WK	0.033	1.331	0.256
5	AR, MK, PC, WK, DLAB	0.048	0.756	0.583

OPI for Spanish Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.251	0.617
2	AR, MK	0.002	0.195	0.823
3	AR, MK, PC	0.006	0.392	0.759
4	AR, MK, PC, WK	0.007	0.065	0.992
5	AR, MK, PC, WK, DLAB	0.019	0.590	0.708

Language Category II: German and Indonesian

German and Indonesian are used as the representatives for the language category II analyses. The models below examine the profile of predictor patterns for the various proficiency outcome measures. As previously, the significance of each predictor is determined when it is added to the model by looking at the change in the R^2 term, and predictors are added according to chronological progression. The descriptive statistics for the predictor variables and the DLI GPA outcome measure by language are shown in Table 68, below. Standard deviations are shown in parenthesis. There are significant differences between the means of the predictor variables and the outcome GPA measure for the German and Indonesian groups. The Indonesian group scored higher on all measures. The only measure that does not differ significantly between the groups is MK. These differences should not affect the analyses here since this research examines predictive profiles within the language groups.

Table 68. Descriptive statistics for predictors of language proficiency for German and Indonesian groups.

Language	N	AR	MK	PC	WK	DLAB	DLI GPA
German	116	58.3 (6.7)	60.8 (6.4)	56.9 (5.1)	56.9 (6.9)	105.5 (13.5)	3.07 (0.48)
Indonesian	104	60.4 (6.2)	61.1 (5.2)	59.7 (5.5)	59.5 (7.1)	110.3 (11.6)	3.54 (0.34)

Figure 31 depicts the predictive models for foreign language GPA for the German and Indonesian language groups. There are no significant predictors in the model for the German learners. For the Indonesian group, MK is a highly significant predictor of foreign language GPA, and DLAB is also significant. With all five predictors in the German model, the standardized β coefficients are 0.026, -0.099, 0.000, 0.173, and 0.219 for AR, MK, PC, WK, and DLAB, respectively. For the Indonesian model the standardized β coefficients are -0.016, 0.095, 0.013, -0.091, and 0.341 for AR, MK, PC, WK, and DLAB, respectively. The German model does not reach significance, and only accounts for 6.1% of the variance, but the Indonesian model is highly significant (F = 3.562, P = 0.005). Examining the magnitude and direction of the standardized β coefficients shows that the predictive profiles are completely different for the two language models. In fact, the only coefficient with similar magnitude and direction is the coefficient for the DLAB term.

Figure 31. Predictive Models for Foreign Language Grade Point Average by Language for German and Indonesian Learners.

Foreign Language GPA for German Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.007	0.796	0.374
2	AR, MK	0.007	0.000	1.000
3	AR, MK, PC	0.015	0.459	0.712
4	AR, MK, PC, WK	0.019	0.152	0.962
5	AR, MK, PC, WK, DLAB	0.061	1.241	0.295

Foreign Language GPA for Indonesian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.012	1.196	0.277
2	AR, MK	0.066	5.839	0.004
3	AR, MK, PC	0.066	0.000	1.000
4	AR, MK, PC, WK	0.071	0.179	0.949
5	AR, MK, PC, WK, DLAB	0.154	2.428	0.040

Listening DLPT scores over the three year period after graduating from DLI also show significant differences for the two language groups (see Table 69). Additionally, the scores trend in opposite directions over time with the average German listening DLPT score increasing and the average Indonesian listening DLPT score decreasing. The initial listening DLPT scores for the Indonesian group, however, were much higher than those for the German group.

Table 69. Descriptive statistics for listening DLPT scores for German and Indonesian learners.

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
German	116	19.9 (5.1)	19.7 (5.2)	20.4 (4.5)	20.5 (5.0)
Indonesian	104	26.5 (3.7)	25.4 (4.1)	25.5 (3.3)	24.8 (2.6)

The models for the four listening DLPT testing cycles demonstrate substantial differences between the predictive pattern profiles for the two language groups (see Figure 32). For example, DLAB is significant for the 1st and 4th DLPT scores listening and marginally significant for the 2nd and 3rd DLPT scores in the German model. In the Indonesian model, DLAB is not significant for any of the testing cycles. MK is significant for DLPT 3 in the German model and highly significant for DLPT 4. MK is marginally significant for DLPT 1 and significant for DLPT 4 in the Indonesian model. Finally, AR is marginally significant only for DLPT 1 in the Indonesian model.

Figure 32. Predictive Models for Listening DLPT scores by Language for German and Indonesian Learners.

Listening DLPT 1 for German Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.011	1.270	0.262
2	AR, MK	0.026	1.740	0.180
3	AR, MK, PC	0.027	0.058	0.982
4	AR, MK, PC, WK	0.028	0.038	0.997
5	AR, MK, PC, WK, DLAB	0.108	2.489	0.036

Listening DLPT 1 for Indonesian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.036	3.770	0.055
2	AR, MK	0.058	2.359	0.100
3	AR, MK, PC	0.062	0.215	0.886
4	AR, MK, PC, WK	0.081	0.689	0.601
5	AR, MK, PC, WK, DLAB	0.122	1.158	0.355

Listening DLPT 2 for German Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.014	1.571	0.213
2	AR, MK	0.033	2.220	0.113
3	AR, MK, PC	0.035	0.117	0.950
4	AR, MK, PC, WK	0.037	0.078	0.989
5	AR, MK, PC, WK, DLAB	0.104	2.075	0.074

Listening DLPT 2 for Indonesian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.012	1.264	0.264
2	AR, MK	0.021	0.928	0.399
3	AR, MK, PC	0.021	0.000	1.000
4	AR, MK, PC, WK	0.028	0.240	0.915
5	AR, MK, PC, WK, DLAB	0.06	0.843	0.523

Listening DLPT 3 for German Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.006	0.724	0.397
2	AR, MK	0.035	3.396	0.037
3	AR, MK, PC	0.043	0.472	0.702
4	AR, MK, PC, WK	0.045	0.078	0.989
5	AR, MK, PC, WK, DLAB	0.117	2.262	0.053

Listening DLPT 3 for Indonesian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	0.174	0.678
2	AR, MK	0.003	0.101	0.904
3	AR, MK, PC	0.01	0.357	0.784
4	AR, MK, PC, WK	0.036	0.899	0.468
5	AR, MK, PC, WK, DLAB	0.056	0.524	0.758

Listening DLPT 4 for German Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.005	0.519	0.473
2	AR, MK	0.059	6.485	0.002
3	AR, MK, PC	0.061	0.120	0.948
4	AR, MK, PC, WK	0.062	0.040	0.997
5	AR, MK, PC, WK, DLAB	0.142	2.587	0.030

Listening DLPT 4 for Indonesian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	0.198	0.658
2	AR, MK	0.037	3.671	0.029
3	AR, MK, PC	0.05	0.691	0.560
4	AR, MK, PC, WK	0.08	1.087	0.367
5	AR, MK, PC, WK, DLAB	0.085	0.135	0.984

Table 70 demonstrates the differing patterns for the predictor models for the two languages. In the table, an asterisk indicates a significant predictor, and a check mark indicates a marginally significant predictor. The pattern profiles for each of the languages are relatively stable over the four exam cycles, but they clearly differ from one another when compared across the two languages.

Table 70. Listening DLPT Model Standardized β Coefficients for German and Indonesian learners.

Standardized β Coefficients for Listening DLPT Models

	Language	AR	MK	PC	WK	DLAB
DLPT 1	German	-0.131	0.071	-0.113	0.077	0.300 *
	Indonesian	0.073 √	0.069	-0.143	0.165	0.239
DLPT 2	German	-0.145	0.103	-0.095	0.035	0.274 √
	Indonesian	0.009	0.017	-0.043	0.097	0.211
DLPT 3	German	-0.093	0.144 *	-0.159	0.040	0.285 √
	Indonesian	-0.184	-0.024	0.019	0.193	0.167
DLPT 4	German	-0.154	0.201 *	-0.118	0.076	0.302 *
	Indonesian	-0.269	0.187 *	0.041	0.212	0.083

Reading DLPT scores over the three year period after graduating from DLI differ significantly between the two language groups for every testing cycle except DLPT 2. The scores, however, remain stable around the ILR scale score of 2+ over the three-year period for both groups (see Table 71).

Table 71. Descriptive Statistics for Reading DLPT Scores for German and Indonesian learners.

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
German	116	27.4 (4.9)	25.8 (6.0)	27.0 (4.9)	26.4 (5.5)
Indonesian	104	26.0 (3.2)	24.7 (3.9)	24.9 (2.9)	25.2 (1.9)

Figure 33. Predictive Models for Reading DLPT scores by Language for German and Indonesian Learners.

Reading DLPT 1 for German Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	0.193	0.661
2	AR, MK	0.033	3.623	0.030
3	AR, MK, PC	0.038	0.294	0.746
4	AR, MK, PC, WK	0.042	0.156	0.960
5	AR, MK, PC, WK, DLAB	0.109	2.087	0.072

Reading DLPT 1 for Indonesian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.027	2.784	0.098
2	AR, MK	0.047	2.120	0.125
3	AR, MK, PC	0.047	0.000	1.000
4	AR, MK, PC, WK	0.069	0.788	0.536
5	AR, MK, PC, WK, DLAB	0.126	1.614	0.163

Reading DLPT 2 for German Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	0.252	0.616
2	AR, MK	0.019	1.958	0.146
3	AR, MK, PC	0.020	0.058	0.982
4	AR, MK, PC, WK	0.021	0.038	0.997
5	AR, MK, PC, WK, DLAB	0.084	1.909	0.099

Reading DLPT 2 for Indonesian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.017	1.731	0.191
2	AR, MK	0.017	0.000	1.000
3	AR, MK, PC	0.019	0.103	0.958
4	AR, MK, PC, WK	0.026	0.240	0.915
5	AR, MK, PC, WK, DLAB	0.049	0.599	0.701

Reading DLPT 3 for German Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.005	0.609	0.437
2	AR, MK	0.041	4.242	0.017
3	AR, MK, PC	0.045	0.237	0.870
4	AR, MK, PC, WK	0.045	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.103	1.794	0.120

Reading DLPT 3 for Indonesian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.064	0.801
2	AR, MK	0.001	0.000	1.000
3	AR, MK, PC	0.011	0.571	0.635
4	AR, MK, PC, WK	0.047	1.259	0.291
5	AR, MK, PC, WK, DLAB	0.077	0.804	0.550

Reading DLPT 4 for German Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.003	0.368	0.545
2	AR, MK	0.031	3.265	0.042
3	AR, MK, PC	0.031	0.000	1.000
4	AR, MK, PC, WK	0.031	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.062	0.917	0.473

Reading DLPT 4 for Indonesian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.008	0.826	0.366
2	AR, MK	0.023	1.735	0.182
3	AR, MK, PC	0.057	2.037	0.114
4	AR, MK, PC, WK	0.118	2.582	0.042
5	AR, MK, PC, WK, DLAB	0.122	0.126	0.986

The models for the four reading DLPT testing cycles also demonstrate substantial differences between the two language groups (see Figure 33). For the German group, Math Knowledge is a significant predictor of reading DLPT 1, 3, and 4 scores. DLAB scores are marginally significant predictors of German reading DLPT 1 and 2 scores. The only significant predictor of reading scores for the Indonesian group, however, is Word Knowledge for DLPT 4. AR is a marginally significant predictor of Indonesian reading DLPT 1 scores, but no other predictors reach even marginal significance.

Table 72 stresses the differing patterns for the predictor models for the two languages. Once again in the table, an asterisk indicates a predictor that had a significant impact when added to the predictive model, and a check mark indicates a marginally significant predictor. As seen, the predictive patterns for the two languages are substantially different. The pattern for each of the individual languages changes with time, but remains more or less stable when compared across the two languages. MK and DLAB are the best predictors of reading proficiency for the German language over time, and the model continues to account for a significant amount of the variance. For the Indonesian language, the predictive pattern is less stable, although WK does remain a stable predictor and even becomes significant in the DLPT 4 model for that group.

Table 72. Reading DLPT Model Standardized β Coefficients for German and Indonesian learners.

Standardized β Coefficients for Reading DLPT Models

	Language	AR	MK	PC	WK	DLAB
DLPT 1	German	-0.165	0.121 *	0.065	0.096	0.277 √
	Indonesian	0.018 √	0.041	-0.098	0.175	0.284
DLPT 2	German	-0.126	0.076	-0.021	0.071	0.268 √
	Indonesian	0.064	-0.100	0.003	0.103	0.180
DLPT 3	German	-0.127	0.162 *	-0.157	0.100	0.255
	Indonesian	-0.112	-0.099	0.017	0.230	0.204
DLPT 4	German	-0.141	0.144 *	-0.036	0.046	0.185
	Indonesian	-0.347	0.124	0.086	0.302 *	0.073

The models for the Oral Proficiency Interview for the two languages are shown in Figure 34. The only predictor to reach significance in either of the two language groups is MK for the Indonesian learners. The model for predicting speaking proficiency levels for the German group does not reach significance ($F = 1.250$, $P = 0.291$) and only accounts for 5.4% of the variance. The model for the Indonesian group is not significant either ($F = 1.588$, $P = 0.171$) and only accounts for 7.5% of the variance. With all five predictors in the German model, the standardized β coefficients are 0.054, -0.057, -0.329, 0.223, and 0.113 for AR, MK, PC, WK, and DLAB, respectively. For the Indonesian model the standardized β coefficients are -0.108, 0.276, -0.107, 0.148, and 0.006 for AR, MK, PC, WK, and DLAB, respectively. The overall profile of the predictors for the OPI models is very different between the language groups. The only similarities are the direction and rough magnitudes of the standardized β coefficients for PC and WK.

Figure 34. Predictive Models for Oral Proficiency Interview scores by Language for German and Indonesian Learners.

OPI for German Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	0.224	0.637
2	AR, MK	0.003	0.113	0.893
3	AR, MK, PC	0.027	1.394	0.248
4	AR, MK, PC, WK	0.042	0.585	0.674
5	AR, MK, PC, WK, DLAB	0.054	0.352	0.880

OPI for Indonesian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.088	0.767
2	AR, MK	0.059	6.225	0.003
3	AR, MK, PC	0.060	0.054	0.983
4	AR, MK, PC, WK	0.075	0.541	0.706
5	AR, MK, PC, WK, DLAB	0.075	0.000	1.000

Language Category III: Russian and Tagalog

The representative languages for the category III analyses are Russian and Tagalog. The models below examine the profile of predictor patterns for the various proficiency outcome measures. As in the two previous language category analyses, the significance of each predictor is determined when it is added to the model by looking at the change in the R^2 term, and predictors are added according to chronological progression. Table 73 (below) shows the descriptive statistics for the predictor variables and the DLI GPA outcome measure by language. Standard deviations are shown in parenthesis. There are significant differences between the means of the predictor variables and the outcome GPA measure for the Russian and Tagalog language groups. The Russian group scored higher on all predictor measures, but the Tagalog group has the higher foreign language GPA. The only measure that does not differ significantly between the groups is PC. These differences should not affect the analyses here since this research examines predictive profiles within the language groups.

Table 73. Descriptive statistics for predictors of language proficiency for Russian and Tagalog groups.

Language	N	AR	MK	PC	WK	DLAB	DLI GPA
Russian	190	61.0 (5.4)	62.8 (4.4)	58.8 (4.0)	59.0 (4.5)	109.5 (11.5)	3.26 (0.44)
Tagalog	186	59.0 (7.1)	61.1 (5.7)	58.1 (5.6)	57.0 (6.3)	107.0 (10.9)	3.44 (0.34)

The predictive models for foreign language GPA for the Russian and Tagalog language groups are shown in Figure 35. DLAB is a highly significant predictor of foreign language GPA for both groups. For the Tagalog language group, Math Knowledge is also a marginally significant predictor of foreign language GPA. With all five predictors in the Russian model, the standardized β coefficients are -0.018, 0.011, 0.105, -0.176, and 0.322 for AR, MK, PC, WK, and DLAB, respectively. For the Tagalog model the standardized β coefficients are -0.107, 0.090, -0.035, 0.156, and 0.268 for AR, MK, PC, WK, and DLAB, respectively. The Russian model is highly significant ($F = 5.503$, $P = 0.000$) and accounts for 13.0% of the variance. The Tagalog model is also highly significant ($F = 3.944$, $P = 0.002$) and accounts for 9.9% of the variance. The magnitude and direction of the standardized β coefficients demonstrate that the predictive profiles are completely different for the two language models. The only coefficient with similar magnitude and direction is the coefficient for the DLAB term.

Figure 35. Predictive Models for Foreign Language Grade Point Average by Language for Russian and Tagalog Learners.

Foreign Language GPA for Russian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.000	0.010	0.921
2	AR, MK	0.002	0.375	0.688
3	AR, MK, PC	0.004	0.188	0.905
4	AR, MK, PC, WK	0.030	1.662	0.161
5	AR, MK, PC, WK, DLAB	0.130	5.316	0.000

Foreign Language GPA for Tagalog Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.174	0.677
2	AR, MK	0.016	2.790	0.064
3	AR, MK, PC	0.016	0.000	1.000
4	AR, MK, PC, WK	0.032	1.003	0.407
5	AR, MK, PC, WK, DLAB	0.099	3.365	0.006

After graduating from DLI and for the subsequent three-year period for the participants, the mean Listening DLPT scores for the two language groups are nearly identical (see Table 74). The models for the four listening DLPT test cycles, however,

Table 74. Descriptive statistics for listening DLPT scores for German and Indonesian learners.

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
Russian	190	23.7 (4.7)	22.9 (4.3)	23.0 (4.6)	23.5 (3.7)
Tagalog	186	23.6 (4.2)	22.9 (4.3)	23.0 (4.6)	23.5 (3.7)

have substantial differences in the predictive pattern profiles for the two language groups (see Figure 36). There are no significant predictors for any of the Russian language group listening DLPT scores. For the Tagalog group, MK is a significant predictor of the DLPT 1 and 4 scores. MK is a marginally significant predictor of

Figure 36. Predictive Models for Listening DLPT scores by Language for Russian and Tagalog Learners.

Listening DLPT 1 for Russian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.005	1.033	0.311
2	AR, MK	0.006	0.188	0.829
3	AR, MK, PC	0.008	0.189	0.904
4	AR, MK, PC, WK	0.009	0.063	0.993
5	AR, MK, PC, WK, DLAB	0.039	1.444	0.211

Listening DLPT 1 for Tagalog Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.163	0.687
2	AR, MK	0.020	3.548	0.031
3	AR, MK, PC	0.025	0.469	0.704
4	AR, MK, PC, WK	0.025	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.065	1.936	0.091

Listening DLPT 2 for Russian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.004	0.671	0.414
2	AR, MK	0.006	0.376	0.687
3	AR, MK, PC	0.027	2.018	0.113
4	AR, MK, PC, WK	0.030	0.192	0.942
5	AR, MK, PC, WK, DLAB	0.052	1.073	0.377

Listening DLPT 2 for Tagalog Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0	0.062	0.804
2	AR, MK	0.016	2.976	0.054
3	AR, MK, PC	0.017	0.093	0.964
4	AR, MK, PC, WK	0.017	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.038	0.988	0.427

Listening DLPT 3 for Russian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.008	1.514	0.220
2	AR, MK	0.008	0.000	1.000
3	AR, MK, PC	0.011	0.284	0.837
4	AR, MK, PC, WK	0.013	0.126	0.973
5	AR, MK, PC, WK, DLAB	0.018	0.235	0.947

Listening DLPT 3 for Tagalog Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.000	0.047	0.829
2	AR, MK	0.010	1.848	0.161
3	AR, MK, PC	0.011	0.093	0.964
4	AR, MK, PC, WK	0.013	0.123	0.974
5	AR, MK, PC, WK, DLAB	0.055	2.011	0.079

Listening DLPT 4 for Russian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.012	2.331	0.128
2	AR, MK	0.013	0.189	0.828
3	AR, MK, PC	0.019	0.572	0.634
4	AR, MK, PC, WK	0.021	0.127	0.973
5	AR, MK, PC, WK, DLAB	0.03	0.429	0.828

Listening DLPT 4 for Tagalog Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.000	0.080	0.778
2	AR, MK	0.021	3.925	0.021
3	AR, MK, PC	0.021	0.000	1.000
4	AR, MK, PC, WK	0.023	0.124	0.974
5	AR, MK, PC, WK, DLAB	0.044	0.994	0.423

listening DLPT 2 scores. DLAB is also a marginally significant predictor of Tagalog listening DLPT 1 and 3 scores.

The differing patterns for the predictor profile models for the two languages are also demonstrated in Table 75. In the table, an asterisk indicates a significant predictor, and a check mark indicates a marginally significant predictor. The pattern profiles for Tagalog is relatively stable over the four exam cycles, but the Russian model coefficients show much more variation in magnitude and direction over the four test cycles. In any case, the predictor pattern profiles clearly differ from one another when compared across the two languages.

Table 75. Listening DLPT Model Standardized β Coefficients for Russian and Tagalog learners.

Standardized β Coefficients for Listening DLPT Models

	Language	AR	MK	PC	WK	DLAB
DLPT 1	Russian	0.038	0.008	0.051	-0.033	0.176
	Tagalog	-0.155	0.118 *	0.100	-0.018	0.207 √
DLPT 2	Russian	-0.030	0.039	0.181	-0.063	0.150
	Tagalog	-0.098	0.107 √	0.050	-0.032	0.150
DLPT 3	Russian	0.057	-0.029	0.029	0.058	0.071
	Tagalog	-0.107	0.072	-0.007	0.064	0.210 √
DLPT 4	Russian	0.065	0.018	0.106	-0.051	0.093
	Tagalog	-0.091	0.132 *	0.014	0.050	0.151

Unlike the listening DLPT scores, the Reading DLPT scores over the three-year period after graduating from DLI do differ significantly between the two language groups for every testing cycle. The scores, however, remain stable in the mid-2 range of the ILR scale over the three-year period for both groups (see Table 76).

Table 76. Descriptive Statistics for Reading DLPT Scores for Russian and Tagalog learners.

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
Russian	190	24.8 (4.5)	24.4 (5.1)	24.5 (5.1)	24.9 (4.5)
Tagalog	186	23.9 (4.1)	23.2 (4.1)	23.2 (4.0)	23.4 (3.7)

The models for the four reading DLPT testing cycles demonstrate substantial differences between the two language groups (see Figure 37). For the Russian group, DLAB is a marginally significant predictor of reading DLPT 1 scores, and PC is a significant predictor of reading DLPT 4 scores. For the Tagalog group, DLAB is a significant predictor of the reading DLPT 1 scores, and Math Knowledge is a significant predictor of all 4 reading DLPT scores.

The differing predictor pattern profiles are shown for the Russian and Tagalog language groups in Table 77. Once again in the table, an asterisk indicates a predictor that had a significant impact when added to the predictive model, and a check mark indicates a marginally significant predictor. As seen in the other language category analyses, the predictive patterns for these two languages are substantially different as well. The pattern for the Tagalog group over time appears more stable than the Russian group, but in any case, the patterns remain more or less stable when compared across the two languages. MK is the best predictor of reading proficiency

Figure 37. Predictive Models for Reading DLPT scores by Language for Russian and Tagalog Learners.

Reading DLPT 1 for Russian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.006	1.217	0.271
2	AR, MK	0.010	0.756	0.471
3	AR, MK, PC	0.011	0.095	0.963
4	AR, MK, PC, WK	0.012	0.063	0.993
5	AR, MK, PC, WK, DLAB	0.055	2.104	0.067

Reading DLPT 1 for Tagalog Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.005	0.999	0.319
2	AR, MK	0.031	4.910	0.008
3	AR, MK, PC	0.037	0.570	0.636
4	AR, MK, PC, WK	0.050	0.830	0.508
5	AR, MK, PC, WK, DLAB	0.111	3.105	0.010

Reading DLPT 2 for Russian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.003	0.603	0.438
2	AR, MK	0.006	0.564	0.570
3	AR, MK, PC	0.016	0.950	0.418
4	AR, MK, PC, WK	0.016	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.020	0.189	0.967

Reading DLPT 2 for Tagalog Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.000	0.028	0.866
2	AR, MK	0.018	3.354	0.037
3	AR, MK, PC	0.022	0.374	0.772
4	AR, MK, PC, WK	0.039	1.073	0.371
5	AR, MK, PC, WK, DLAB	0.059	0.961	0.443

Reading DLPT 3 for Russian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.006	1.084	0.299
2	AR, MK	0.007	0.188	0.829
3	AR, MK, PC	0.024	1.629	0.184
4	AR, MK, PC, WK	0.030	0.384	0.820
5	AR, MK, PC, WK, DLAB	0.036	0.288	0.919

Reading DLPT 3 for Tagalog Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.011	2.058	0.153
2	AR, MK	0.031	3.777	0.025
3	AR, MK, PC	0.033	0.189	0.904
4	AR, MK, PC, WK	0.037	0.252	0.908
5	AR, MK, PC, WK, DLAB	0.069	1.555	0.175

Reading DLPT 4 for Russian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.004	0.734	0.393
2	AR, MK	0.010	1.133	0.324
3	AR, MK, PC	0.047	3.630	0.014
4	AR, MK, PC, WK	0.051	0.261	0.903
5	AR, MK, PC, WK, DLAB	0.051	0.000	1.000

Reading DLPT 4 for Tagalog Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.002	0.319	0.573
2	AR, MK	0.028	4.895	0.009
3	AR, MK, PC	0.031	0.283	0.838
4	AR, MK, PC, WK	0.046	0.954	0.434
5	AR, MK, PC, WK, DLAB	0.054	0.383	0.860

for the Tagalog language group, and PC grows to be the best predictor for reading proficiency for the Russian group.

Table 77. Reading DLPT Model Standardized β Coefficients for Russian and Tagalog learners.

Standardized β Coefficients for Reading DLPT Models

	Language	AR	MK	PC	WK	DLAB
DLPT 1	Russian	0.012	0.042	0.004	0.056	0.210 √
	Tagalog	-0.119	0.143 *	0.044	0.142	0.256 *
DLPT 2	Russian	-0.022	0.056	0.107	0.006	0.060
	Tagalog	-0.154	0.140 *	0.015	0.159	0.146
DLPT 3	Russian	-0.003	-0.035	0.103	0.098	0.079
	Tagalog	-0.029	0.128 *	0.023	0.080	0.185
DLPT 4	Russian	-0.050	0.083	0.245 *	-0.073	0.012
	Tagalog	-0.124	0.180 *	0.000	0.156	0.093

The Oral Proficiency Interview predictive models for the two languages are shown in Figure 38. DLAB is a marginally significant predictor of OPI scores for the Tagalog group, and it is the only predictor to reach even marginal significance in either of the two language group models. The model for predicting speaking proficiency levels for the Russian group, however, is a highly significant with all five predictors in the model ($F = 3.588$, $P = 0.004$), and it accounts for 8.9% of the variance. The model for the Tagalog group is only marginally significant ($F = 2.205$, $P = 0.056$) and only accounts for 5.8% of the variance. With all five predictors in the Russian model, the standardized β coefficients are 0.216, -0.071, -0.104, -0.149, and

0.193 for AR, MK, PC, WK, and DLAB, respectively. For the Tagalog model the standardized β coefficients are -0.112, 0.085, 0.076, 0.041, and 0.208 for AR, MK, PC, WK, and DLAB, respectively. The predictor profiles for the OPI models are very different between the language groups. The only similarity is the direction and magnitude of the standardized β coefficient for DLAB.

Figure 38. Predictive Models for Oral Proficiency Interview scores by Language for Russian and Tagalog Learners.

OPI for Russian Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.012	2.373	0.125
2	AR, MK	0.014	0.379	0.685
3	AR, MK, PC	0.035	2.035	0.111
4	AR, MK, PC, WK	0.053	1.178	0.322
5	AR, MK, PC, WK, DLAB	0.089	1.827	0.110

OPI for Tagalog Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.000	0.045	0.832
2	AR, MK	0.011	2.080	0.128
3	AR, MK, PC	0.016	0.475	0.700
4	AR, MK, PC, WK	0.018	0.126	0.973
5	AR, MK, PC, WK, DLAB	0.058	1.964	0.086

Language Category IV: Arabic and Chinese

The category IV analyses use Arabic and Chinese as the representative languages. The predictor pattern profiles for the various proficiency outcome measures are examined in this section. Similar to the previous language category analyses, the significance of each predictor is determined when it is added to the model by looking at the change in the R^2 term, and predictors are added according to chronological progression. The descriptive statistics for the predictor variables and the DLI GPA outcome measure by language are shown in Table 78. Standard deviations are shown in parenthesis. There are no significant differences between the means of the predictor variables, but the outcome GPA measure for the Arabic and Chinese language groups are significantly different. The Chinese group graduated from DLI with a higher average grade point average than the Arabic group.

Table 78. Descriptive statistics for predictors of language proficiency for Arabic and Chinese groups.

Language	N	AR	MK	PC	WK	DLAB	DLI GPA
Arabic	197	61.1 (5.9)	62.9 (5.0)	59.4 (4.9)	59.7 (5.6)	118.2 (12.1)	3.17 (0.42)
Chinese	201	61.2 (6.0)	63.5 (4.0)	59.3 (4.9)	59.4 (5.5)	119.1 (13.3)	3.32 (0.39)

The predictive models for foreign language GPA for the Arabic and Chinese language groups are shown in Figure 39. Arithmetic Reasoning, Paragraph Comprehension, and DLAB are significant predictors of foreign language GPA for the Arabic group. Math Knowledge is a marginally significant predictor in that model, as well. For the Chinese language group, AR is a significant predictor, and DLAB is a highly significant predictor of foreign language GPA. The Arabic model is highly significant ($F = 5.111$, $P < 0.001$), explaining 11.8% of the variance, and with all five predictors in the model, the standardized β coefficients are -0.028, 0.057, 0.171, 0.013, and 0.245 for AR, MK, PC, WK, and DLAB, respectively. The Chinese model is also highly significant ($F = 5.599$, $P < 0.001$), accounting for 12.6% of the variance, and the standardized β coefficients with all five predictors in the model are 0.123, -0.040, -0.014, -0.074, and 0.340 for AR, MK, PC, WK, and DLAB, respectively. The magnitude and direction of the standardized β coefficients demonstrate that the predictive profiles are completely different for the two language models. The only coefficient with similar magnitude and direction is the coefficient for the DLAB term.

Figure 39. Predictive Models for Foreign Language Grade Point Average by Language for Arabic and Chinese Learners.

Foreign Language GPA for Arabic Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.026	5.234	0.023
2	AR, MK	0.038	2.420	0.092
3	AR, MK, PC	0.068	3.122	0.027
4	AR, MK, PC, WK	0.069	0.069	0.991
5	AR, MK, PC, WK, DLAB	0.118	2.667	0.024

Foreign Language GPA for Chinese Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.020	3.992	0.047
2	AR, MK	0.020	0.000	1.000
3	AR, MK, PC	0.020	0.000	1.000
4	AR, MK, PC, WK	0.021	0.067	0.992
5	AR, MK, PC, WK, DLAB	0.126	5.887	0.000

At DLI graduation and for the three-year period following graduation, the mean Listening DLPT scores for the two language groups differ significantly (see Table 79). The mean scores for the Chinese are higher than those for the Arabic participants. The scores for both groups follow a similar pattern where the highest mean scores are at graduation, then the scores drop slightly on DLPT 2, and finally remain fairly constant for the final two exam cycles.

Table 79. Descriptive statistics for listening DLPT scores for Arabic and Chinese learners.

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
Arabic	197	22.1 (5.8)	20.3 (6.5)	20.3 (6.6)	20.2 (5.8)
Chinese	201	23.1 (4.9)	21.6 (5.7)	21.5 (6.2)	21.9 (5.8)

Over the four testing cycles, the two language groups also show substantial differences in the predictive pattern profiles (see Figure 40). On the first listening DLPT, MK is highly significant for the Arabic learners, but adds very little to the Chinese model. And, although the DLAB adds incremental predictive validity in both models, it is highly significant in the Chinese model and only marginally significant in the Arabic model. The second DLPT testing cycle shows a similar pattern, but now DLAB is only marginally significant in both models. For the DLPT 3 models, MK is once again highly significant for the Arabic participants, but now AR is the only significant predictor in the case of the Chinese participants. For DLPT 4, both groups appear to converge somewhat when looking at the predictor pattern profiles, and AR is highly significant for the Arabic group and marginally significant for the Chinese group. In any case, the predictor pattern across the four testing cycles appears much more stable for the Arabic learners than the Chinese learners; although the incremental predictive validity of the DLAB all but disappears in the case of the Arabic learners while it continues to account for around 3% of the variance for the Chinese learners.

Figure 40. Predictive Models for Listening DLPT scores by Language for Arabic and Chinese Learners.

Listening DLPT 1 for Arabic Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.011	2.141	0.145
2	AR, MK	0.055	8.986	0.000
3	AR, MK, PC	0.075	2.086	0.103
4	AR, MK, PC, WK	0.077	0.139	0.968
5	AR, MK, PC, WK, DLAB	0.115	2.050	0.074

Listening DLPT 1 for Chinese Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.009	1.764	0.186
2	AR, MK	0.01	0.200	0.819
3	AR, MK, PC	0.01	0.000	1.000
4	AR, MK, PC, WK	0.014	0.266	0.900
5	AR, MK, PC, WK, DLAB	0.083	3.687	0.003

Listening DLPT 2 for Arabic Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.007	1.321	0.252
2	AR, MK	0.026	3.784	0.024
3	AR, MK, PC	0.026	0.000	1.000
4	AR, MK, PC, WK	0.028	0.132	0.971
5	AR, MK, PC, WK, DLAB	0.071	2.222	0.054

Listening DLPT 2 for Chinese Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.005	0.914	0.340
2	AR, MK	0.007	0.399	0.672
3	AR, MK, PC	0.015	0.804	0.493
4	AR, MK, PC, WK	0.018	0.201	0.938
5	AR, MK, PC, WK, DLAB	0.057	2.027	0.077

Listening DLPT 3 for Arabic Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.005	0.987	0.322
2	AR, MK	0.040	7.073	0.001
3	AR, MK, PC	0.040	0.000	1.000
4	AR, MK, PC, WK	0.041	0.067	0.977
5	AR, MK, PC, WK, DLAB	0.048	0.353	0.880

Listening DLPT 3 for Chinese Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.019	3.875	0.050
2	AR, MK	0.019	0.000	1.000
3	AR, MK, PC	0.029	1.020	0.385
4	AR, MK, PC, WK	0.032	0.204	0.936
5	AR, MK, PC, WK, DLAB	0.061	1.513	0.187

Listening DLPT 4 for Arabic Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.035	6.984	0.009
2	AR, MK	0.039	0.807	0.448
3	AR, MK, PC	0.04	0.101	0.959
4	AR, MK, PC, WK	0.041	0.067	0.992
5	AR, MK, PC, WK, DLAB	0.043	0.100	0.992

Listening DLPT 4 for Chinese Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.018	3.560	0.061
2	AR, MK	0.025	1.422	0.244
3	AR, MK, PC	0.033	0.819	0.485
4	AR, MK, PC, WK	0.036	0.204	0.936
5	AR, MK, PC, WK, DLAB	0.067	1.628	0.154

The standardized β coefficients for the different testing cycles are summarized by language group in Table 80, below. The table demonstrates that the various aptitude measures differ radically in how they predict the listening DLPT scores for the two language groups over the first three testing cycles. DLPT 4 shows very similar predictor pattern profiles across the two groups. Examining this in more detail, the pattern for the independent variable predictors remains more stable over the four testing cycles for the Arabic group. Notably, AR shows a steady upward trend in magnitude and direction for the Arabic group, while the other predictors show a steady decline in magnitude. For the Chinese group, the independent variables show a much more erratic predictive pattern, but by DLPT 4 the patterns for both groups end up looking very similar.

Table 80. Listening DLPT Model Standardized β Coefficients for Arabic and Chinese learners.

Standardized β Coefficients for Listening DLPT Models

	Language	AR	MK	PC	WK	DLAB
DLPT 1	Arabic	-0.082	0.173 *	0.200	-0.095	0.213 √
	Chinese	0.103	-0.068	0.046	-0.139	0.274 *
DLPT 2	Arabic	0.007	0.076 *	0.057	-0.097	0.230 √
	Chinese	0.077	0.008	-0.144	0.034	0.209 √
DLPT 3	Arabic	-0.011	0.172 *	0.031	-0.061	0.096
	Chinese	0.165 *	-0.024	-0.153	0.044	0.181
DLPT 4	Arabic	0.169 *	0.055	-0.075	0.042	0.041
	Chinese	0.128 √	0.056	-0.144	0.039	0.184

The reading DLPT scores over the three-year period after graduating from DLI follow a very similar pattern to the listening DLPT scores over that same period for these two groups. The scores differ significantly between the two language groups for every testing cycle. The scores, however, remain stable in the 2 range of the ILR scale over the three-year period for both groups (see Table 81). The mean scores for the Chinese participants are higher than those for the Arabic participants. The scores for both groups follow a similar pattern where the highest mean scores are at graduation, and then the scores drop slightly for the second testing cycle; finally, the scores remain fairly constant for the final two exam cycles.

Table 81. Descriptive Statistics for Reading DLPT Scores for Russian and Tagalog learners.

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
Arabic	197	23.3 (5.3)	21.6 (5.9)	21.3 (6.0)	21.3 (5.8)
Chinese	201	25.7 (4.3)	23.2 (5.3)	22.8 (5.8)	22.8 (5.7)

Over the four testing cycles, the two language groups also show substantial differences in the predictive pattern profiles (see Figure 41). On the first listening DLPT, however, the two language groups appear similar based on the significant predictors. MK is highly significant and AR is marginally significant for the Arabic learners, and MK and AR are both significant in the Chinese model. Then, the models begin to diverge. MK remains significant, and AR remains marginally significant for the Arabic model. DLAB is also highly significant for the Arabic group. For the Chinese group the only significant predictor is DLAB. For the DLPT 3 models, MK is still highly significant for the Arabic participants, but there are no other significant

Figure 41. Predictive Models for Reading DLPT scores by Language for Arabic and Chinese Learners.

Reading DLPT 1 for Arabic Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.017	3.273	0.072
2	AR, MK	0.053	7.375	0.001
3	AR, MK, PC	0.056	0.308	0.820
4	AR, MK, PC, WK	0.056	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.089	1.739	0.128

Reading DLPT 1 for Chinese Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.020	3.973	0.048
2	AR, MK	0.036	3.286	0.040
3	AR, MK, PC	0.043	0.724	0.539
4	AR, MK, PC, WK	0.072	2.052	0.089
5	AR, MK, PC, WK, DLAB	0.085	0.696	0.627

Reading DLPT 2 for Arabic Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.019	3.704	0.056
2	AR, MK	0.048	5.910	0.003
3	AR, MK, PC	0.048	0.000	1.000
4	AR, MK, PC, WK	0.051	0.203	0.937
5	AR, MK, PC, WK, DLAB	0.123	3.941	0.002

Reading DLPT 2 for Chinese Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.013	2.672	0.104
2	AR, MK	0.014	0.201	0.818
3	AR, MK, PC	0.014	0.000	1.000
4	AR, MK, PC, WK	0.019	0.335	0.854
5	AR, MK, PC, WK, DLAB	0.065	2.411	0.038

Reading DLPT 3 for Arabic Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.012	2.275	0.133
2	AR, MK	0.057	9.258	0.000
3	AR, MK, PC	0.059	0.206	0.892
4	AR, MK, PC, WK	0.059	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.071	0.620	0.685

Reading DLPT 3 for Chinese Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.010	2.044	0.154
2	AR, MK	0.010	0.000	1.000
3	AR, MK, PC	0.011	0.100	0.960
4	AR, MK, PC, WK	0.011	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.079	3.618	0.004

Reading DLPT 4 for Arabic Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.042	8.569	0.004
2	AR, MK	0.042	0.000	1.000
3	AR, MK, PC	0.051	0.920	0.432
4	AR, MK, PC, WK	0.052	0.068	0.992
5	AR, MK, PC, WK, DLAB	0.064	0.615	0.689

Reading DLPT 4 for Chinese Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.017	3.402	0.067
2	AR, MK	0.024	1.420	0.244
3	AR, MK, PC	0.027	0.305	0.822
4	AR, MK, PC, WK	0.027	0.000	1.000
5	AR, MK, PC, WK, DLAB	0.045	0.924	0.467

predictors for the Arabic group. For the Chinese group, DLAB is highly significant with no other significant predictors. Finally, for DLPT 4, both groups appear to converge somewhat again. AR is a highly significant predictor of the Arabic reading DLPT scores, and the measure is a marginally significant predictor of the Chinese reading DLPT scores. No other aptitude measures reach significance in the reading DLPT 4 models.

Table 82. Reading DLPT Model Standardized β Coefficients for Arabic and Chinese learners.

Standardized β Coefficients for Reading DLPT Models

	Language	AR	MK	PC	WK	DLAB
DLPT 1	Arabic	-0.015 √	0.152 *	0.050	-0.004	0.198
	Chinese	0.071 *	0.133 *	0.234	-0.255	0.119
DLPT 2	Arabic	0.054 √	0.092 *	0.051	-0.124	0.296 *
	Chinese	0.109	-0.005	0.050	-0.138	0.225 *
DLPT 3	Arabic	0.030	0.191 *	-0.046	-0.023	0.124
	Chinese	0.088	-0.031	-0.026	-0.047	0.273 *
DLPT 4	Arabic	0.242 *	-0.030	-0.139	0.027	0.120
	Chinese	0.121 √	0.070	-0.034	-0.058	0.139

The standardized β coefficients for the four reading testing cycles are summarized by language group in Table 82, above. The table demonstrates that the predictor profile patterns are relatively similar for the reading tests between the two language groups with a couple of exceptions. The most notable difference is in the predictive capabilities of MK. The standardized β coefficient for MK appears more consistent for the Arabic group over the four testing cycles, although is drops off for the DLPT 4. Also, AR is more consistent for the Chinese group, while there is steady growth in the magnitude of the AR standardized β coefficient for the Arabic group. The other predictors are more erratic, but the magnitude and direction of their standardized β coefficients are similar across the two groups.

The Oral Proficiency Interview predictive models for the two languages are shown in Figure 42. DLAB is a significant predictor of OPI scores for the Arabic group. No other predictor reaches significance in either of the two language group models. The model for predicting speaking proficiency levels for the Arabic group is significant with all five predictors in the model ($F = 2.512$, $P = 0.031$), and it accounts for 6.2% of the variance. The model for the Chinese group is not significant and only accounts for 2.4% of the variance. With all five predictors in the Arabic model, the standardized β coefficients are -0.014, -0.069, -0.033, 0.042, and 0.263 for AR, MK, PC, WK, and DLAB, respectively. For the Chinese model the standardized β coefficients are 0.031, 0.018, 0.097, -0.126, and 0.100 for AR, MK, PC, WK, and DLAB, respectively. The magnitudes of the standardized β coefficient are near zero, but do show some differences. Notably, the language groups differ in regards to the coefficients for PC and WK.

Figure 42. Predictive Models for Oral Proficiency Interview scores by Language for Russian and Tagalog Learners.

OPI for Arabic Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.001	0.270	0.604
2	AR, MK	0.001	0.000	1.000
3	AR, MK, PC	0.002	0.097	0.962
4	AR, MK, PC, WK	0.005	0.194	0.941
5	AR, MK, PC, WK, DLAB	0.062	2.917	0.015

OPI for Chinese Learners

Model	Predictors	R^2	F of ΔR^2	P of ΔR^2
1	AR	0.003	0.548	0.460
2	AR, MK	0.004	0.199	0.821
3	AR, MK, PC	0.004	0.000	1.000
4	AR, MK, PC, WK	0.010	0.398	0.810
5	AR, MK, PC, WK, DLAB	0.024	0.703	0.622

7.5 Discussion

7.5.1 Combined Language Groups

The investigation conducted on the combined language groups in this chapter sheds some light on many of the earlier research findings that pertain to the Defense Language Aptitude Battery (DLAB), which is an appropriate starting point for this discussion. As mentioned, the Defense Language Aptitude Battery was designed by Petersen and Al-Haik to serve as the primary selection tool for military personnel to train at the Defense Language Institute (DLI). It was intended to replace the older Defense Language Aptitude Test (DLAT) and provide an equal or higher predictive validity than that of concurrently available commercial foreign language aptitude tests (Peterson & Al-Haik, 1976).

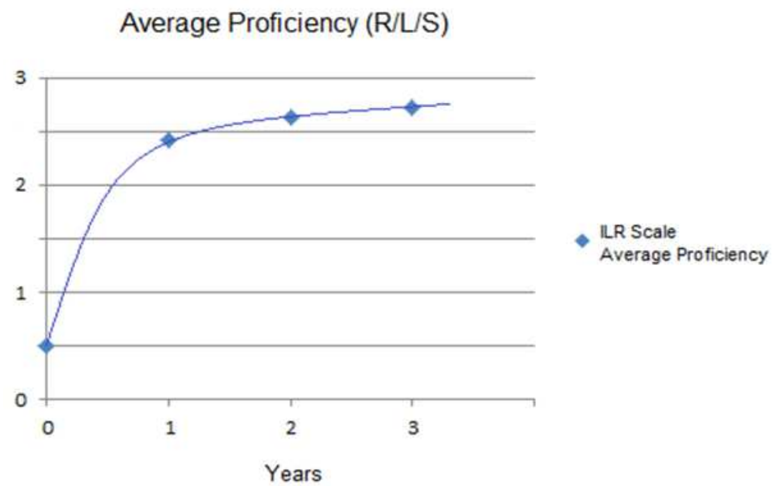
Peterson & Al-Haik (1976) examined the predictive validity of the DLAB using DLI grade point average (in the current study referred to as foreign language GPA) as the outcome measure. The current study shows that DLAB is a highly significant predictor of DLI grade point average, even after introducing other aptitude measures into the predictive model. This not only replicates the findings of Peterson & Al-Haik (1976), but also demonstrates the incremental predictive validity of the DLAB measure. Additionally, the low correlations between DLAB and the other aptitude measures in the model provide some evidence for the componential nature of foreign language aptitude since the predictors are unrelated and both Arithmetic Reasoning and Paragraph Comprehension are also significant in the model.

Silva and White (1993) were concerned that the initial development of the DLAB was intended to predict success at DLI using foreign language GPA as the outcome measure and not a more specific measure of foreign language proficiency. They also intended to investigate whether or not the DLAB was a useful tool above and beyond other measures of general aptitude, thus initiating a discussion of specific cognitive aptitudes. Silva and White (1993) commented that the DLAB measures the existence of strategies consisting of a specific kind of crystallized ability with predictive power beyond that of "g." They showed that the DLAB did provide incremental predictive validity when added to predictive models of foreign language proficiency. The outcome measures that they used were listening and reading Defense Language Proficiency Tests. They also used ASVAB scores as the other independent variables. Their findings support the hypothesis that if language learning is componential in nature and both general and specific cognitive abilities contribute to learning then DLAB should still have predictive power in the presence of other individual difference measures. In the current research, the findings of Silva and White (1993) were replicated. This highlights two points. First, it provides additional evidence to support the hypothesis just mentioned. But second, and perhaps more importantly, it suggests that predictive validity is maximized when DLAB scores are combined with the other general measures of cognitive ability.

In addition to examining incremental predictive validity in a static environment, a time element is also added. The interest in change over time is twofold. First, if the same pattern of cognitive factors can be shown to predict higher proficiency independent of a time element, then higher scores on the cognitive measure would indicate consistently higher rates of learning, thus allowing for easier identification of high-level aptitude learners. If the predictive pattern changes, then that would be more in line with the claims of Linck et al. (2012) which suggest that aptitude measures such as the DLAB are useful in distinguishing higher rates of learning, but only in lower level learners. Second, a time element is important precisely to examine changes in predictive patterns which may better indicate how language learning takes place. For example, one may anticipate that basic, general, cognitive memory elements may be more important for initial vocabulary learning, but the predictive power of their measures may fade with time as other learners catch up and more specific associative memory elements become increasingly important. This is a hypothetical example, but studies like Carson et al. (1990) and Vandergrift (2006) demonstrate that L2 vocabulary knowledge and L1 comprehension skills predict L2 comprehension, indicating that a certain level of L2 vocabulary knowledge is required for L1 transfer of comprehension skills. In other words, changes in the predictive power of individual difference measures would be the expected result for studies like the current one according to their findings.

For the listening DLPT results over the four testing cycles, the predictor pattern profile remains fairly consistent. For the first exam cycle, the significant predictors are AR, MK, and DLAB. The predictive power of AR appears to be masked (coincident or collinear) by MK, once MK is added to the model, since the standardized β coefficient of the AR term drops to zero upon its addition, at least for the first two exam cycles. These same three terms continue to have the largest impact over the next three test cycles, but the magnitude of their standardized β coefficients continues to decrease with each successive testing cycle. This finding provides evidence to support the claims of Linck et al. (2012) that the predictive ability of these aptitude measures appears most useful in distinguishing rates of learning in lower level learners. This finding is also in line with the findings of Carson et al. (1990) and Vandergrift (2006), but adds clarity to the example used above. The same learning elements are involved in developing listening proficiency over time, but the ability of the individual difference measures to distinguish higher level learners diminishes as slower learners catch up and learning rates decrease as proficiency levels increase. Figure 43, below, is taken from Wagener (2014) which shows average DLPT scores for learners in the Olmsted Program over their 3 year course of study. This figure is useful since the intensive course of study for Olmsted Scholars condenses the foreign language learning timeline. It demonstrates how learners progress more rapidly through the lower level ILR scale scores, but learning rate (as measured by change in ILR scale score) decreases exponentially over time. Thus, smaller differences in proficiency scores later in the learning process make it more difficult to differentiate higher level learners. Additionally, the lack of granularity of the ILR scale diminishes the likelihood of distinguishing higher level learners.

Figure 43. Proficiency vs. Time Across Languages For Olmsted Scholars



In any case, the findings of this listening proficiency portion of the study would tend to indicate that learners continue to utilize the same learning abilities that are measured by the AR, MK and DLAB to develop foreign language listening proficiency over time.

The reading DLPT results over the four testing cycles nearly replicate the findings of the listening DLPT section. The predictor pattern profile remains fairly consistent. For the first exam cycle, the significant predictors are AR, MK, and DLAB. Once again, the predictive power of AR appears to be masked by MK, and these same three terms continue to have the largest impact over all of the test cycles. As in the listening DLPT section, the magnitude of their standardized β coefficients continues to decrease with each successive testing cycle. Also, and more notably so in the reading DLPT section, the MK measure has the largest impact of any of the predictor measures. This indicates that a measure of general cognitive memory may be the best indicator of long term language learning success as measured by listening and reading proficiency across languages.

7.5.2 Language Categories

Since its inception the use of the DLAB has evolved from predicting success at DLI to helping the institution sort learners into specific language categories applicable for the range of DLAB scores attained. Lett & O'Mara (1990) describe how the DLAB is used to determine probable success of learners in a particular category of languages. Of course, the notion that the DLAB is the appropriate tool to predict successful placement in a language category program has been challenged by authors such as Child (1998) and Lowe (1998). For example, Child (1998) goes into detail explaining that the DLAB may be the preferred foreign language aptitude measure for category I and II languages, but it is not a credible aptitude test for category III and IV languages due to significant differences in syntactic patterns and structures between English and the languages in those categories. Additionally, Lowe indicates that language categories were formed mainly by grouping languages with a similar time to train learners at DLI to an appropriate proficiency level in two of the three modalities. He claims that if a language aptitude measure can be tailored to specific linguistic features that vary across languages, then the possibility exists that certain aptitude components could play a larger role in one language category than another. This claim is explored in the current research by looking at the predictive patterns of several general and specific cognitive aptitude measures within each of the four language categories.

First, examining the predictive patterns of the individual difference measures for foreign language GPA, the main result was very much expected; DLAB is a highly significant predictor of foreign language GPA for all language categories. As stated by Petersen and Al Haik (1976), the DLAB was specifically designed to predict success at DLI where the outcome measure of success was DLI GPA, so the current findings replicate the findings of the DLAB developmental study. However, a deeper exploration of the current data also shows that Arithmetic Reasoning is a significant predictor of foreign language GPA for the category I, II, and IV languages, but not the category III languages. And, although Paragraph Comprehension is not significant in any of the language category models, the standardized β coefficient suggests that it also has an important impact on the predictive validity of the category II, III, and IV models, but not the category I model where L1 Word Knowledge has a much larger impact. Interestingly, L1 Word Knowledge actually has a negative impact on learning category III and IV languages as evidenced by the standardized β coefficients, at least relative to the other predictors in the models. In any case, some interesting differences in the predictive patterns of the ID measures begin to emerge from the DLI GPA outcomes. These are further examined in the discussion of the listening and reading proficiency models that follows.

The listening proficiency models accentuate the differing predictor pattern profiles between the language category models. Specifically, the current findings indicate that AR is a significant predictor of listening DLPT scores for language category I and IV learners only. As a measure of general cognitive reasoning ability, AR was originally anticipated to have predictive success across all language categories, but the current results demonstrate a focus in only two of the categories. According to the Personnel Testing Division of DMDC, Arithmetic Reasoning is an aptitude component that measures logical thinking and predicts success in the mathematics domain (ASVAB Technical Bulletin, 2012; The ASVAB Career Exploration Program, 2011). The AR test examines the ability of a candidate to solve basic mathematical problems encountered in day-to-day life. The candidate must select the appropriate math functions and perform operations in the correct order. Additionally, in a factor analysis performed be Alderton et al. (1997), AR loads nearly equally on a math factor and a non-verbal reasoning factor. Translating this to language learning, AR indicates a sort of symbolic assembly. Here I will define symbolic assembly as the ability to group and organize items with a particular parameter set into meaningful structures. This definition would seem to imply that AR should differentiate learner success across all languages. However, in the results of this research, this is not the case. This may indicate that learners rely on this ability for languages that are dramatically different than their own L1 or for languages where strong L1 influence takes place. Since AR shows optimum predictive success in the category I and IV languages, this argument stands to reason.

Next, looking at MK, it is a significant predictor for the category IV language learners' listening DLPT scores, but it is only significant on the last of the four tests for the category II and III learners; and it is not significant for any of the listening DLPT test scores for the category I learners. Higher MK scores also slow attrition for the category II and IV languages. As an unrelated measure of crystallized knowledge and logical thinking (Alderton et al., 1997; ASVAB Technical Bulletin, 2012; The ASVAB Career Exploration Program, 2011), MK was not expected to have any predictive ability for language learning in any of the categories. Although since it does have a general cognitive memory component, this finding could be interpreted as evidence of a long term memory element required for language learning. According to Child (1998) the category I languages have the most overlap with L1 English. This may be allowing the linguistic associations between L1 and L2 to negate much of the discriminatory effect of non-verbal long term memory since learners can rely on verbal associations. But, as languages become more distant from English, with fewer and weaker associations, long term memory differences are able to differentiate learner proficiencies.

Remaining on the topic of L1-L2 associations, PC and WK were expected to predict differences in language categories where a high degree of L1 transfer is expected. In other words, PC and WK were expected to have some predictive ability for category I and II languages. The current results show that these two measures do not significantly impact any of the language category models. Interestingly, however, when examining the trend of the standardized β coefficients, Word Knowledge demonstrates the expected profile where higher L1 WK scores are better at discriminating L2 proficiency in the category I and II languages than the category III and IV languages. But, higher scores in WK appear to speed attrition in category III languages. Additionally, higher WK scores show a large negative impact when predicting category IV listening proficiency scores. This indicates that greater L1 word knowledge increases the difficulty of making gains in category III and IV listening proficiency.

Finally, for the listening DLPT section, DLAB adds significant incremental predictive validity for the category II and IV models only. Although in all fairness, DLAB generally shows up as a significant predictor in post hoc models for all of the language categories, especially in the earlier testing cycles. This may indicate some collinearity in the measures although the low correlations between them would indicate limited overlap in what they are measuring. Differences in the standardized β coefficients, however, once again points to differential predictive patterns across the language categories.

The reading models also add evidence to support differing predictor pattern profiles between the language category models. As in the listening DLPT section, the current findings indicate that AR is a significant predictor of reading DLPT scores for language category I and IV learners only. Additionally, AR slows attrition in reading DLPT scores for category IV languages. Here again, the reason that a measure of general cognitive reasoning ability would differentially predict foreign language learning between language categories and why its predictive abilities are specifically found in the category I and IV languages is not readily apparent. This result is addressed further in the following section on individual languages.

As mentioned earlier MK, as a measure of crystallized math knowledge (Alderton et al., 1997), was not expected to have any predictive ability for language learning in any of the categories. The findings for MK as a predictor of reading DLPT scores, however, show that not only does it significantly predict scores across language categories, but it also significantly predicts scores over time in all categories. Specifically, higher MK scores slow attrition in the category II and IV reading DLPT models. The logical explanation is that it is a measure of general cognitive memory, and as such it also may be used as a measure of orthographic memory. Additionally, its limited correlation with DLAB in the reading score models provides strong evidence to support a model of specific and general aptitudes.

As in the listening models, PC and WK fail to predict reading DLPT scores. WK, however, is retained as a significant predictor in post hoc models early on for categories I and III. When examining the trend of the standardized β coefficients, Word Knowledge acts as expected and is better at discriminating L2 proficiency in the category I and II languages than the category III and IV languages. WK also significantly slows attrition in the category I and II models. As in the listening discussion above, WK scores have a large negative impact when predicting category IV reading proficiency scores. PC has little effect on any of the category models. The findings for PC and WK indicate that crystallized L1 knowledge may assists learners in L2 reading, but fluid L1 abilities may not transfer to the L2.

Finally, DLAB adds significant incremental predictive validity for the first testing cycle in all categories, but its predictive validity clearly diminishes in subsequent testing cycles. The DLAB, however, does continue to successfully predict reading proficiency scores in post hoc models through the third testing cycle when other predictive measures are removed from the model in stepwise linear regression. This demonstrates the value of the DLAB as a tool to predict reading proficiency, but also provides evidence to support the findings of Linck et al. (2012) that the DLAB is more likely a better predictor of proficiency among lower level learners.

The varying predictive patterns for all outcome measures provide evidence to support the claims of Child (1998) and Lowe (1998). Distance from English appears to stand out as a valid explanation specifically when examining the attrition models since higher scores on measures that are associated with the L1 (PC and WK) slow attrition in the category I and II models while speeding it in the category III and IV models. Also, higher scores on the general cognitive measures (AR and MK) slow attrition in the category IV models where L1 linkages are less readily available. Additionally, the correlations between the ASVAB predictors and the DLAB are minimal when controlling for all outcome measures used in the current analysis. This demonstrates that the measures are unrelated which could be interpreted as evidence supporting differential aptitudes, especially considering the incremental predictive validity of the unrelated measures in the models.

7.5.3 Individual Languages

When looking at all languages grouped together, the predictive power of different aptitude measures is apparent and lends support to an argument for a model of specific and general aptitudes. As the focus narrows to language categories, it is more apparent that general cognitive aptitudes differentially find their way into language learning which tends to support a model of aptitude components in line with the claims of Linck et al. (2012). Taking that one step further and analyzing individual languages within the language categories, there appear to be certain predictor profiles that emerge for each particular language. This is in line with the arguments of Child (1998) and Lowe (1998) that state varying distances from the L1 may call for different aptitude measures to determine probability of success in a language. A line of reasoning drawn from Child (1998) and Lowe (1998) leads to the basis for the current study and the claim that different components of aptitude differentially predict success depending on the language being learned. In this analysis of individual languages, support is found for this claim.

The primary analysis here is looking at the magnitude and direction of the standardized β coefficients, but significant incremental predictive validity of the aptitude components is also addressed. Starting with the two category I languages, French and Spanish, the predictor pattern profiles are completely different. In the listening models AR and MK have the largest positive magnitudes of the standardized β coefficients and add significant incremental predictive validity for the French learners. The direction and magnitude is stable over time. The standardized β coefficients for the Spanish learners are all near zero and fluctuate in direction over time. In the reading models MK and DLAB have the largest positive magnitudes for the French group and generally add significant predictive validity. For the Spanish group, there are no significant predictors although the standardized β coefficient for DLAB has a strong positive magnitude. Not to rehash all of the findings from the results section, but it is important to emphasize how different the capabilities of the

ID measures are in predicting proficiency levels of learners for two languages in the same category.

For the category II languages the patterns of standardized β coefficients for the two languages are more similar than those for the category I languages, but there are still notable differences. The predictive power for both the listening and reading models for the German group relies heavily on MK and DLAB. Also the pattern is stable over time. The predictive power for the Indonesian group, on the other hand, appears to rely mainly on WK and DLAB, and the predictor pattern profile is erratic. So, the category II models, once again, provide evidence that the cognitive learning tools necessary to develop foreign language proficiency may vary between languages within the same category although the evidence is not as strong as in the case of the category I languages.

The predictor pattern profiles for Russian and Tagalog, the category III representatives, as in the case of the category I languages, are substantially different. Of the predictors used in this analysis, proficiency in Russian is best determined by PC and DLAB scores. The standardized β coefficient for the PC score is stable over time where the coefficient for DLAB is more variable. In the case of Tagalog, MK and DLAB are the best predictors of proficiency level, and they add significant predictive validity to the Tagalog listening and reading models for the most part.

Finally, the category IV language models also provide evidence to support the arguments of Child (1998) and Lowe (1998). There are some very notable differences especially in the listening models. The prediction of listening proficiency scores in Arabic is heavily dependent on MK and DLAB while the Chinese models are more reliant on AR and DLAB. The reading models also reflect this weighting. More interestingly, however, is that the direction and magnitudes of the standardized β coefficients for the two language group models appear to converge by the fourth testing cycle. This may indicate that the long term maintenance of the two category IV languages is based on similar cognitive factors, general cognitive reasoning ability (AR) and foreign language aptitude (DLAB), even though earlier proficiency levels were predicted by different measures across the two languages.

The analysis of the individual languages highlights the fact that predictive models of proficiency are not consistent within language category. The individual languages are also not consistent across categories. This provides evidence to support a model of differential impact of cognitive factors on foreign language learning that is dependent on the language to be learned. In essence, this is in line with the arguments of Child (1998) and Lowe (1998). Child (1998) parses languages based on "distance" from English. Child (1998) argues that languages should be matched according to language difficulties based on phonology, with provision made for written representation, grammatical system covering morphology and syntax, and semantics. He says each of these three should be rated with respect to their distance from English, and then languages should be compiled into their language category based on these distances. He also discusses how "learning difficulty is tied to the degree in which the object of learning resembles something already known" (Child, 1998, p.6), and suggests that aptitude tests should be designed to predict success in a specific language or language category (once the languages are aligned in their new categories). He states that aligning the predictor to the criterion is the best way to identify aptitude for a particular language. The current research supports the argument that certain predictors align better with certain languages. Further research is still needed to determine if these predictors are in some way indicative of this “distance” from English to which Child (1998) refers. One additional factor to examine is the differential effect that motivation may have across languages. Motivation levels may have to be higher in the harder languages for students to succeed relative to “easier” languages.

Chapter 8: Summary and Conclusions

This intent of this research was to investigate the ability of predictive measures to differentiate levels of language proficiency among learners across language categories and learning contexts. The findings here support the claim that the performance of language learning predictive measures is influenced by both language category and learning context. Additionally, this research provides evidence that predictor profiles of language learning success vary across individual foreign languages. Based on these findings future research should be done to determine if a re-categorization of languages would better align predictor success within the language category structure.

In examining the models of the three contexts in this study where the outcome measure is foreign language GPA, differences are readily apparent between the groups (see Figure 44, below). First, the predictive ability of SATM appears to demonstrate that higher aptitude math scores tend to be detrimental to achieving a higher GPA in foreign language for USNA students but higher quantitative scores appear to be beneficial for DLI students. This would seem to indicate differences in the language programs or language learning focus of the students between the two institutions. Second, the differences between the domestic programs and the study abroad group in predictor significance indicate a heavier reliance on L1 abilities in the domestic environments. This may indicate explicit instruction in a foreign language has a greater benefit to students with greater L1 verbal aptitude. Third, the DLAB demonstrates greater success in distinguishing rates of learning at lower proficiency levels in a classroom environment. Additionally, this finding may indicate that the DLAB measures an aptitude for language learning in an explicit language instruction environment since the participants in the SA group are taking content classes in the foreign language, but do not receive explicit language-focused instruction.

Figure 44. Summary of Results: Effects of Context on Predictive Ability of Individual Difference Measures in SLA

Summary of Results: Effects of Context

Outcomes	Predictors	FLC	INI	SA
Foreign Language GPA	GREQ/SATM	-	+*	-
	GREV/SATV	+*	+*	-
	DLAB	+*	-	+
	CQPR	+*	+	+*
Listening DLPT[1]	GREQ/SATM	-	-	+
	GREV/SATV	+	+	-
	DLAB	+√	-	+
	CQPR	+	+	+√
Reading DLPT	GREQ/SATM	+	-	+
	GREV/SATV	+	+	-
	DLAB	+	-	-
	CQPR	+	+	+

+/- indicates directional trend of the predictor coefficient

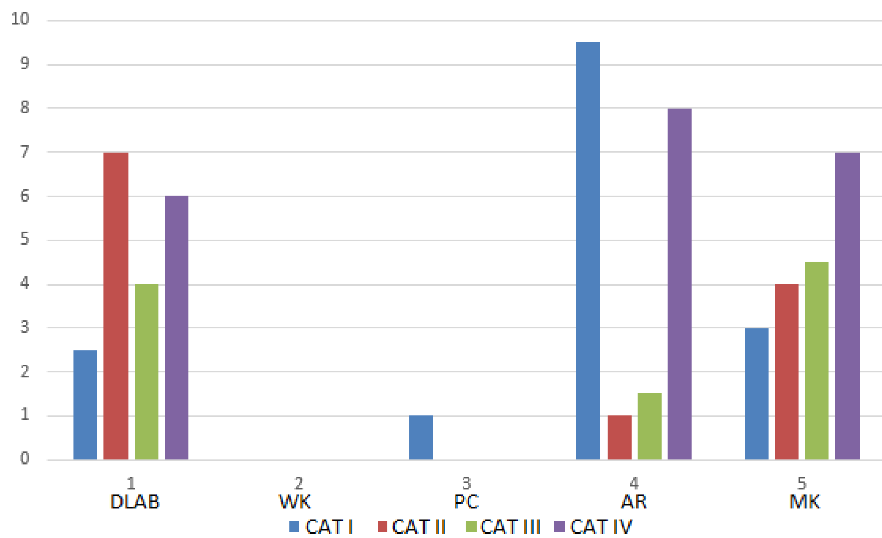
* indicates significant predictor / √ indicates marginal significance

[1] - marginal significance found post hoc

Next, using the DLPT scores as outcome measures, the trend in direction of the predictors is similar to the foreign language GPA models with the exception of SATM/GREQ. A smaller N-size and the lack of a significant impact on the models could explain this difference. The predictive direction of the verbal scores, however, continues to demonstrate that L1 verbal skills positively impact the domestic (classroom) programs, but not study abroad. DLAB scores again have their largest impact on the FLC group and little to no impact on the other two groups. Finally, undergraduate GPA has a positive influence across the board. The lack of granularity in combination with the small n-size for the DLPT measures limits the ability of the models to demonstrate more definitive differences in the context groups, but the fact that the trends are similar to the foreign language GPA outcome offers some confidence in the findings. In summary, Chapter 5 challenges the default assumption that aptitude and other individual difference measures ought to be context independent by providing evidence to the contrary.

A similar story surfaces when examining the effects of language category on the predictive patterns of ID measures. Starting with FL GPA as the outcome measure, differing predictor pattern profiles immediately emerge. These differences are further supported by the DLPT and OPI outcome models. For example, AR, a measure of general cognitive reasoning ability, is a significant predictor of nearly all outcomes for language category I and IV learners only. Another predictor, MK, is a significant predictor mainly for the category IV language learners. DLAB has some predictive ability across language categories, but its success is focused in the category II and IV models. Figure 45, below, demonstrates the relative occurrence of significance for each predictor by language category.

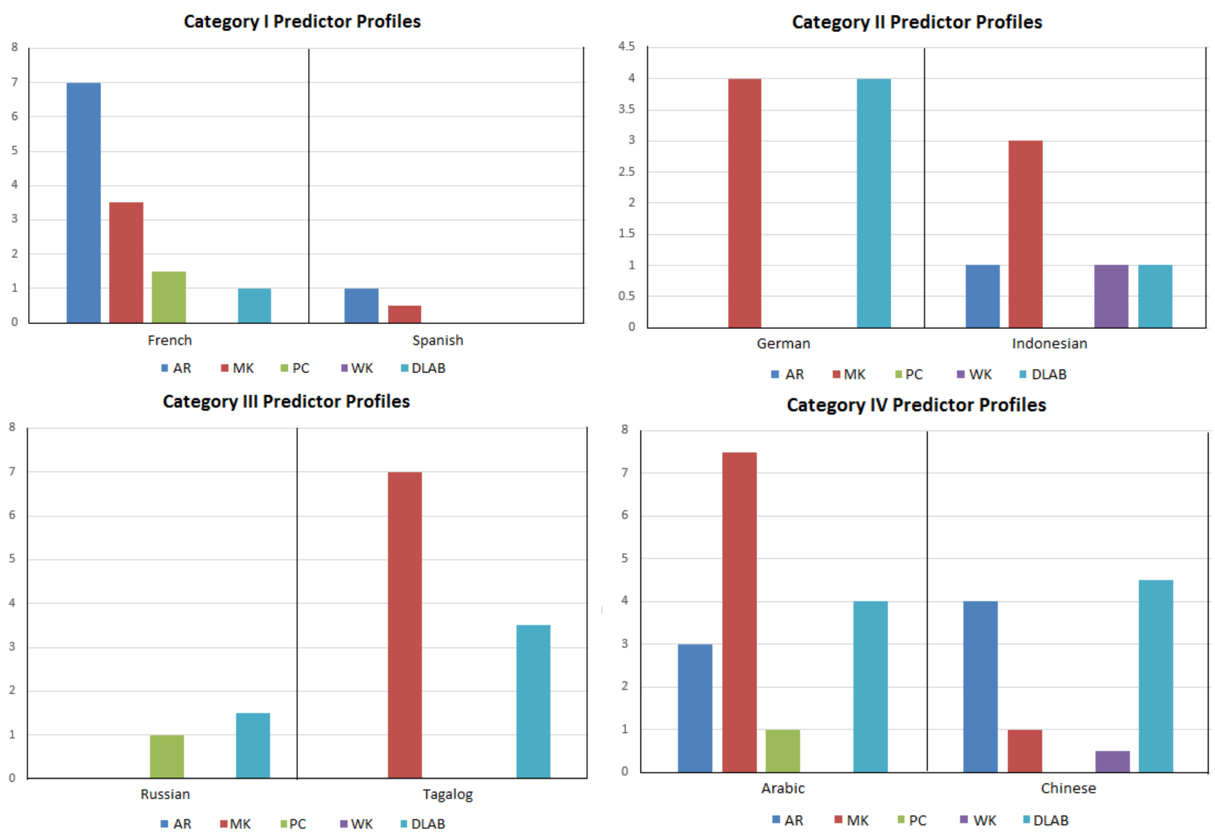
Figure 45. Summary of Occurrence in Predictive Models of Individual Difference Measures in SLA by Language Category



Occurrence of each predictor was determined by its level of significance in each of the ten models of foreign language performance analyzed in Chapter 7. Clearly, the predictive ability of the individual difference measures presented in this study is dependent on language category. Additionally, the detailed analyses of the magnitude and direction of the standardized β coefficients for the models presented in Chapter 7 provide further evidence to support this claim. Finally, distance from English appears to stand out as a valid explanation when examining the attrition models since higher scores on measures that are associated with the L1 (PC and WK) slow attrition in the category I and II models while speeding it in the category III and IV models. Also, higher scores on the general cognitive measures (AR and MK) slow attrition in the category IV models where L1 linkages are less readily available.

Next, the analysis of the individual languages further highlights the fact that predictive models of proficiency are not consistent within language category. The results presented in Chapters 6 and 7 provide evidence of the differential impact of cognitive factors on foreign language learning that is dependent on the individual language. Figure 46 depicts the summary of occurrences of the individual difference

Figure 46. Summary of Occurrence in Predictive Models of ID Measures in SLA by Individual Language



measures in the language predictor models. Each time a predictor was significant in one of the ten outcomes analyzed, a value of one was added to the number of occurrences. A value of 0.5 was added for each time a predictor had a marginally significant impact on a model. More evidence is added when examining the individual languages that show a trend towards attrition over the four testing cycles, and exploring the individual differences that affect attrition in the language category models. Here, distance from English appears to stand out as a valid explanation for the rates of attrition since higher scores on measures that are associated with the L1 (PC and WK) slow attrition in the category I and II models while speeding it in the category III and IV models. Also, higher scores on the general cognitive measures (AR and MK) slow attrition in the category IV models where L1 linkages are less readily available.

This last point demonstrates the possibility to better align learners with specific abilities into languages that are more demanding of those abilities. For example, symbolic assembly, as defined earlier and measured by AR, is an ability that assists the learner in aligning parametrically based memory items within a protocol, such as a syntactic structure, that is difficult for the learner to transfer from the L1 to the L2. This skill is a demonstrated, general cognitive ability that allows for the mental manipulation of items into a rule guided structure. Therefore, it is more relevant in languages where those patterns are more difficult to induce. The difficulty in inducing those patterns may be due to stress pattern differences, as is the case in French, or larger grammatical differences as found in languages like Chinese or Arabic. AR has shown relatively high correlations with measures of inductive reasoning, so this argument is logical (see Alderton et al., 1997, for correlations).

Therefore, languages like Turkish, Swahili, and Thai would also be expected to exhibit a larger reliance on symbolic assembly.

In summary, this research provides evidence to support the findings of Silva & White (1993) that the DLAB adds incremental predictive validity, in most cases, to the more general cognitive measures used on the ASVAB. It also provides some evidence to support the claims of Linck et al. (2012) that measures like the DLAB are better indicators of rates of learning in a classroom environment. To clarify their claims, however, the DLAB also adds incremental predictive validity to intermediate learner proficiency models as well and continues to serve its original purpose by predicting results at DLI graduation. In addition, this research analyzes the current DOD language categorization system, examines individual language proficiency models, and tests the performance of ID measures in several different learning contexts. As a result, this research provides evidence to support new claims that the predictive validity of individual difference measures in language learning proficiency models is dependent on language category, learning context, and the individual language being learned. Finally, this research fills a gap in the literature concerning the current language categorization system and calls for additional research to redefine the language categories and pair them with improved predictive measures of language learning success.

Therefore, these findings in combination with the literature imply that a future measure of language learning aptitude (i.e., DLAB III) should be flexible enough to account for the particular foreign language and the learning context in order to optimize predictive validity. It should include additional measures of general cognition that have been overlooked in the development of earlier foreign language aptitude measures. Among these overlooked measures are indicators of symbolic assembly and logical reasoning (AR and MK). The inclusion of the additional measures follows similar logic to that behind the progression from the DLAB I to the DLAB II, where personality and motivation were added to the battery. For the DOD in particular, increased predictive validity of aptitude measures leads to large savings in required resources (Welsh et al., 1990).

Additionally, Welsh et al. (1990) in their review of the ASVAB explain how the DOD builds “Occupational Composites” from the 10 subtest scores of the ASVAB. The “Occupational Composites” are the most predictive combination of the subtests for success in a particular occupational specialty. Subsequently, occupations are then grouped by the occupational composites that predict them. This method of organizing occupational specialties into clusters follows from the theory of differential classification (Brogden, 1955). Similarly, Linck et al. (2012) look for the best combination of language learning aptitude components to predict high level language success. Using this same logic, then, a foreign language aptitude measure that maintains it componential nature could be assembled into composite scores. Languages could be categorized by the aptitude components that best predict learner success in the language. In other words, the components of the new DLAB should be sculpted into composite scores matched to a particular language or group of languages. Once the language categories are established, composites that predict those categories could also be adjusted to find the most predictive sub-composites of the desired outcome via the prescribed learning context. In conclusion, this research calls for a new language aptitude battery developed in concert with the restructuring of the DOD language categorization system to provide a more meaning tool for DOD language selection protocols.

Appendix A

The specific answers to the research questions proposed in Chapter 3 are addressed here.

- **Research Question 1: Do individual differences in verbal, quantitative, and foreign language aptitude and individual differences in L1 skills predict differences in foreign language proficiency in a foreign language classroom (FLC) environment?** Yes. When all languages are grouped together, SAT Verbal, DLAB scores, and undergraduate GPA predict differences in Foreign Language GPA. Examining two individual languages, Arabic and Chinese, shows that English Composition grades also predict Foreign Language GPA. Additionally, SAT Math predicts Foreign Language GPA for Chinese majors. Foreign Language GPA, subsequently, demonstrates some success in predicting reading and listening DLPT scores. English Composition grades are also successful in predicting reading DLPT scores for Arabic majors.

- **Research Question 2: Do individual differences in verbal, quantitative, and foreign language aptitude and individual differences in L1 skills predict differences in foreign language proficiency in an intensive instruction (INI) environment?** Yes, SAT Math and SAT Verbal are able to predict differences in Foreign Language GPA for the INI students. The measures, however, are unable to predict differences in the DLPT scores. In

Chapter 7, DLAB scores, Math Knowledge scores, and Arithmetic Reasoning scores, generally, are strong predictors of foreign language proficiency using FL GPA, reading and listening DLPT scores, and OPI scores.

- **Research Question 3: Do individual differences in verbal, quantitative, and foreign language aptitude and individual differences in L1 skills predict differences in foreign language proficiency after a semester study abroad (SA)?** Verbal, quantitative, and foreign language aptitude do not predict differences in foreign language proficiency. Undergraduate GPA, on the other hand, does predict differences in Foreign Language GPA and is marginally successful at predicting listening DLPT scores.

- **Research Question 4: Do the magnitudes of the coefficients in predictor models of language success vary for the independent variables (L1 skills, measures of verbal, quantitative, and foreign language aptitude) in different learning environments?** Yes. Not only do the coefficients vary in magnitude between the context models, but also in polarity (direction). See Figure 44.

- **Research Question 5: Do individual differences in verbal, quantitative, and foreign language aptitude and individual differences in L1 achievement differ in how they predict foreign language proficiency based on language category?** The most notable difference is with the

Arithmetic Reasoning measures which predicts success in the Category I and IV models mainly. DLAB scores are better predictors in the Category II and IV models. Finally, Math Knowledge scores best predict for Category IV models.

- **Research Question 6: Do the magnitudes of the coefficients in predictor models of language success vary for the independent variables (L1 skills, measures of verbal, quantitative, and foreign language aptitude) within the same language category?** Yes. Reference Figure 46. The ID measures vary substantially across the languages. Within each language category there are large variations in the coefficient magnitudes and their directions (polarity). For more specific information on the magnitudes, see Tables 65, 67, 70, 72, 75, 77, 80, and 82.

- **Research Question 7: Do the patterns in the magnitudes of the coefficients of achievement measures, measures of verbal, quantitative, and foreign language aptitude and foreign language proficiency vary across language categories?** Yes. Again reference Figure 46. Although the figure is not specifically based on magnitude and direction of the coefficients, it quickly demonstrates the predictors that are significant for each of the languages and language categories. For more specific information on the magnitudes, see Tables 65, 67, 70, 72, 75, 77, 80, and 82.

Appendix B

Descriptive Statistics for the DLPT Listening and Reading Scores for each study.

Study 1 (Chapter 5) DLPT Descriptive Statistics

	Listening DLPT			Reading DLPT		
	N	Mean	SD	N	Mean	SD
FLC	47	18.68	9.22	47	19.79	7.72
INI	144	23.28	6.08	145	26.22	5.43
SA	57	15.02	10.28	57	16.53	9.57

Study 2 (Chapter 6) DLPT Descriptive Statistics

	Listening DLPT			Reading DLPT		
	N	Mean	SD	N	Mean	SD
Arabic	39	8.97	9.00	39	12.80	8.90
Chinese	42	15.21	7.46	42	15.20	7.50

Study 3 (Chapter 7) DLPT Descriptive Statistics (Listening)

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
French	197	24.4 (4.5)	23.3 (5.3)	23.7 (5.1)	23.4 (5.0)
Spanish	198	23.8 (4.7)	23.7 (5.1)	24.4 (4.7)	24.7 (4.6)
German	116	19.9 (5.1)	19.7 (5.2)	20.4 (4.5)	20.5 (5.0)
Indonesian	104	26.5 (3.7)	25.4 (4.1)	25.5 (3.3)	24.8 (2.6)
Russian	190	23.7 (4.7)	22.9 (4.3)	23.0 (4.6)	23.5 (3.7)
Tagalog	186	23.6 (4.2)	22.9 (4.3)	23.0 (4.6)	23.5 (3.7)
Arabic	197	22.1 (5.8)	20.3 (6.5)	20.3 (6.6)	20.2 (5.8)
Chinese	201	23.1 (4.9)	21.6 (5.7)	21.5 (6.2)	21.9 (5.8)

Standard Deviations in parenthesis.

Study 3 (Chapter 7) DLPT Descriptive Statistics (Reading)

Language	N	DLPT1	DLPT2	DLPT3	DLPT4
French	197	24.8 (5.0)	23.3 (5.5)	24.5 (5.1)	24.4 (4.6)
Spanish	198	27.5 (3.3	27.1 (4.0)	27.0 (4.5)	27.7 (3.5)
German	116	27.4 (4.9)	25.8 (6.0)	27.0 (4.9)	26.4 (5.5)
Indonesian	104	26.0 (3.2)	24.7 (3.9)	24.9 (2.9)	25.2 (1.9)
Russian	190	24.8 (4.5)	24.4 (5.1)	24.5 (5.1)	24.9 (4.5)
Tagalog	186	23.9 (4.1)	23.2 (4.1)	23.2 (4.0)	23.4 (3.7)
Arabic	197	23.3 (5.3)	21.6 (5.9)	21.3 (6.0)	21.3 (5.8)
Chinese	201	25.7 (4.3)	23.2 (5.3)	22.8 (5.8)	22.8 (5.7)

Standard Deviations in parenthesis.

Bibliography

Alderton, D.L., Wolfe, J.H., & Larson, G.E. (1997). The ECAT Battery. Military Psychology, 9, 5–37.

ASVAB Career Exploration Program (2011). The ASVAB Career Exploration Program: Theoretical and Technical Underpinnings of the Revised Skill Composites and OCCU-Find. Department of Defense. See www.asvabprogram.com.

ASVAB Technical Bulletin (2012). ASVAB Technical Bulletin No.4; P&P-ASVAB Forms 23-27. March 2009, Revised May 2012. Seaside, California: Personnel Testing Division, Defense Manpower Data Center.

Atkinson, D. (2002). Toward a sociocognitive approach to second language acquisition. Modern Language Journal, 86, 525-545.

Ayers, J. and Quattlebaum, R. (1992). TOEFL performance and success in a master's program in engineering. Educational and Psychological Measurement, 52, 973-975.

Bamford, K.W., & Mizokawa, D.T. (1991). Additive-bilingual (immersion) education: Cognitive and language development. Language Learning, 41, 413-429.

Batstone, R. (2002). Contexts of engagement: A discourse perspective on "intake" and "pushed output." System, 30, 1-14.

Brecht, R. D., Davidson, D. E., & Ginsberg, R. B. (1993). Predictors of foreign language gain during study abroad. Washington D.C.: The National Foreign Language Center.

Brogden, H. E . (1955) . Least-squares estimates and optimal classification. Psychometrica, pp. 249-252.

Brunner, J. and Austin, P. (2009). Inflation of Type I error rate in multiple regression when independent variables are measured with error. The Canadian Journal of Statistics, 37, 33-46.

Bygate, M. (1988). Units of oral expression and language learning in small group interaction. Applied Linguistics, 9, 59-82.

Carroll, J. (1967). Foreign language proficiency levels attained by language majors near graduation from college. Foreign Language Annals, 1, 131–151.

Carroll, J. (1973). Implications of aptitude test research and psycholinguistic theory for foreign language teaching. International Journal of Psycholinguistics, 2, 5–14.

Carson, J., et al. (1990). Reading-writing relationships in first and second language. TESOL Quarterly, 24 (2), 245-266.

Child, J. (1998). Language aptitude testing: learners and applications. Applied Language Learning, 9, 1-10.

Cho, Y. and Bridgeman, B. (2012). Relationship of TOEFL iBT scores to academic performance: Some evidence from American universities. *Language Testing, 29(3),* 421-443.

Collentine,J. (2004). The effects of learning contexts on morphosyntactic and lexical development. Studies in Second Language Acquisition, 26, 227-248.

Collentine,J., & Freed, B. (Eds.). (2004). Learning context and its effects on second language acquisition [Special Issue]. Studies in Second Language Acquisition, 26, 153-171.

Cooper, T.C. (1987). Foreign Language Study and SAT-Verbal Scores. *Modern* Language Journal, 71(4): 381-387.

Corder, S.P. (1983). Strategies of communication. In C. Faerch & G. Kasper (Eds.), Strategies in interlanguage communication. (pp. 15-10). New York, NY: Longman, Inc.

Defense Language Institute Foreign Language Center (2013). *Defense Language* Proficiency Test 5 System: Familiarization Guide for Multiple-Choice *Format.* Retrieved from http://www.dliflc.edu/archive/documents/.

DeKeyser, R. (1986). From learning to acquisition? Foreign language development in a US classroom and during a semester abroad. Stanford, CA: Stanford University dissertation.

DeKeyser, R. (1991). Foreign language development during a semester abroad. In B. F. Freed (Ed.), Foreign language acquisition: Research and the classroom (pp. 104–119). Lexington, MA: D.C. Heath.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. Studies in Second Language Acquisition, 22, 499–533.

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams, Theories in second language acquisition (pp. 97–113). Mahwah, NJ: Erlbaum.

DeKeyser, R. (2010). Monitoring processes in Spanish as a second language during a study abroad program. Foreign Language Annals, 43, 80-92.

Dewey, D. (2004). A comparison of reading development by learners of Japanese in intensive domestic immersion and study abroad contexts. Studies in Second Language Acquisition, 26, 303-327.

Dewey, D. P. (2008). Japanese vocabulary acquisition by learners in three contexts. Fontiers: The Interdisciplinary Journal of Study Abroad, 15, 127-148.

Diaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. Studies in Second Language Acquisition, 26, 249- 274.

Dufon M. and Churchill, E. (2006). Language learners in study abroad contexts. Clevedon, UK: Multilingual Matters.

Dörnyei, Z. (2005). The Pyschology of the Language Learner: Individual Differences in Second Language Acquisition. Mahwah, NJ: Lawrence Erlbaum Associates.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty, & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Oxford: Blackwell.

Educational Testing Service (2013). *Test Fairness and Validity.* Retrieved from http://www.ets.org/gre/institutions/about/fairness/

Ellis, R. (1994). Second language acquisition. Oxford: Oxford University Press.

Faerch, C. and Kasper, G. (1983). Plans and strategies in foreign language communication. In C. Faerch & G. Kasper (Eds.), Strategies in interlanguage communication. (pp. 20-60). New York, NY: Longman, Inc.

Feyten, C.M. (1991). The power of listening ability: an overlooked dimension in language acquisition. The Modern Language Journal, 75, 173-180.

Freed, B. F. (1990). Language learning in a study abroad context: The effects of interactive and noninteractive out-of-class contact on grammatical achievement and oral proficiency. In J. E. Alatis (Ed.), Georgetown University Round Table on Languages and Linguistics: Linguistics, language teaching,and language acquisition: The interdependence of theory, practice, and research (pp. 459–477). Washington, DC: Georgetown University Press.

Freed, B. F. (1991). Current realities and future prospects in foreign language acquisition research. In B.F. Freed (Ed.), Foreign language acquisition research and the classroom (pp. 3-27). Lexington, Massachusetts: D.C. Heath and Company.

Freed, B. F. (1995). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), Second language acquisition in a study abroad context (pp. 123–148). Amsterdam: Benjamins.

Freed, B. F. (1998). An overview of issues and research in language learning in a study abroad setting. Frontiers, 4, 31–60.

Freed, B. et al. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. Studies in Second Language Acquisition, 26, 275-301.

Follmann, D. (2005). Subgroups and Interactions. In N. L. Geller (Ed.), Advances in clinical trial biostatistics (pp. 121-139). New York: Marcel Dekker, Inc.

Ginsburg, R. B., & Miller, L. (2000). What do they do? Activities of students during study abroad. In R. D. Lambert & E. Shohamy (Eds.), Language policy and pedagogy: Essays in honor of A. Ronald Walton (pp. 237–261). Philadelphia: Benjamins.

Humes-Bartlo, M. (1989). Variation in children's ability to learn second languages. In Hyltenstam, K., and Obler, L., editors, *Bilingualism across the lifespan*. Cambridge: Cambridge University Press, 41–54.

Institute for International Education. (2007). Open doors online: Report on international educational exchange. Retrieved January 10, 2008, from http://opendoors.iienetwork.org/

Interagency Language Roundtable. (1985). *Interagency Language Roundtable Skill Level Descriptions.* Retrieved from http://www.govtilr.org/.

Isabelli-García, C. L. (2006). Study abroad social networks, motivation, and attitudes: Implications for SLA. In M. DuFon & E. Churchill (Eds.), Language learners in study abroad contexts (pp. 231–258). Clevedon, UK: Multilingual Matters.

Keating, K. (1994). Findings from a qualitative study of adult learners of Spanish: Implications for approaches to second language programs for future teachers. Journal of Instructional Psychology, 21, 57-63.

Kinginger, C. (2008). Language learning in study abroad: Case studies of Americans in France. Modern Language Journal, 92, Monograph.

Lafford, B. (1995). Getting into, through and out of a survival situation: a comparison of communicative strategies used by students studying Spanish abroad and 'at home.' Second language acquisition in a study abroad context, ed. by B. Freed, 97-121. Philadelphia: John Benjamins.

Lafford, B. (2004). The effect of context of learning on the use of communicative strategies by learners of Spanish as a second language. Studies in Second Language Acquisition, 26, 201-226.

Lafford, B. (2006). The effects of study abroad vs. classroom contexts on Spanish SLA: Old assumptions, new insights and future research directions. In C. A. Klee & T. L. Face (Eds.), Selected proceedings of the 7th conference on the acquisition of Spanish and Portuguese as first and second languages (pp. 1-25). Somerville, MA: Cascadilla Proceedings Project.

Lantolf, J. (1994). Sociocultural theory and second language learning: introduction to the special issue. Modern Language Journal, 78, 418-420.

Lantolf, J. (2000). Sociocultural theory and second language learning. Oxford: Oxford University Press.

Lantolf, J. and Appel, G. (1994). Vygotskian approaches to second language research. Norwood, NJ: Ablex.

Lett, J.A. and O'Mara, F.E. (1990). Predictors of success in an intensive foreign language learning context: correlates of language learning at the Defense Language Institute Foreign Language Center. In Parry, T. and Stansfield, C. (eds.) *Language Aptitude Reconsidered.* Englewood Cliffs, NJ: Prentice Hall, 222-260.

Li, S. (2015). The Associations Between Language Aptitude and Second Language Grammar Acquisition: A Meta-Analytic Review of Five Decades of Research. Applied Linguistics 36, 3, 385-408.

Linck, J., et al. (2012). Hi-LAB Identifies Successful High-proficiency Language Learners: Results of the Group Discrimination Study. Technical Report: Center for Advanced Study of Language. University of Maryland, College Park.

Lohman, D.F. (2004). Aptitude for College. In R. Zwick (ed.), *Rethinking the SAT:* The Future of Standardized Testing in University Admissions. New York, NY: Routledge Falmer.

Long, M. (1997). Construct validity in SLA research: A response to Firth and Wagner. Modern Language Journal, 81, 318-323.

Lowe, P. (1998). Zero-based language aptitude test design: where's the focus for the test? Applied Language Learning, 9, 11-30.

Markee, N. (2000). Conversation Analysis. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Martinsen, R. A. et al. (2010). Exploring diverse settings for language acquisition and use: Comparing study abroad, service learning abroad, and foreign language housing. Applied Language Learning, Vol. 20, Nos. 1 & 2, 45-69.

Maxwell, J. A. (1996). Qualitative research design: an interactive approach. Thousand Oaks, CA: Sage Publications, Inc.

Miyake, A. and Friedman, D. (1998). Individual differences in second language proficiency: working memory as language aptitude. In A. Healy and L. Bourne (eds), Foreign Language Learning: Psycholinguistic Studies on *Training and Retention.* Mahwah, NJ: Lawrence Erlbaum Associates.

O'Malley, J. M. & Chamot, A. U. (1990). Learning Strategies in Second Language Acquisition. Cambridge: Cambridge University Press.

Petersen, C. & Al-Haik, A. (1976). The development of the Defense Language Aptitude Battery (DLAB). Educational and Psychological Measurement, 36, 369-380.

Pica, T. (1987). Second language acquisition, social interaction and the classroom. Applied Linguistics, 8, 3-21.

Pimsleur, P. (1996). Testing foreign language learning. In Valdman, A., editor, *Trends in language teaching*. New York: McGraw Hill, 175–214.

Pimsleur, P. (1968). Language aptitude testing. In Davies, A., editor, *Language* aptitude symposium: A linguistic approach. New York: Oxford University Press, 98–106.

Robinson, P. (2013). Aptitude in second language acquisition. In Chapelle, C.A., editor, The Encyclopedia of Applied Linguistics. Oxford: Blackwell Publishing Ltd., 1-5.

Rodriguez, S. (2001). The perception of requests in Spanish by instructed learners of Spanish in the second- and foreign-language contexts: a longitudinal study of acquisition patterns. Bloomington: Indiana University dissertation.

Schmidt, R. (1990). The role of consciousness in second language learning. Applied Linguistics, 11, 129-158.

Schmidt, R. (1993). Awareness and second language acquisition. Annual Review of Applied Linguistics, 13, 206-226.

Schoonen, R., Hulstijn, J. & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension. An empirical study among Dutch students in grades 6, 8 and 10. *Language Learning*, 48(1), 71-106.

Segall, D. and Moreno , K. (1999). Development of the CAT-ASVAB. In F. Drasgow & J. B. Olson-Buchanan (Eds.). Innovations in Computerized Assessment (pp. 35—65). Hillsdale, NJ: Lawrence Erlbaum Associates.

Segalowitz, N. & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition: learning Spanish in at home and study abroad contexts. Studies in Second Language Acquisition, 26, 173-199.

Selinker, L., & Douglas, D. (1985). Wrestling with "context" in interlanguage theory, Applied Linguistics, 6, 190-204.

Silva, J. M. & White, L. A. (1993). Relation of cognitive aptitudes to success in foreign language training. *Military Psychology, 5,* 79-93.

Simões, A. (1996). Phonetics in second language acquisition: an acoustic study of fluency in adult learners of Spanish. Hispania, 79, 87-95.

Skehan, P. (1986). Cluster analysis and the identification of learner types. In V. Cook (ed.), Experimental Approaches to Second Language Acquisition. Oxford: Pergamon Press, 81-94.

Skehan, P. (1989). Individual Differences in Second Language Learning. London: Edward Arnold.

Skehan, P. (1991). Individual differences in second language learning. *Studies in* Second Language Acquisition, 13, 275–98.

Sparks, R. and Ganschow, L. (1991). Foreign language learning difficulties: affective or native language aptitude differences? *Modern Language Journal,* 75, 3–16.

Sparks, R. and Ganschow, L. (1993). Searching for the cognitive locus of foreign language learning difficulties: linking native and foreign language learning. *Modern Language Journal,* 77, 289–302.

Sparks, R. and Ganschow, L. (1995). A strong inference approach to causal factors in foreign language learning: a response to MacIntyre. *Modern Language Journal* 79, 235–244.

Sparks, R., et al. (1998). Differences in native-language skills, foreign-language aptitude, and foreign language grades among high-, average-, and low-proficiency foreign language learners: two studies. *Language Testing,* 15 (2), 181-216.

Stanhope, D., & Surface, E. (2014). Examining the Incremental Validity and Relative Importance of Specific Cognitive Abilities in a Training, Journal of Personnel Psychology, 13(3), 146-156.

Stevens, J. (2001). The acquisition of L2 Spanish pronunciation in a study abroad context. Los Angeles: University of Southern California dissertation.

Surface, E. A., Dierdorff, E. C., & Donnelly, J. (2004). Modeling second language proficiency change for U.S. Special Operations personnel. Paper presented at the 19th annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

Tarone, E. (1994). Interlanguage. In R. Asher & S. Simpson (Eds.), Encyclopedia of Language and Linguistics (Vol. 4, pp. 1715-1719). Oxford: Pergamon.

Tarone, E. (2007). Sociolinguistic Approaches to Second Language Acquisition Research-1997-2007. *The Modern Language Journal*, 91, Focus Issue, 837-848.

Tarone, E. & Swain, M. (1995). A sociolinguistic perspective on second language use in immersion classrooms. *The Modern Language Journal*, 79, 166-178.

Torres, J. (2003). A cognitive approach to the acquisition of clitics in L2 Spanish: insights from study abroad and classroom learners. Ithaca, NY: Cornell University dissertation.

Vandergrift, L. (2006). Second language listening: listening ability or language proficiency. The Modern Language Journal, 90 (1), 6-18.

Wagener, T. (2014). Aptitude and achievement measures as predictors of growth in second language proficiency. Unpublished.

Welsh, J., Kucinkas, S., and Curran, L. (1990). Armed Services Vocational Battery (ASVAB): Integrative Review of Validity Studies. San Antonio, Texas: Operational Technologies Corporation.

Wesche, M. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (ed.), Individual Differences and Universals in Language Learning Aptitude. Rowley, MA: Newbury House, 119-54.

Zwick, R. (2004). Rethinking the SAT: The Future of Standardized Testing in *University Admissions.* New York, NY: RoutledgeFalmer.