2019

Second Language Capability in the Army Linguist Community

Kenneth John Hutchinson
Walden University

Walden University

College of Social and Behavioral Sciences

This is to certify that the doctoral dissertation by

Kenneth Hutchinson

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Marcel Kitissou, Committee Chairperson,
Public Policy and Administration Faculty

Dr. Kent Sieg, Committee Member,
Public Policy and Administration Faculty

Dr. Olivia Yu, University Reviewer,
Public Policy and Administration Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2019

Abstract

Second Language Capability in the Army Linguist Community

by

Kenneth John Hutchinson, Sr.

MA, American Military University, 2011

BS, American Military University, 2009

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Management and Leadership

Walden University

February 2019

Abstract

Developing the ability to understand one's adversary is a critical task for any professional in the military. In the army, this understanding is partially realized through a second language capability. This study involved quantifying results from different sites and methods of second language training for army linguists, using proficiency scores measured by the Defense Language Proficiency Test to determine if quantitative differences between methods of instruction existed. The hypothesis that trainees at Site 1 achieved significantly higher proficiency levels than trainees at Site 2 was used as the primary building block for this research. The objective of this study was to aid leaders in the linguist community in making evidence-based policy decisions. Social representation theory was used as the theoretical framework for understanding the norms and beliefs formed by the subgroups of linguists within the target population who were beneficiaries of the different methods of second language training. Secondary data were obtained through a Freedom of Information Act request to the major army command used as the target population for the study. The study found no significant difference between the language training sites as measured by the posttraining Defense Language Proficiency Test. Further study recommendations consist of investigating qualitative aspects of second language training. The primary social change impact of this research for the army linguist community may reside in the study's promotion of the best use of resources. To fulfill the national security role presented to the army, the most effective and efficient methods of second language training must be used.

Second Language Capability in the Army Linguist Community

by

Kenneth John Hutchinson, Sr.

MA, American Military University, 2011

BS, American Military University, 2009

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Management and Leadership

Walden University

February 2019

Dedication

This work is dedicated to my children in the hope to show them anything is possible.

Acknowledgments

I would like to begin by thanking my dissertation committee. Dr. Kitissou was the right fit to guide me through this work, understanding that language is more than words and syntax and that a real understanding resides in the understanding of culture. Dr. Sieg's attention to detailed, thoughtful, and challenging review resulted in higher quality work.

I also need to thank the unsung heroes who go about their daily duties to make this world and our nation safer, our army linguists. This exceptional group of men and women indeed is at the edge of our nation's defense, and without their keen understanding and ability to put the languages of our many advisories in context, our country might not enjoy the safety that we currently live in today.

My wife is my rock, without whom I could never achieve anything close to this level of professional and academic success.

Lastly, I have to thank the professional team of dedicated army civilians who helped by motivating me, providing data, and supporting me over the past years. Mr. Perkins, the Deputy G-2 for U.S. Forces Command, presented the push that helped this study flourish.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Through this study, I sought to quantify the results of different methods of second language (L2) training for army linguists by comparing their proficiency scores measured by the Defense Language Proficiency Test (DLPT), which was used to determine if there was a significant difference between methods of training. The social change aspect of this study came in the form of improved readiness for the army and the best utilization of funding. The findings may help to build a foundation for army readiness based on training. Without a training strategy, a commander can realize real reductions in the ability to meet the mission. In times of conflict, this reduction may cost lives. There are many elements of readiness, which range from a soldier's ability to perform individual task to the largest of collective tasks. A discussion on second language training in the army must first establish an understanding of how the army trains.

This chapter begins with the problem statement and purpose for the study. Continuing beyond the introduction, this section provides information on the framework for the study by presenting the research question, hypothesis, and theoretical framework. Finally, the chapter expands on the significance, assumptions, and limitations of the research.

Background of the Study

There is a body of research on L2 acquisition and sustainment within the wider language education community. Major themes found in the research are training, testing, proficiency, theories on acquisition and sustainment strategy, and the abstract concept of learning a foreign language. The primary gap that I sought to address in the study

concerned the best strategy for the sustainment of an L2 capability by a professional military linguist population. The study found that neither method of language training had an impact on the proficiency levels of the trainees by quantitatively comparing two different training sites used by the army to provide L2 training. These sites differed in cost, teacher selection, and location. The funding associated with the L2 training program is limited, and this body of work may aid in the validation of the best utilization of funding. There also exists a social belief that Site 1 produces a higher level of proficiency, which this study aimed to validate through statistical analysis.

A critical public policy concern for the U.S. is that of national security. Every structured national security policy has an aspect of training. The ability to understand the threats in the current operational environment (OE) in which the U.S. military operates requires a level of language and cultural understanding. The language and cultural training design must also include an assessment strategy. Subjective assessments are common but hold little quantitative value for policy makers.

Problem Statement

The ability to understand one's adversary is a critical task for any professional in the military. In the 2015 National Security Strategy, this crucial task was highlighted with the requirement to monitor global threats to prevent impacts from misunderstandings and poor communication (Obama, 2015). National security is a critical subcomponent of public policy, potentially affecting the entire population of the U.S. There is a problem with current U.S. Army policy that in part meets this public policy requirement. Funding was extremely limited for linguist training within the study organization, with the funding

received in fiscal year 2017 at just half of the level requested. Secondly, the method of training may have an impact on the efficiency of the overall program, which is also impacted by funding. Specifically, there is a misunderstanding of the most effective method and time required for foreign language training (Bauer, Braun, & Muller, 2017). Currently, there are subjective assessments on how different methods of training affect proficiency levels (Stepanoviene, 2016). However, potential misconceptions lead to poor resource management and fail to adhere to the army's principles of training (Beam & Hodges, 2015). To best fill the gap left by the subjective assessments currently being used and to provide a quantitative assessment, a thoughtful study should provide an understanding of methods of training and the formal measures of success for each training model, offering an objective assessment of one training method as superior to another. The literature reviewed for this study identified that language training, as a theme for research studies, is complex (Dulipovici & Vieru, 2015). To truly understand the nature of warfare and the current OE in which the military must operate, it is necessary to understand the culture and language in the area of operations and beyond. Research literature, however, has not addressed professional army linguist training from a social representation perspective. This study addressed this problem by providing data to aid in the formation of public policy decisions and to help change policies on how the army maintains foreign language capability.

Purpose of the Study

In this study, I sought to quantify the results of different methods of L2 training for army linguists by examining their proficiency scores as measured by the DLPT,

which were used to determine whether there was a significant difference between methods of training. Specifically, I sought to conduct research on the multiple types of instruction used for L2 maintenance within the selected major command under the U.S. Army. The proficiency level as measured by the Interagency Language Roundtable (ILR) scale was used as the dependent variable. Among the sites of instruction, there are multiple methods and modes of instruction employed. The first method of instruction involves the use of educators provided by a contract company who have no formal program of instruction, with instruction designed based on student needs. The second method of instruction is provided by a structured educational staff with a set and approved program of instruction. The hypothesis was that there was a significant statistical difference in proficiency levels resulting from the training offered at a Department of Defense (DOD) organized school, which is referred to as Site 1, compared with Site 2, where instruction was provided by contractors.

Research Question and Hypotheses

The research question was the following: Is there a significant difference between the types of training events measured by the posttraining proficiency levels?

Hypothesis: Trainees at Site 1 achieve significantly higher proficiency levels than trainees at Site 2.

Null hypothesis: There is no significant difference between Site 1 and Site 2 in posttraining proficiency levels.

The dependent variable was proficiency levels for calendar years 2014-2018, with an ordinal level of measurement from 0-5, with the + measurement on the ILR Scale

denoted by .6. The independent variable was the location of training, with ordinal level assigned as 1 or 2. Site 1 provided professional education through a DOD higher education formal school, and Site 2 was a full-time contract instruction, with no formal program of instruction, with instruction designed based on student needs.

Theoretical Foundation

Social representation theory explains that within any group, there are multiple subgroups, and these subgroups form assumptions about each other (Moloney & Walker, 2007). This theoretical framework allowed for thoughtful comparisons of the two groups corresponding to the sites in this study. Within the professional linguist community, there are multiple models for training offered by different organizations. The two sites selected for this study represented two organizations that offered different models of instruction. With social representation theory serving as a framework for this study, statistical data could be used to determine if the preconceived notions of many linguists and commanders were correct by ascertaining whether the different models of instruction produced different outcomes. This may allow public policy decision makers to focus limited resources on the program that best meets national security requirements.

The graduates of these programs make up two distinct populations within the army linguist community. Social representation theory allows observers to input these subgroups into a framework and validate the belief by the population that Site 1 offers superior training. With social representation theory as a framework for this study, it is possible to use statistical data to determine if the preconceived notions of many linguists

and commanders are correct by determining if the different models of instruction produce different outcomes.

Nature of the Study

This quantitative study used repeated measurements to determine how methods of training affect linguists' proficiency levels. Using a simple chart developed by the Web Center for Social Research Methods (n.d), I determined that the study fell into the nonexperimental category of research design. To expand on this classification of research design, the time-series study may provide a detailed structure for the collection of data. O'Sullivan, Rassel, Berner, and Taliaferro (2017) detailed how a time-series study should collect data to evaluate trends and determine if there is a pattern.

Secondary data sources used for the study were collected on army linguists in fiscal years 2014 and 2018. The same major command was used for the entire population. The members of the population for the study acquired their L2 capability from the same source.

Definitions

Independent variable (X): Location of training. There were two primary types of location at which the members of the population for this study received training. The first location was a professional education center that focused solely on foreign language instruction, which was serviced by professional educators. The second site was the linguist home station through a contracted civilian language instructor, with instruction performed in a semistructured setting.

Dependent variable (Y): Proficiency levels: This measurement was an ordinal level of measurement from 0-5, with the + measurement on the ILR Scale denoted by .6

Defense Language Proficiency Test (DLPT): The instrument used to provide the above proficiency levels in the listening and reading categories.

Defense Language Aptitude Battery (DLAB) Test: Used to provide a level of aptitude for prospective students to enter the military linguist program. The test uses a made-up language to test the candidate's ability to understand language structure.

Interagency Language Roundtable (ILR) Scale: Used to measure proficiency levels on a scale from 0-5:

0 = No proficiency

0+ = Memorized proficiency

1 = Elementary proficiency

1+ = Elementary proficiency, plus

2 = Limited working elementary proficiency

2+ = Limited working elementary proficiency, plus

3 = General professional proficiency

3+ = General professional proficiency, plus

4 = Advanced professional general proficiency

4+ = Advanced professional general proficiency, plus

5 = Functionally native proficiency

Assumptions

The study used a quantitative research approach that centered on statistical analysis. The statistical analysis offers scientifically supported evidence in an attempt to either accept the null hypothesis or validate an alternative hypothesis. With statistical analysis, there is a level of objectivity over qualitative analysis. Participants are less likely to interject subjective data into a quantitative study. Additional assumptions connected with the quantitative research method include the concept that a sample population will be representative of the entire population. The assumption that special cause and common cause variation on the statistical data can be understood and, in the case of special cause variation, be assigned meaning by the researcher, is present in quantitative analysis. The measurement plan and methods also assume a level of validity and precision. Finally, in this quantitative research, the research design was developed to offer decision makers a definitive recommendation for which training method connected to the study should be prioritized for linguists in the population of the study.

A specific assumption for a study using a quantitative approach is that the method of data collection will prevent subjective data from the population from affecting the determination to accept or reject the hypothesis. An example of a subjective data point was the individual linguists' motivation to pass their annual DLPT. To mitigate this risk for the study, it is important to understand that those linguists who pass their annual DLPT are authorized to receive bonus pay. Under the current regulatory structure for the study's population, there was also a requirement for them to be separated from the army or reclassified into a new job after their second failure, however (U.S. Army, 2016). In a

study of a similar population of military linguists, Kurum (2011) found that there was no significant difference between income levels and motivation to learn a second language.

The test-taking ability of the individual linguists should be identified as a potential assumption. A large body of work has been published on the factors that surround an adult learner's ability to test. For the selected population of linguists, the assumption was made that they had the ability to successfully pass the DLPT. The entire population acquired L2 capability from the Defense Language Institute and Foreign Language Center, where students were exposed to study and testing strategies as part of the curriculum. Each student had to pass a final test successfully. Each semester, students were also exposed to several mock exams.

Scope and Delimitations

The specific focus of the research problem was the need to gain a better understanding of any statistical differences between two different training methods and sites for a population of professional army linguists. This specific purpose was chosen to help best utilize the funding associated with the program. One method of training incurs additional costs compared to training conducted at the home station. The population for this study consisted of professional army linguists with the ranks of Private First Class through Sergeant First Class. The population graduated from the Defense Language Institute Foreign Language Center, which is the DOD language school. The entire population also used the same testing procedures administered through an army testing center. This study involved looking at army linguists assigned to one major command

within the army. The inclusion of other linguists from this command was done to ensure that on-the-job language experience was similar across the population.

Limitations

The research plan called for the measure of secondary data gathered, which were based on annual DLPT results. The test may present an internal validity issue; as the researcher, I did not control test creation or administration. The internal validity issue connected with testing was mitigated through an understanding of the way that the test was created, validated, and administered.

The test was created and validated by the Defense Language Institute and Foreign Language Center, which is an accredited DOD school. The test is administered at multiple education centers located on each military installation. Most testing centers are operated through a contract or agreement with a local university. An example of this military–academic relationship is found at Ft. Bragg, NC were testing support and execution are provided by Campbell University at this location.

The only bias anticipated for this study was that of the individual linguist assuming that one training site was better than the other. This bias constituted one of the main reasons to use social representation theory as a framework for this study. The choice to use quantitative analysis on secondary data should mitigate this bias.

Significance of the Study

The original contribution to the field of L2 maintenance that this study may offer resides in the fact that it involved specific research directed toward adult professional linguists in the U.S. Army. Unlike their civilian counterparts, army professional linguists

must meet significant requirements beyond simply maintaining a level of language proficiency. The study hypothesis was that the different training sites provided different degrees of success. Again, this study was conducted to validate or disprove this hypothesis.

This research provides detailed quantitative data to army commanders and policy writers on whether outside contractors or in-house instruction is more effective for training for army professional linguists. This may allow for the best use of resources. This research also tested assumptions made in Total Army Language Program regulations and guidance. If it is disproven that there is a significant difference between outcomes for different types of training, then commanders may be able to better utilize the limited training time that soldiers with foreign language skills have.

The positive social change impact that this study may provide will be for those soldiers who make up the population of professional military linguists. In the selected major command, the linguists, junior in rank, range from the rank of Specialist (E-4) to Sergeant (E-5). In the army structure, those who are junior in rank have less say in how training occurs. The current structure is one where those making the training decisions are not professional linguists. A quantitative study on two differing methods of training for army professional linguists may provide for change within the army training community. This change, I hope, will better focus resources on methods of training that have a higher impact on army readiness. Readiness is the fundamental factor in the national security community's ability to meet public policy requirements to protect the nation.

Significance to Practice

This study may aid in the development and execution of language training policy in the U.S. Army. A potential contribution to the language and greater army training community may reside in a new or updated training assessment model, which may be used in the future for other training requirements within the same community. Again, this study has the potential to allow commanders to best utilize the limited resources provided to maintain the language capability of the U.S Army.

Significance to Social Change

Social change in the population of linguists studied may not be drastic. However, small changes may occur in the way in which training is selected. With these changes, the program will realize cost savings, which then can be used to improve the overall program. As an example, Kim and Cha (2017) detailed how experience abroad is a worthwhile investment for a L2 program. Beyond advances in the L2 training program, the understanding that may be gained from this study may provide commanders and managers with the ability to best use valuable training time.

An aspect of social change directly connected to the army's ability to maintain readiness rests in the concept of evidence-based policy making. Policies enacted related to the L2 program must not be an exception to this rule. Marchi, Lucertini, and Tsoukias (2016) explain that policy creation has become more complex and the method of policy making has shifted away from an opinion-based model. The waste of time and public resources should be considered as a social change issue. More organizations and departments are moving to efficiency models in the management of public organizations.

My aim in this research is to add to the body of work used in evidence-based policy making connected to the L2 program in the army.

Summary and Transition

This chapter provided the problem statement and purpose of the study. This section began with the framework for the study, providing the research question, hypothesis, and theoretical framework. Finally, this chapter addressed the significance, assumptions, and limitations of the study.

In Chapter 2, I attempt to provide a complete review of the literature associated with language training, testing, proficiency levels, and the abstract concept second language instruction in the army linguist community.

Chapter 2: Literature Review

The purpose of this quantitative study was to expand the understanding of the culture within the army professional linguist community. Understanding of the culture was developed by specifically researching the proficiency impact from multiple types of instruction used for L2 maintenance within the selected command under the U.S. Army. The training impact above is highlighted in Bauer et al.'s (2017) work that shows the misunderstanding that surround the most effective methods and time required for foreign language training. A core concept found in the work of Bauer et al. is that L2 acquisition is a complex science. Further, the ability to sustain L2 capability requires a dedicated design that consists of an understanding not only of the structure of a language, but also of the culture behind the language. Again, there is a problem with current U.S Army policy for training army linguists. Funding is insufficient for linguist training within the organization. Through this research, I hope to support the best utilization of financing in this underfunded program. Found in Army Regulation 11-6 for the Total Army Language Program (2013), the calculation of $7,500 is used as the cost calculation to fund each individual linguist. An estimate for this study was that the program was only funded at approximately half of the requested amount. Based, then, on funding, leadership is required to prioritize the individual linguists who receive training. Second, the method of training may have an impact on the efficiency of the overall program and the commander's military readiness. This research aimed to show through quantitative analysis which method of L2 maintenance training offers the most benefit.

The synopsis of the current literature centers on the elements of training and testing design. The relationship between the student and the instructor serves as the trend throughout the literature, with a focus on authentic language material used in many of the successful classrooms. Throughout the literature, the different models of assessment and testing have been studied in depth. Most published work in this area has involved attempts to show ways to increase proficiency. The gap in the current literature that this work was intended to fill centers on the comparison of methods of instruction.

This chapter is intended to aid in the understanding of the literature search strategy used during research and planning for the study. This information is followed by an overview of social representation theory, which was used as the theoretical framework for the study. Finally, this chapter includes a review of current literature connected to the variables for the study and the quantitative approach proposed as a model to examine the data. The chapter concludes with a thorough review of the formative work and the challenges faced by other researchers.

Literature Search Strategy

A study on how to best plan and execute language training in the army must take into consideration several factors. To meet the requirement to thoroughly research the data, a series of different search techniques were used to conduct the literature review. The databases used consisted of several resources located in the Walden University Library, as well as resources found at the Defense Language Institute, and through publications within the linguist community. Using the Military and Government Collection and expanding to the Political Science Complete collection aided in

developing a broad perspective through multiple training studies associated with the military. In using these databases for literary research, I was able to look beyond language training in the U.S. Army and review ways in which other professional military forces maintain L2 capability. The scope of the literature review encompassed peer-reviewed work published over the past 10 years, as well as a look back to 1961 to Mosconvici's original work on social representation theory. Using the following examples search terms provided for a more detailed review of the current literature in the fields of adult learning:

- *training perceptions and theoretical framework*
- *language and proficiency*
- *language training in the army*
- *language proficiency testing*

Theoretical Foundation

Social representation theory provided a framework that allowed for an understanding of how individuals and groups create meaning in concepts, ideas, and objects. Expanding beyond the language community to find a theoretical framework for this study makes it possible to seek broader application of the findings. Social representation theory was initially designed by Serge Mosconvici in 1961. Mosconvici was a Romanian who immigrated to France after his disillusionment with the Soviet takeover of the eastern European region, which provided him with the base ideas for his theory (Moloney & Walker, 2011). Mosconvici's general theory proposed that a subgroups transfer abstract ideas into assumed factual data (Hoijer, 2011).

Social representation theory naturally expanded into the study of training impacts resulting from a policy. Dulipovici and Vieru (2015) used both social representation and a socio-material practice perspective to offer a framework for a study on user perceptions of collaborative training models. These perceptions are a critical aspect of this study, in that there are user perceptions in the population of professional linguist soldiers that a certain model and venue for language maintenance training are preferred. Specifically, there is a perception that that Site 1 produces a higher quality of language training than Site 2. Therefore, social representation theory offered the best fit as the framework for the study.

The group dynamics and possibly the class type system that are created within the population of professional army linguists also led naturally to the selection of social representation theory for this study. Moloney and Walker (2011) provided an expansion on Mosconvici's theory of social representation, suggesting that subgroups within a larger group strive to reach a higher level of identification within the group. BenAlaya (2016) proposed the inclusion of the idea that behavioral rules created through processes within social representation theory may be caused by a group's place in the community. Through this study, I sought to investigate this very idea. Between the two training sites, there was a subjective view that one site provides better linguists. This perception has caused a subgroup of linguists to assign meaning to the belief that they are better than what they feel is a lower class of linguist.

Social representation theory served as an effective framework for a study by Karamanoli and Papachristopoulos (2007), who found that different categories of military

personnel groups had distinct attitudes toward seeking help to address psychological and medical issues. Dulipovici and Robey (2013) conducted a study using social representation theory that demonstrated how management groups at differing levels placed meaning on knowledge management systems. In a manner similar to the current study, these groups had similar vocabularies, understanding, and concepts but put different values on knowledge management systems.

All government endeavors must adhere to ethics; beyond the academic ethics requirement for this study, the linguist training program is funded by taxpayer funding, and the government must strive to use these dollars in the most efficient way. Markova (2013) connected back to Mosconvici's original theory to highlight the connection to ethics. Like Mosconvici, Markova contended that when one is observing intra- and inter-group dynamics, the human cognitive ability to think and express ideas is important. There is also an important ethical standard that must be followed by the instructors. In any formal training system, a teacher can teach to the test. Salloum (2016) researched the impact of social representation theory on teachers and found that like other groups, they formed opinions, collective behavior and beliefs, and that this construct impacted their teaching. All groups, including instructors, assign meaning to objects that they then also define ethically.

Public policy has a wide definition. However, national security and foreign affairs both are well-defined areas that have deep policy concerns that directly impact the nation or group being governed. Based on this idea, for this research, it was important to find a theoretical framework that connected to policy issues. Social representation theory has

been used in other scholarly work to study policy initiatives and their impact on groups within society. O'Dwyer, Lyons, and Cohrs, (2016) used social representation theory to study how groups within the country of Ireland formed views of foreign policy initiatives, finding that the theory helped to define the idea of identity and gave members of the national group a sense of who they were. Social representation theory was not the sole theory that could have been used for the study, however.

Other theories found in language studies could have contributed insights to the study. Political philosophy has had many forms over history. Some events and philosophers have had deep impacts on society, and others have aided in the clarification of these impacts and theories. Robin Lakoff is one of the individuals who has provided for the clarification of political change and theory. Lakoff (1975) provided a theory of sociolinguistics, which has an explicit assumption that people's use of language reveals their true beliefs. In her work, she described how sexism was observed in language and phrasing. This work connects to today's culture through the use of phrases in speech that show gender identification or language that alienates a group within society. This concept translates to L2 acquisition directly in that many foreign languages have gender grammar rules and structure that differ from one another. The beliefs of the student must be set aside, and a true cultural understanding of the L2 that the student is attempting to be acquired must be gained.

Oxford (2013) provided the theory of strategic self-regulation (S2R) in adult language learning. Oxford's S2R theory relies on self-motivation and a structure centered on the types of metaknowledge as part of metastrategies. Motivations play a role in all

army training paradigms. The S2R model could be used as a subframework in the study of the army's ability to maintain a second language capability. However, it is important to understand that the forms of instruction being compared in this study are structured and are directed rather than self-employed.

Oxford's (2013) the S2R model also has two basic assumptions. The S2R model assumes that everyone has the aptitude to learn an L2 through dedicated strategies and that these strategies can be taught (Oxford, 2013). The current army philosophy for identification for language training counters Oxford's first assumption that anyone can learn a second language. The army has developed a Defense Language Aptitude Battery (DLAB) test that is administered. A soldier must show an aptitude for language acquisition on the DLAB test before acceptance into formal language training. Studies have shown that graduation rates improve at the Defense Language Institute Foreign Language Center when DLAB waivers are eliminated.

Literature Review

Training Method

For the study, the independent variable was the method of L2 maintenance training. The population for the study received differing forms of training. The first type of training approach to training used temporary duty away from the subject's home station. The second consisted of a contract instructor-led classroom setting at the soldier's home station. Both approaches involved an attempt to use the target language throughout instruction. Mills (2011) reflected on this concept as early as 1950 in Great Britain, where he found that the method of instruction that worked best for L2 acquisition was use of

only the target language in the classroom. Stepanoviene (2016) provided a detailed study on how professional linguistics education may require a different approach to training. A key aspect of L2 education is the method of instruction. A review of the topic would be incomplete without a review of methods of teaching.

Beyond L2 acquisition training, there is a requirement to advance beyond one's initial L2 capability. Stepanoviene (2016) continued in his summary to show how although imitation or repetition of vocabulary may be suitable for gaining initial linguistic ability, this method provides minimal benefit in gaining the desired higher level of proficiency.

Morgan and Vandrick (2009) described the need to structure an L2 class to meet the desired level of proficiency by providing an example illustrating how the text selected for the course should match the desired level. With both training methods investigated for this study, the program of instruction incorporated learning beyond repetition and used material designed for the target level of proficiency.

Other aspects exist that are within and support the classroom that impact the ability to gain a high level of L2 proficiency. There is a large body of research on L2 acquisition focused on early education; however, many of the key variables cross over to a study on adult learners. Kim and Cha's (2017) study showed that traveling abroad appears to impact students' ability to improve language proficiency. Another aspect of L2 education, beyond the method of instruction, consists of the level awareness in the linguist practitioners of their thought process and how they best learn. Henter (2012) conducted a study using metacognition for an individual adult learner as a framework.

Even though this study was done by a student at the Romania Air Force Academy, the findings offer valid insight about second language acquisition for my study. Henter found that even metacognition activities outside the classroom increased proficiency level by 9%. Some self-study must be built into army linguists' language training plan, serving as a building block toward formal training. Goodridge (2017) articulated that students who approached L2 acquisition for pragmatic purposes may show a higher level of motivation. Self-study was a fundamental building block for the target population. The proposed study did not take into consideration the self-study habits of the individual linguist.

To expand understating of the language classroom, other aspects of the training model should be explored. Castrillon (2017) summarized that between the student and teacher, in the L2 acquisition classroom, there must be a structure built that produces material that is slightly above the student's level to improve proficiency. Holman (2013) provided a model for curriculum alignment specific to language instruction within the DOD. Holman found that inclusion of formative assessments that directly connected to objectives provided a more productive model than an earlier attempted summative model. Song (2014) demonstrated the existence of three forms of interaction in an L2 classroom: learner-to-learner interaction, learner-to-instructor interaction, and learner-to-content interaction. All of these forms of interaction occurred at both training sites in the current study.

To continue the discussion on the relationship between the instructor and the student and the effect this relationship may affect proficiency, Tyler (2012) found that the

trust level between a student and a teacher had a weak correlation to the student's ability to communicate during instruction. Zhou and Ma (2014) articulated that L2 instructors should closely monitor students' delivery of language, which is the physical manifestation of their learning, to develop strategies to improve students' speaking skills. All of these factors build toward a higher level of language capability.

In relation to the communication skills of linguists, Shih (2013) found that a level of immersion training improved both oral and written communication, while also expanding on the idea that to be able to speak as a native, one must think like a native as well. Wu (2013) further advanced this idea by offering an example of an approach to oral language instruction that includes elicited imitation where the student repeats short sentences as precisely as possible.

Finally, a frequent topic found in discussions of L2 training is the use of computer-assisted language training. Burston and Arispe (2018) performed a study to specifically validate the claims made by the computer-assisted language learning (CALL) community using CALL studies from 1983 to 2015, finding that, in fact, CALL has a negligible impact on L2 proficiency. Hur (2012) studied the effect of using a technology-integrated classroom, specifically investigating the impact of this approach on different groups based on technology experience; this again ties back to social representation theory, in that different groups assign meaning to objects and ideas. Elements of CALL were incorporated into both training models for the study.

Testing

The dependent variable for the study are the results on the DLPT. In both the testing and training realms of L2 acquisition and sustainment anxiety plays a key role. Gerencheal (2016) found using the foreign language classroom anxiety model (FLCAS), with third-year English students at Miuzn-Tepi University, a higher level of anxiety was found in female students over males.

Horwitz (2016) expanded upon the idea that anxiety plays a large role in the language education field finding that both nationality and target language are factors in the FLCAS measurement of anxiety in the language classroom. In the study there was no control for test anxiety in the study would be difficult. As a mitigation strategy, an understanding of how army linguists prepare for testing as part of their training was an important aspect to understand.

As part of the curriculum at the Defense Language Institute and Foreign Language Center, students are taught several methods of test taking. They are given several mock exams that simulate final testing. Won and MacDonald (2014) offered that language instructors at the Defense Language Institute should use diagnostic assessments that consider not only the listening, reading, and speaking skills of the student but also differentiate instructional approaches. Finally, the DLPT test results are reported using the ILR scale. Cox and Clifford (2014) argued that using the ILR scale to test the listening modality violates the academic assessment practices in the field.

Understanding how the results of the DLPT are reported is another important aspect of the study. Cox and Clifford (2014) described that the ILR scale as designed to

measure ability levels for linguists at varying levels of difficulty. There are alternatives to the ILR Scale, Nolan (2014) discussed the STANAG 6001 level descriptions, as used by NATO, as an alternative to the ILR scale.

Test design can impact the understanding of an organization's actual capability. Nolan (2014) completed a detailed study on how Sweden assesses their military professional linguist arguing that testing should be predictive and standardized. Nolan made a key point in his study of Sweden professional military linguist in that assessments should be based on real-life professional language use. Cox and Clifford (2014) provided a detailed description on how language assessments should be structured in a short verses long ladder design, which is similar to what exists for professional army linguists today there are lower level DLPT and higher level DLPTs. These methods may provide a better gauge on determining proficiency levels. The DLPT is designed in such a way that the questions become more technical and the concepts behind the development of the questions become more complex the higher level of professional.

Graduation criteria from the Defense Language Institute and Foreign Language Center requires the student to obtain a 2/2 on their end of course DLPT, which can serve as a benchmark for later training needs for the student. Ellington, Blume, Surface, and Wilson (2015) showed that the time from graduation and end of course score on the DLPT had an impact on L2 maintenance over time. Time and experience as a linguist may help you to improve overtime, however. The linguist most actively engage in the use of a L2 in an effort sustain and improve on the level of proficiency.

The discussion on how to train and evaluate must include a strategy for assessment. Ainsworth and Viegut (2006) discussed the creation of formative assessments in support of classroom instruction. Even though Ainsworth and Viegut's analysis centers on primary education in a K-12 setting the conclusions and framework have connection to all L2 learning. The criticisms of the formative assessment model are that teacher may be prone to teach to the test (Ainsworth & Viegut, 2006). For teachers to teach to the test, on the DLPT, would be extremely complicated. The formative assessment model is valid for secondary research for the professional linguist in the army.

In developing a strategy to study second language maintenance, the type of acquisition strategy must be considered. Ellington, Surface, Blume, and Wilson (2015) used a framework of training transfer to study how time and graduation testing levels affect soldier's ability to maintain a certain level of proficiency. Training transfer as defined in the Ellington et al. as expanding the idea that instead of using postformal education testing to evaluate proficiency levels, evaluators should use job performance to determine a soldier's level of proficiency. Training transfer is an important part of the target population of this study but does not address the formal classroom evaluation.

Proficiency Levels

The ultimate goal of any L2 acquisition or maintenance training is to achieve a higher level of proficiency. Again, proficiency levels serve as the dependent variable in the study. This process starts by recruiting individuals who have an aptitude for acquiring a second language. Santizo (2017) provided quantitative research that showed comparing initial aptitude battery test and the results of the oral proficiency test at the end of

instruction had a moderate correlation in certain language categories. Elfiky (2017) studied military linguist students and their ability to accurately self-predict their level of oral proficiency. Kim and Cha (2017) also found that experience abroad was a predictor for comprehension of self-efficacy in the targeted L2. Further, expanding on predictors that may connect to proficiency levels Sparks et al. (2006) research showed that linguist students who had a degree of native written language proficiency serves as a good predictor of L2 proficiency. Combined, these findings show a correlation between the self-predicted proficiency and the results from the formal proficiency testing. Interestingly, the army linguist was able to predict their ability.

Again, the army standard for determining proficiency levels is based on the 1-5 ILR scale. Higher level of proficiency on the ILR scale at a 4 and 5, with level 4 approaching a professional level, was achieved by portion of the population. Army linguists who desires to reach this higher level must take a second DLPT, which is administered with higher level questions.

Language in the Abstract

A professional military's ability to understand the adversary, has been present since the beginning of military conflict. There are other aspects to the art of mastering a foreign language beyond the classroom and testing. Mills (2011) provides a historical example on research in which the Government Communications Headquarters in the 1950s, tasked with listening to Russian intercepts, found that women made better linguist than men. The head of the Commission and Warrant Branch concluded that not only did women have better acquisition skills but also required less inducement to maintain their

proficiency (Mills, 2011). Miller (2016) in a study of students at the U.S. Service Academy, found that civilian instructors are more knowledgeable and military instructors are more relatable. Miller also found that as the L2 learner gained a higher level of proficiency, based off on the increased perception that civilian instructors were more knowledgeable.

The ability to communicate is an advantage in military conflict and when overlaid with national security policy the, art of language, becomes a matter of life and death. Morgan and Vandrick (2009) expanded their pursuit of language studies to detail how language is also a medium. Similar to any other medium and in most cases more so, language transforms into a form of expression. Johnson and Berrett (2011) warn against the practice, for professional army linguists, to generalize a culture. Johnson and Berrett's research support the argument that for professional linguist attempting to determine the meaning or intent behind an action, requires understanding the culture not just the capability to transliterate text.

The concept of learning a L2 stems, for humankind, from our first native language, this knowledge builds a framework for later learning in life. Miller (2017) found that with native Spanish speakers, who spoke Spanish as a primary language in the home, were affected by the parent's attitude toward speaking their native language versus English. Slabakova (2013) summarized that within a comparison between a second and first language acquisition there is little difference, while further stating that while young there is a critical period for first language acquisition.

Human motivation again serves as another abstract concept that may affect L2 acquisition and improvement. Kurum (2011) provided a unique study of Turkish military linguists in their seven years of dedicated study of English as a second language, showing that motivation level had a significant impacts on proficiency levels. Ferrer (2014) provided research on how to structure a day in a language classroom. Ferrer expressly found that by delivering tailored homework for their students that perceptions of increased motivation grew in the classroom. All of the previous factors in L2 instruction lend themselves to build on operational readiness.

Operational Requirement

A language capability within the force is a crucial combat multiplier. The current term for defining the environment in which the army will fight is called the Operational Environment. Knight (2012) discussed how a vital part of planning for the Operational Environment is the ability to communicate. Knight provided a framework on how the linguist capability can be used to perform tasks such as medical support and logistic coordination within a host country. Ellington's et al. (2015) studied a population tasked to perform these missions, which consisted of special operations soldiers who formed in a cohesive team that all had the same L2. In this environment, a leader has a better ability to evaluate soldiers on L2 skill levels. The target population for this study was not organized in the same manner, with leaders that may have no language skill or not have the same language as the soldier being evaluated. Therefore, a more strategic approach has to be considered.

Again, cultural understanding as a variable is essential in the discussion of the operational environment. Waldvogel, Youtz, and Laser (2013) highlighted that the requirement for soldiers deployed to foreign nations must have a cultural understanding and the ability to communicate for all current and future military operations. Kim (2013) provided a successful example on how to meet this cultural and language requirement, describing his participation in the Foreign Area Officer program and explicitly serving as a military planner in the Pacific theater of operations. Goodridge (2017) expands this theory by summarizing that a second language capability informs policy to aid in a strategy to overcome communication barriers within the international community. To connect this understanding back to the linguist classroom, Zhang and Zhou (2014) found, through observation of classes at the Defense Language Institute, that students that lack cultural knowledge, directly affected linguistic potential.

The ability to communicate is not only a requirement in today's military operations but will continue to be a requirement in the future. Dilianian and Akiwowo (2016) relayed the idea, which is supported by senior leaders in the U.S. Army, that future operations will take place over a vast geographical area and require organizational and regional culture understanding to be able to operate fully. Carver (2015) further articulated that army units must work to shed old paradigms and designs. This force of the future must be focused on readiness as a holistic force.

Summary and Conclusions

The dedicated study on L2 acquisition and maintenance serves little purpose if not incorporated into the more extensive body knowledge. Holman (2017) captured several

key aspects of L2 acquisition and maintenance that are reflective of the entire body of knowledge. Grounded in a review of the history of how language proficiency was measured in the DOD. Also according to the Holman, dissimilar to academia, which is focused on understanding literature in a target language, government language schools are concentrated on proficiency in the reading and listening domains. In addition, Holman found that that the government language education system, compared to that of academia, failed to publish the research and findings on success and failures. This study was designed to help fill this gap on aspects of L2 maintenance that translate to the differing methods of training.

Chapter 3: Research Method

The purpose of this quantitative study was to expand the understanding of the culture within the army professional linguist community. Specifically, I conducted research on the multiple types of instruction used for L2 maintenance within the selected major command under the U.S Army. This chapter begins with a detailed discussion on the research design, methodology, and population for the study. This chapter then presents a review of data collection and analysis procedures, as well as any validity threats to the study.

Research Design and Rationale

The nonexperimental design was used for the study, which involved two training locations that used different methods of training, with the research structured to determine whether there was a significant difference between the two sites. There was no control group for this study, and no variables were manipulated as part of the research. The data used to study the above variables were secondary data. Throughout the body of work presented in the previous chapter, there were comparison studies that reviewed individual aspects of the L2 classroom and their impact on proficiency levels. This study, using the nonexperimental research design, was conducted to add to the body of knowledge on training assessment and the ability to maintain L2 capability within an organization.

In that I used secondary data, the time constraints for this study resided in the Freedom of Information Act (FOIA) process. An assumption made for this study was that the FOIA process could take several months to complete before I would receive data back

from a government organization. To mitigate this constraint, I discussed with the organization its requirements to fulfill the FOIA request and, with consultation from the Institutional Review Board, developed a plan for how to handle the data once received. The FOIA request was sent to the identified points of contact (see Appendix A). Again, the variables for the study were the following:

Dependent variable: Proficiency level for calendar years 2014-2018, which was an ordinal level of measurement from 0-5, with the + measurement on the ILR Scale denoted by .6.

Independent variable: Location of training; ordinal level assigned as 1 or 2.

- Site 1: Professional education from a DOD higher education formal school
- Site 2: Contract instruction, full-time, with precise schedule and level

Methodology

Population

The target population for this study was army professional linguists in the ranks of Specialist to Sergeant First Class. The total population of professional military linguists in the selected major command numbered approximately 460 in 2016 and 2017. This population further consisted of active duty soldiers who held the military occupational specialty of Cryptologic linguist (35P). Data were gathered for only those soldiers who had graduated from the Defense Language Institute and Foreign Language Center; the data population was further reduced by including only those linguists from each site who had completed posttraining testing within the same year as the training.

Due to such a low population size, there was no sampling procedure, and the study used already collected secondary data from the entire population.

Archival Data

The secondary data for this study consisted of data on the testing site, proficiency test levels, primary control language, and dates. Each data record was given a unique identification number prior to release of the data to ensure that the linguist remained anonymous. To obtain these data records, a FOIA request was submitted to the major command that this study aimed to analyze. These data were archived by the organization, and the request was for data from 2014 to present day to ensure that enough data were collected to show any level of significance between the two sites. The FOIA request sent to the major command is located in Appendix A. The secondary data were collected as part of official military training records. Pay actions are generated from these data. Therefore, it is in the best interest of the army to ensure that the data has a high level of validity.

Instrumentation and Operationalization of Constructs

The location of training served as the independent variable for the study. There were two primary types of locations where the population received training. Site 1 consisted of a professional education center that focused solely on foreign language instruction, which was serviced by professional educators. Site 2 consisted of the linguist's home station, with training provided by contracted civilian language instructors in a semistructured setting. Proficiency levels served as the dependent variable, which was measured as an ordinal level of measurement from 0-5, with the + measurement on

the ILR Scale denoted by .6. The DLPT was the instrument used to provide the above proficiency levels in the listening and reading modalities. The DLAB test is used to provide a level of aptitude for prospective students to enter the military linguist program. The DLAB test uses a made-up language to test candidates' ability to understand language structure. The ILR Scale is used to measure proficiency levels on a scale from 0-5. The ILR scale categories are as follows:

- 0 = No proficiency
- 0+ = Memorized proficiency
- 1 = Elementary proficiency
- 1+ = Elementary proficiency, plus
- 2 = Limited working elementary proficiency
- 2+ = Limited working elementary proficiency, plus
- 3 = General professional proficiency
- 3+ = General professional proficiency, plus
- 4 = Advanced professional general proficiency
- 4+ = Advanced professional general proficiency, plus
- 5 = Functionally native proficiency

No collection instruments were used for this study. However, Minitab software was used to perform the statistical analysis of the secondary data collected. Minitab is one of several statistical software packages that allow a researcher to perform statistical analysis. Shirey, Sullivan, Lines, and Smithwick (2017) offered a description of how Minitab was used as a successful application in the medical and Lean Six Sigma process

improvement fields in performing detailed statistical analysis. To show validity, the final study presents the raw output charts with a comprehensive analysis of the statistical metrics created. Since the reported data was nonnormal the Mann-Whitney and chi square test for association was selected from Figure 1 to answer the research question.

Data Analysis Plan

Similar secondary data sets for the dependent variable have been reported in two different ways. The most common is the ordinal 1 to 5 scale with a plus symbol indicating a half level in between the levels. Other examples of similar data are represented in whole numbers, with an example of a score of 26, constituting the 2+ category. Much of the data analysis depended on which form these data were reported. First, purposeful verification to determine if the data was continuous or ordinal took place. Then, the Minitab version 17 software package was used to perform the following analysis.

The data was analyzed to determine the baseline measurements consisting of the mean and standard deviation for the subsets of data composed of proficiency levels from the multiple years and by training site. A graphical summary of Minitab was used to determine if the data had a normal distribution.

If the data were reported as continuous data, a two-sample *t* test would have been used to determine if there is a statistical difference between the two sites of training. Xu, Fralick, Zheng, Tu, and Feng (2017) stressed that for a two-sample *t* test, the two samples must be independent. Depending on the variance of the data, a *Z*-test may also be used. Further, Figure 1 shows that the data must have a normal distribution to use these

statistical tests. Since the secondary data showing proficiency levels were measured on a 5-point scale and these records were reported as attribute data, the following test were used. Multiple Mann-Whitney and chi-square tests for association were used as statistical test for the study. The data also required the use of a Pearson correlation coefficient to determine if there is linear correlation between the two variables. Further analysis was used to determine if there was variation and to assign cause to this variation. The below procedure, which is based on the Lean Six Sigma (LSS) training hypothesis tree selection, was used to account for multiple statistical tests (LSS, 2015).

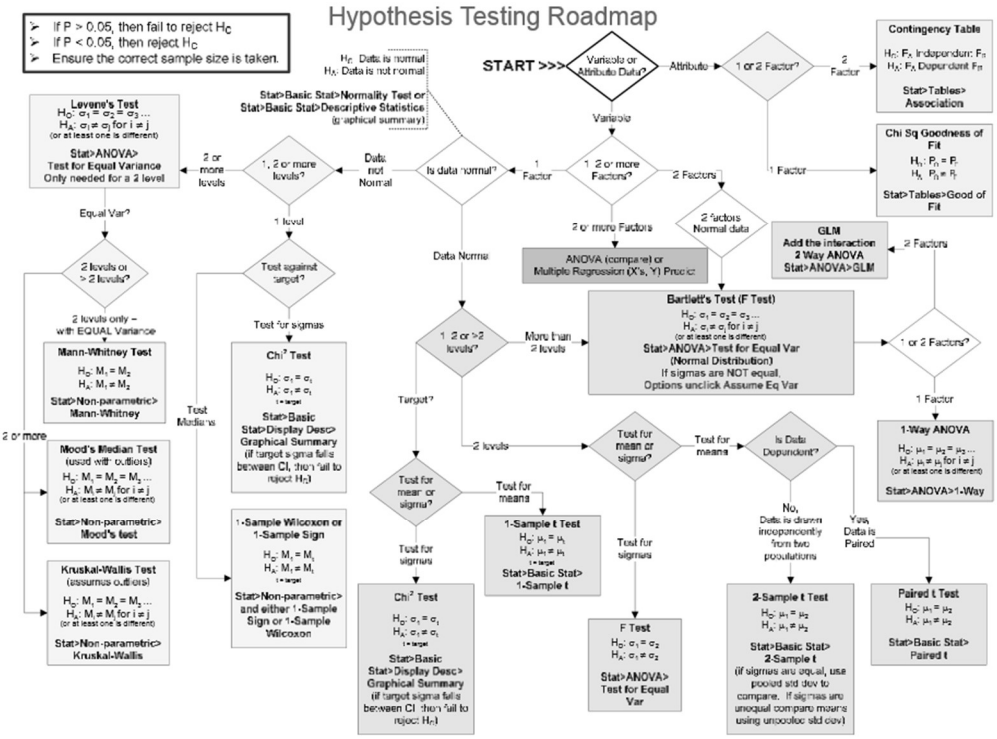

Figure 1. Hypothesis testing roadmap.

In review, below are the research question and the hypothesis that I sought to answer through the study:

RQ: Is there a significant difference between the types of training events measured by the posttraining proficiency levels?

Hypothesis: Trainees at Site 1 achieve significantly higher proficiency levels than trainees at Site 2.

Null hypothesis: There is no significant difference between Site 1 and Site 2 posttraining proficiency levels.

There was a level of data cleaning required once the data were received through the FOIA request. Authentication of the data required a comparison of language taught to the control language for the linguist. Additionally, dates of training and testing dates was needed to be compared and correlated. Anticipating the need to align and scrub the data, preparation and time was spent on the following actions:

- Data were correlated by training site locations with the test results.
- Data were then correlated by the year of training.
- Data fields needed to be converted into a standard format.
- The data were sorted by fiscal years and by quarter.
- Removal of any duplicate records was conducted.
- The naming of columns was checked.

Threats to Validity

External Validity

The population of army linguists may have different control languages; this threat to validity is mitigated by the same model of acquisition training, which was standardized by the Defense Language Institute and Foreign Language Center. There were no sampling biases because the entire population was used for this study. To limit the generalization of the findings an effort was made to determine if there are any mitigating factors that lead to any post DLPT effect on the independent variable beyond the limits of the study.

Internal Validity

For internal validity, the study aimed to determine which variables indeed had a cause-and-effect relationship. Because this study aimed to determine the significance between the training sites and testing outcome, the study authentically had an element of cause, the training came before the effect, which was the test. There were no indications that maturation was present; the study will not cause a physical or physiological change in the participants. Again, as found in the literature review, in any training scenario, the threat that teachers and students will become test aware is possible. This internal validity threat can be addressed through awareness.

Construct Validity

As discussed, the primary construct validity issue for the study was the test results. An assumption made for the research was that the DLPT does, in fact, measure the listening and reading ability of the test taker. The DLPT has been used to validate the

language capability of individuals for some time. The test is used to determine proficiency pay rates by the army, which offsets this concern. There were no confounding variables present in the study and no statistical conclusion validity issue found in the study.

Ethical Procedures

Because access to the secondary data used for the study was gained through a FOIA request, there were no ethical concerns connected with the possibility of personal identification data being presented in the study. The data were scrubbed for any information that might identify a person. Further coding was used to ensure that specific sites of training are not recognized by actual name, but by type of instruction. There were no institutional permissions or ethical concerns related to recruitment.

The data will be stored per Walden University Institutional Review Board requirements for 5 years on a separate external hard drive and the hard drive of my primary computer. These data will include any Excel files and the Minitab output file. The data for the study will be destroyed at the end of the 5-year time frame as dictated by the university.

Summary

In summary, this chapter provided the methodology designed for the study. The study used a nonexperimental design to determine if there are any significant differences between the sites of training and their resulting proficiency levels. Using the Minitab software package, the data analysis will consisted of a series of statistical tests to

determine if there is a difference or correlation between factors. In the following chapter, a complete review of the findings is presented.

Chapter 4: Results

The study proposed to quantify the results of two different methods of L2 training for army linguists through their proficiency scores measured by the DLPT, which was used to determine if a significant difference between methods of training was present. Specifically, the design of the study attempted to establish if there was a significant statistical difference between the two sites of training, which might have practical impacts on how the limited funding for this program is managed. The research question that shaped this study was the following: What is the level of significant difference between the types of training events measured by the posttraining proficiency levels? The selected hypothesis for the study was that a significant difference exists between Site 1 and Site 2 in posttraining proficiency levels, with Site 1 producing higher DLPT results than Site 2.

The following chapter details the data collection process, including the method used to gain access to the secondary data and the methods of analysis. Further, this section expands on the results from the study, validating the study's finding that there was no significant difference in either the listening or reading modality on the posttraining DLPT between the two training sites. The chapter concludes with a complete review of the study's results.

Data Collection

The secondary data used for the study were obtained through a FOIA request for training and testing data for a select major command in the U.S. Army. The timeframe for the data collected ranged from March 2014 to February 2018. The data consisted of 3,044

individual testing opportunities over this period, which included results from both the reading and listening portions of the DLPT. The DLPT is offered at the end of the training session and is a yearly requirement for the population of the study. This study was designed to compare these posttraining test scores between the two sites of training. The data records were identified by unique identification numbers to ensure the anonymity of the individual test takers.

The DLPT data were reported as ordinal and not as continuous data, which eliminated the need for *T* test, Z test, or a covariate called for in the data collection plan presented in Chapter 3. In place of the *T* or *Z* test, multiple Mann-Whitney test and chi-square test for association were used. The data provided for Site 1 did not include previous-year testing scores. Therefore, only final DLPT scores were used to understand if there was a significant difference between the two training sites. The Mann-Whitney test was used to compare the ordinal-level DLPT results, which were based on the ILR scale, to understand if Site 1 produced greater posttraining proficiency levels than Site 2. The chi-square test for association was used to determine whether the categorical factors of pass or fail between the two sites had a significant difference. Both tests treated the listening and reading modalities separately.

The setup for the Mann-Whitney test consisted of designing four columns: Reading Site 1, Reading Site 2, Listening Site 1, and Listening Site 2. The test then compared the probability between the two training sites to determine, in this case, if Site 1 produced higher DLPT scores than Site 2. The Mann-Whitney was conducted twice: once for the listening modality and then for the reading modality.

To properly set up the multiple chi-square tests for association, the data had to be converted into a categorical classification, which resulted in the coding the data as pass or fail. The failing category consisted of those scoring 0, 6, 10, or 16 on the posttraining DLPT. The passing category consisted of those with DLPT scores of 20, 26, 30, 36, and 40 on the posttraining DLPT.

The population included soldiers holding the military occupational job title of cryptologic linguist who ranged from the rank of Specialist to that of Sergeant First Class. All test scores for both training sites, again consisting of 3,004 individual testing events, were used for this study. Tables 1 and 2 provide distribution data for the target population.

Table 1
Distribution DLPT Scores for Reading at Site 1

DLPT score	6	10	16	20	26	30	36	40	All
Count	0	0	0	25	36	25	1	0	87
Distribution %	0	0	0	28.7	41.4	28.7	1.2	0	100

Note. $N = 87$. DLPT scores are shown as whole numbers that correlate to the ILR scale.

Table 2
Distribution DLPT Scores for Reading at Site 2

DLPT score	6	10	16	20	26	30	36	40	All
Count	8	5	87	733	1160	906	22	36	2,957
Distribution %	0.3	0.2	3.0	24.8	39.2	30.6	0.7	1.2	100

Note. $N = 2{,}957$. DLPT scores are shown as whole numbers that correlate to the ILR scale.

Tables 1 and 2 provide the general distribution data for the reading modality test results scored at the end of training. Site 1 had $N = 87$ total testing opportunities from

March 2014 to February 2018, compared to N = 2957 for Site 2. Again, the data were reported on an ordinal scale that included 0, 6, 10, and 16 representing failing scores and 20, 26, 30, 36, and 40 representing passing scores. The tables show the ordinal DLPT scores across the top row. The tables also include, in descending order, count data for individual scores on the DLPT for the two sites in the columns and the associated distribution, shown as a percentage. Site 1 had no failures for the reading modality, compared to Site 2 having a 3.111% failure rate. No statistical significance can be inferred from this finding.

Table 3
Distribution DLPT Scores for Listening at Site 1

DLPT score	0	6	10	16	20	26	30	36	40	All
Count	0	0	1	4	34	26	22	0	0	87
Distribution %	0	0	1.2	4.6	39	30	25.3	0	0	100

Note. N = 87. DLPT scores are shown as whole numbers that correlate to the ILR scale.

Table 4
Distribution DLPT Scores for Listening at Site 2

DLPT score	0	6	10	16	20	26	30	36	40	All
Count	2	3	39	140	977	805	934	35	22	2,957
Distribution %	0.1	0.1	1.3	4.7	33	27.2	31.6	1.2	0.7	100

Note. *N* = 2,957. DLPT scores are shown as whole numbers that correlate to the ILR scale.

Tables 3 and 4 provide general distribution data for the listening modality test results scored at the end of training. Site 1 had N = 87 compared to Site 2 having N = 2957 total testing opportunities from March 2014 to February 2018. The tables also include, in descending order, count data of individual scores on the DLPT for the two

sites in the columns and the associated distribution shown as a percentage. Site 1 had a 5.747% failure rate in the listening modality, compared to Site 2 having a 6.223% failure rate. Again, no statistical significance can be inferred from this finding.

Because I used the total population rather than applying a sampling strategy, the study showed a direct representation of the target population. The listening and reading scores were separated in the final study. This allowed the findings to even further target the best use of resources. Other descriptive data beyond the soldiers' job classification and rank were not provided. The rank data provided had a direct correlation to how many years of service the individual linguists had in the army. Because the length of service data were not provided, this variable was not considered for this study.

Study Results

Descriptive Statistics

The descriptive statistics for the study must begin with an understanding of how the DLPT scores were reported. The DLPT data were reported on an ordinal scale based on the ILR scale, a 0-5 scale with a + symbol indicating a half score between ILR levels, with 11 levels in total. These ordinal data can then be used to show the categorical comparison of pass or fail on the posttraining DLPT. The data were reported using whole numbers: 0, 6, 10, 16, 20, 26, 30, 36, and 40. None of the test results showed any linguist obtaining a level 50, which is functionally native proficiency. The pass-fail data were categorized by using 16 and below to show failure and 20 and above to show passing. For the DLPT and the target audience, a score of 2 in both modalities is required to pass.

Using a pass-or-fail categorical scale allowed for the design of a 2 by 2 matrix that showed training site compared to the pass or fail rate by modality.

To aid in showing basic descriptive statistics for the study variables, Figures 2 through 5 show a graphical summary between each site and dependent variable shown both by listening and reading. This shows that the data do not have a normal distribution and the data are indeed ordinal and not continuous data. The spread of data the population reaches the passing level on the DLPT and then the distribution of those soldiers who scored higher was expected, with the highest concentration of students at the two level on the ILR scale in both the listening and reading modalities. Site 1 reported no failures for the reading modality and only five failures in the listening modality. Conversely, Site 2 reported more linguists reaching the higher scores in both modalities.

Figure 2 represents a graphical summary for the listening modality at Site 1. The *p*-value of less than .0005 indicates that this subsection of data does not have a normal distribution and, when compared to the binning of data points in the histogram, confirms that the data were categorical and not continuous data. The histogram does show and the data confirm that the mean score was at the 2+ level.

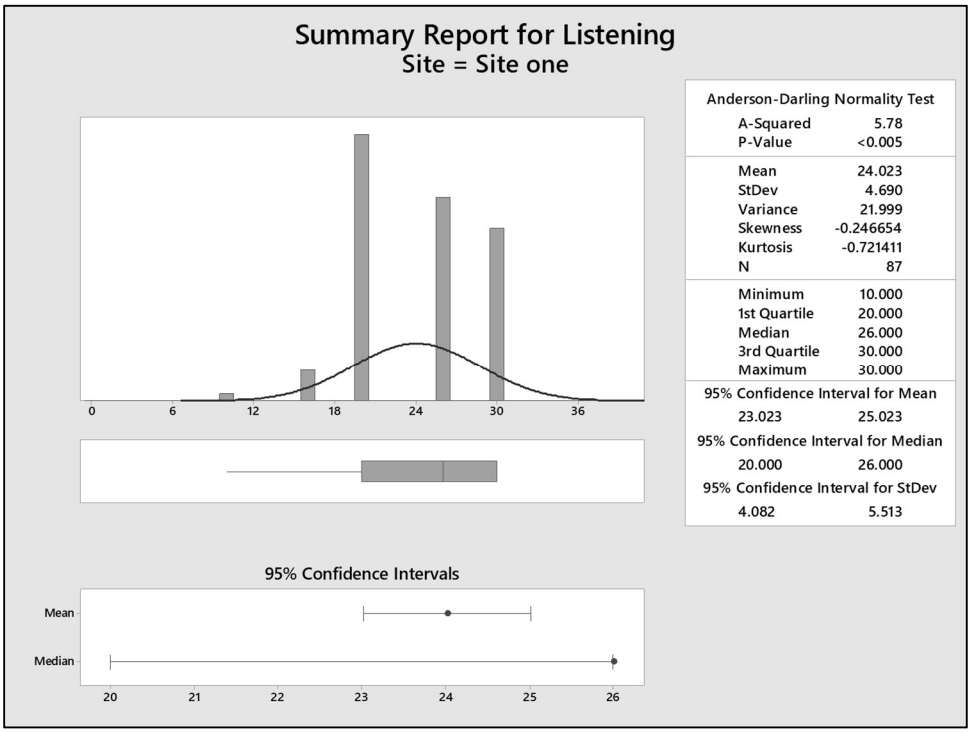

Figure 2. Summary report for listening, Site 1.

Figure 3 represents the graphical summary for the listening modality at Site 2. The *p*-value of less than .0005 indicates that this subsection of data does not have a normal distribution and, when compared to the binning of data points in the histogram, confirms that the data were categorical and not continuous data. In the listening data at Site 2, there is a slightly higher mean than at Site 1.

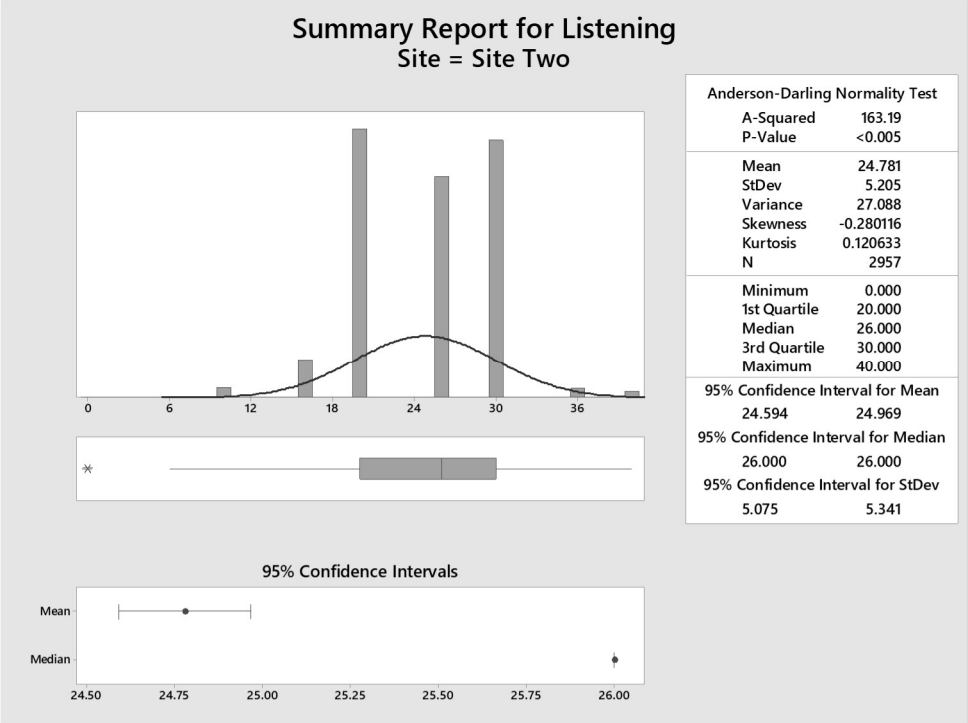

Figure 3. Summary report for listening, Site 2.

Figure 4 represents the graphical summary for the reading modality at Site 1. The *p*-value of less than .0005 indicates that this subsection of data does not have a normal distribution and, when compared to the binning of data points in the histogram, confirms that the data were categorical and not continuous data. The overall data indicated a higher mean for the reading scores than for the listening scores.

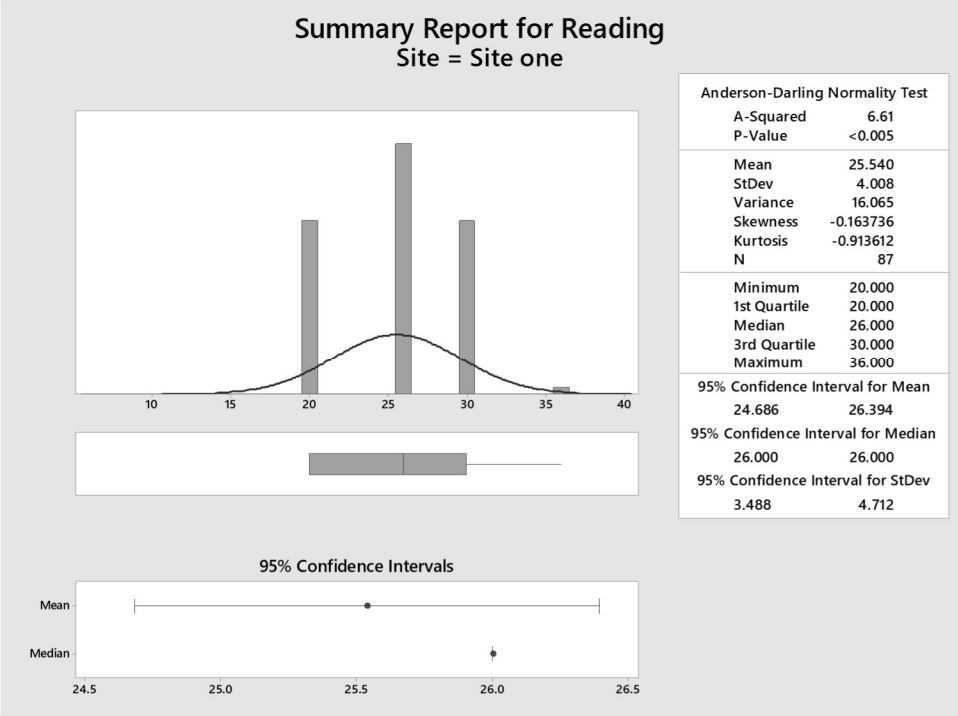

Figure 4. Summary report for reading, Site 1.

Figure 5 represents a graphical summary for the reading modality at Site 2. The *p*-value of less than .0005 indicates that this subsection of data does not have a normal distribution and, when compared the binning of data points in the histogram, confirms that the data were categorical and not continuous data.

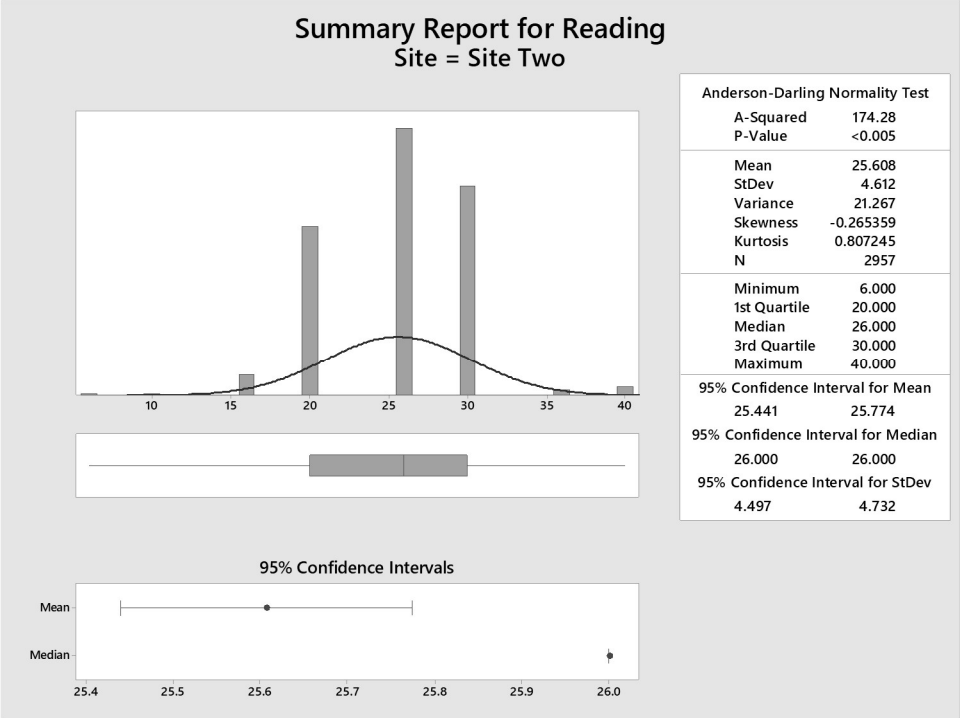

Figure 5. Summary report for reading, Site 2.

Statistical Assumptions

The assumptions for the Mann-Whitney test were as follows: All of the observations from both groups are independent of each other; all responses are ordinal; the null hypothesis is that populations are less than, greater than, or equal; and the alternative hypothesis is that the populations are not equal. The data as originally reported met these assumptions, with each population training at different locations and testing being conducted on an individual basis.

The two primary statistical assumptions for the chi-square test for association were the following: Observations are independent of each other, and there are at least five observations in any one cell. To address the first assumption for the test, based on the test administration rules, individual linguists taking the test had no impact on each other. In

the final 2-by-2 model created for this study with pass-fail as the dependent variable and site of training as the independent variable in the reading modality, the second assumption was not met with less than five failing observations from Site 1.

Findings

In an attempt to quantitatively answer the primary research question, which was to determine if there was a significant difference between the two training sites measured by the posttraining proficiency levels, the Mann-Whitney and chi-square tests for association were used. The hypothesis was that trainees at Site 1 achieve significantly higher proficiency levels than trainees at Site 2.

For both the chi-square test for association and the Mann-Whitney test, the null hypothesis was that there was no significant difference between Site 1 and Site 2 posttraining proficiency levels. Based on a 95% confidence interval the significant *p*-value is calculated at .05 for this test. Based on this calculation the rule used for this test was that if the reported *p*-value is less than .05, the null hypothesis is rejected, therefore accepting the alternative hypothesis.

The two Mann-Whitney test resulted in a reported *p*-values in the listening modality of .915 and .6 in the reading modality. Both of these test result in the finding that we fail to reject the null hypothesis, which indicates that Site 1 is not greater than Site 2 in either modality.

Based on the two tested areas on the DLPT, listening and reading, which again serve as the dependent variable, two chi-square test for associations were conducted. The

first tested between the two independent variables or testing sites and the listening modality tested on the DLPT.

Table 5 showed the results from the Mann-Whitney test that compared the listening modality between the sites of training. The purpose for this test was determine if Site 1 listening testing results are greater than Site 2. The null hypothesis for this test is that both sites are equal. With a reported *p*-value of .915, indicates that we fail to reject the null hypothesis. This supports the conclusion that Site 1 does not produce a higher proficiency rate as measured by the posttraining DLPT.

Table 5
Mann-Whitney: Site 1 Listening (N = 87), Site 2 Listening (N = 2,957)

Sample	*N*	Median
Site 2 listening	2,957	26
Site 1 listening	87	26
Method	W-value	*p*-value
Not adjusted for ties	121350.50	0.915

Note. Ho: n1 - n2 = 0. Ha: n1 - n2 > 0. A 95% confidence level was used for this test.

Table 6 showed the results from the Mann-Whitney test that compared the reading modality between the training sites. The purpose for this test was to determine if Site 1 reading DLPT testing results are greater than Site 2. The null hypothesis for this test was that both sites were equal. With a reported *p*-value of .6 indicates that we fail to reject the null hypothesis. This supports the conclusion that Site 1 does not produce a higher proficiency rate as measured by the posttraining DLPT.

Table 6
Mann-Whitney: Site 1 Reading (N = 87), Site 2 Reading (N = 2,957)

Sample	*N*	Median
Site 2 reading	2,957	26
Site 1 reading	87	26
Method	W-value	*p*-value
Not adjusted for ties	130418.50	0.600

Note. Ho: n1-n2 = 0. Ha: n1-n2 > 0. A 95% confidence level was used for this test.

Table 7 showed the results of the chi-square test for association, this test was used to compare the two training sites to the listening modality on the DLPT. Utilizing the significant *p*-value of 0.05 on the chi-square test for association the reported result was a *p*-value of 0.856. For this test we fail to reject the null hypothesis and conclude that there was no significant difference between the two sites.

Table 8 performed a chi-square test for association, this test was used to compare the two training sites to the reading modality on the DLPT. Utilizing again the significant *p*-value of 0.05 on the chi-square test for association the reported result was a *p*-value of 0.081. By a small margin, we again find that we fail to reject the null hypothesis and conclude that there was no significant difference between the two sites.

Table 7
Chi-Square Test for Association: Listening, Site N = 3,044

Fail	Site 1	Site 2	All
Count	5	184	189
Expected count	5.40	183.60	
Residual	-0.4018	0.4018	
Standardized residual	-0.17287	0.02965	
Adjusted residual	-0.1811	0.1811	
Contribution to chi-square	0.0298832	0.0008792	
Pass	Site 1	Site 2	All
Count	82	2773	2855
Expected count	81.60	2773.40	
Residual	0.4018	-0.4018	
Standardized residual	0.04448	-0.00763	
Adjusted residual	0.1811	-0.1811	
Contribution to chi-square	0.0019783	0.0000582	
All	87	2,957	3,044

	Chi-square	*df*	*p*-value
Pearson	0.033	1	0.856
Likelihood ratio	0.034	1	0.855

Table 8
Chi-Square Test for Association: Reading, Site N = 3,044

Fail	Site 1	Site 2	All
Count	0	100	100
Expected count	2.86	97.14	
Residual	-2.858	2.858	
Standardized residual	-1.6906	0.2900	
Adjusted residual	-1.744	1.744	
Contribution to chi-square	2.85808	0.08409	
Pass	Site 1	Site 2	All
Count	87	2857	2,944
Expected count	84.14	2859.86	
Residual	2.858	-2.858	
Standardized residual	0.3116	-0.0534	
Adjusted residual	1.744	-1.744	
Contribution to chi-square	0.09708	0.00286	
All	87	2957	3044

	Chi-square	*df*	*p*-value
Pearson	3.042	1	0.081
Likelihood ratio	5.898	1	0.015

Summary

The study showed that no statistical significant difference between the two training sites was present when comparing the site to the resulting proficiency score on the DLPT. When using the Mann-Whitney test the reported *p*-value resulted in the failure to reject the null hypothesis showing that Site 1 does not produce higher DLPT score than Site 2. During the chi-square test for association the level of 2 on the DLPT was used to show the pass or fail point for the dependent variable. This chi-square test for association also resulted in the failure to reject the null hypothesis. I concluded that there was no significant difference between the training sites.

The next chapter will provide discussion, conclusions, and recommendations for the study. This chapter has specific sections detailing the interpretation findings, limitations, implications of the study, and recommendations for further research. The chapter concludes with a distinct summary of the complete study.

Chapter 5: Discussion, Conclusions, and Recommendations

This study proposed to quantify the results of different methods of L2 training for army linguists by statistically testing proficiency scores, measured by the DLPT, against locations and differing approaches to training. The hypothesis that there was a significant statistical difference in the proficiency levels resulting from the training offered at a DOD-organized school, which is referred to as Site 1, compared with Site 2, where instruction was provided by contractors, served as the base for the study. The research was aimed to find out whether the two training sites were significantly different in trainees' DLPT results. The study found no significant difference between the training sites and either modality tested on the DLPT. This chapter includes an interpretation of the findings, a review of the potential study limitations, and recommendations, concluding with an analysis of the implications of the research.

Interpretation of Findings

The study found that neither the listening nor the reading modality on the DLPT had a significant difference between the two sites of training. There are only additive elements to this research that can inform the army's linguist community. The study did find that Site 1 had no failures in the reading modality compared to Site 2, which had a 3.111% reading failure rate. Even though this was proven to not be statistically significant, there may be elements of information on how reading is taught at Site 1 that could inform the linguist community. A further review of how that site trains for linguist modality may help to fill a gap in the literature.

The theoretical foundation for the study was social representations theory, which centers on beliefs and norm structures formed by different subgroups within a population. Within the army linguist community, there was a subjective view that the two training sites produced a different quality of linguist. However, this assumption had not been quantitatively tested before this study. Based on this study, this assumption can be rejected. This specific result will aid in filling that gap.

In the study, the listening and reading modalities had differing results in total numbers of students passing and failing, as well as differences in the number of students who reached higher levels of proficiency. This could be used to better manage resources in the form of how the organization builds training programs of instruction. Specifically, at Site 1, listening results were higher than reading results. Even with no statistical significance between the two sites, the finding that there were differences between the two modalities is essential to the management of the program.

Limitations of the Study

In addition to the limitations outlined in Chapter 1, limitations were identified during the study. A significant limitation of the study was the absence of information on the pre-training proficiency level and the understanding of any self-study conducted during the training. This limitation could be resolved with recommended future research that incorporates qualitative methods that were unavailable during this study. With Site 1 reporting no failures in the reading modality on the DLPT, the limitation was not found in the timeframe of the study of 4 years but possible due to having only 87 linguists train at

Site 1 compared to the 2,957 trained at Site 2. There were no generalizability limitations found in this study, in that the entire population was used for the study.

Recommendations

Several areas in the army's L2 sustainment program stand out as potential further quantifiable elements that may provide additional research opportunities. This additional research may provide decision makers with the ability to make resourcing and governance decisions. Further study of this area could use different research designs, possibly capturing qualitative analysis that was lacking in this study.

Research to compare the difference between individual languages and their resulting DLPT proficiency levels may refine the overall understanding of how to best develop training strategies. This study treated all languages in the same way to compare the training sites. Internal to each site there may be a significant difference between the different languages or dialects taught at these sites. Using a similar model centered on the chi-square test for association may lead to a better understanding of the internal training site's best practices. As in this research, using a comparison both at the passing rate of 2 on a DLPT and then a higher level of 3 may even further refine this understanding.

Qualitative research connected to the same linguist population study may result in a better understanding of the human dimension connected both to the testing and training aspects of L2 sustainment. Throughout the literature review, a significant body of knowledge existed on the stresses and test-taking abilities of individual linguists. The DLPT is no exception to these factors. By surveying the target population, we may find

that a linguist more comfortable with the test had less test anxiety and therefore had a better result.

Additionally, a potential further qualitative study of teacher and student interaction could provide valuable insight into the training dynamic. An effort to determine the subclassroom task that leads to better results on a DLPT would also aid in the best use of resources. This concept is again rooted in the literature review but was not a factor in the original study. Lastly, connected to the idea of classroom dynamics, further research on how oral communication is taught and tested may build toward a more considerable body of knowledge. The oral proficiency test, which is the test of record in the army for the speaking modality, was not used as a factor in this study.

Implications

The core positive social change from this research is found in the connection to the theoretical framework, which again was rooted in social representation theory. The subjective idea that Site 1 produced better linguists than Site 2 was used as a foundational assumption for subgroups of linguists and potentially impacted policy decisions. However, this research has led to the statistical rejection of this subjective assumption that site 1 is better than Site 1. The area in which this research may have real value is the attitudes of those who train at Site 2. Previously, they may have assumed that they were being provided substandard training. If the results of this study can be included in the framework of the L2 program, the hope is that the perception may be changed. This may have retention and motivation impacts on the individual linguists.

A potential methodological impact of this research involves how resourcing decisions are made. Future selections of training methods should be grounded in evidence-based policymaking. To support the evidence-based policymaking cycle for training planning, this research could offer a model for posttraining validation. In relying on evidence-based policymaking, it may be possible to use the limited resources in the army's L2 program more effectively, and the resulting proficiency levels might be higher. This would be a leap forward from the current practice of subjective or resource-driven decision cycles. Too often in the linguist community, training is designed around what is affordable. The findings of this study could have a positive impact on the army's L2 program.

Conclusions

This study found that there were no statistical differences between the two L2 training sites and methods used for the selection population of U.S. Army linguists. Based on these findings, the subjective social norm that led to the perception that Site 1 resulted in higher quality linguists can now be rejected. This rejection may have a social impact on the retention and motivation of linguists who receive training from Site 1. This research offers a validation model for training, and a key recommendation is centered on the idea that future L2 training decisions should be rooted in evidence-based policymaking.

References

Ainsworth, L., & Viegut, D. (2006). *Common formative assessments: How to connect standards based instruction and assessment*. Thousand Oaks, CA: Corwin Press.

Bauer, A., Braun, M., & Muller, K. (2017). Accurate maximum-margin training for parsing with context-free grammars. *IEEE Transaction on Neural Networks & Learning Systems*, *28*(1), 44-56.

Beam, A. D., & Hodges, J. (2015). Be realistic: A model for home station training. *Army Sustainment*, *47*(5), 41-45.

BenAlaya, D. (2016). Structural theory of social representation in light of a metaphor: Of clouds and clocks. *Papers on Social Representation, 35*(2), 1-8.

Burston, J. J., & Arispe, K. K. (2018). Looking for a needle in a haystack: CALL and advanced language proficiency. *CALICO Journal, 35*(1), 77-102.

Carver, C. (2015, March-April). Army transformation and the role of tables of distribution and allowances. *Army Sustainment*. Retrieved from https://www.army.mil/article/143446/army_transformation_and_the_role_of_tables_of_distribution_and_allowances

Castrillon, L. L. (2017). The effects of Vygotsky's sociocultural theory on second language acquisition and language input. *Espiral: Revista De Docencia E Investigacion, 7*(1), 91-102.

Cox, T. L., & Clifford, R. (2014). Empirical validation of listening proficiency guidelines. *Foreign Lanugage Annals, 47*(3), 379-403.

Dilianian, A., & Akiwowo, T. (2016, July-August). Improving army readiness for the 21st century: An interview with Lt. Gen. (RET) Robert T. Dail. *Army Sustainment*. Retrieved from https://www.army.mil/article/169567/improving_army_readiness_for_the_21st_century

Dulipovici, A., & Robey, D. (2013). Strategic alignment and misalignment of knowledge management systems: A social represetnation perspective. *Journal of Management Information Systems, 29*(4), 103-126.

Dulipovici, A., & Vieru, D. (2015). Exploring collaboration technology use: How users' perceptions twist and amend reality. *Journal of Knowledge Management, 19*(4), 661-681.

Elfiky, S. (2017). Investigating the relationship between students' self-assessment and ratings obtained from a formal oral proficiency interview. *Dialog on Language Instruction, 27*(2), 23-25.

Ellington, J. K., Blume, B. D., Surface, E. A., & Wilson, M. A. (2015). Foreign language training transfer: Individual and contextual predictors of skill maintenance and generalization. *Military Psychology*, *27*(1), 36-51.

Ferrer, I. R. (2014). Reaching higher levels: How we did it: Homework and diagnostic assessment. *Dialog on Language Instruction, 24*(2), 55-76.

Gerencheal, B. (2016). Gender differences in foreign language anxiety at an Ethiopian university: Mizan-Tepi Univeristy third year English major students in focus. *African Journal of Education and Practice, 1*(1), 1-16.

Goodridge, P. (2017). Second-language acquisition and motivation: A literature review. *Pursuit: The Journal of Undergraduate Research at the University of Tennessee, 8*(1), 79-82.

Henter, R. (2012, May). *Metacognitive training for awareness of reading strategies*. Paper presented at the International Conference of Scientific Paper, Brasov, Romania. Retrieved from http://www.afahc.ro/ro/afases/2012/socio/2.2 /Henter.pdf

Höijer, B. (2011). Social representation theory. *Nordicom Review, 32*(2), 3-16.

Holman, D. (2013). Got alignment? Instructional design for field support programs. *Dialog on Language Instruction, 23*, 55-61.

Holman, D. (2017). Across the threshold: A call for ILR proficiency descriptor banks. *Applied Language Learning, 27*(1&2), 1-17.

Hur, H. (2012). Technology and foreign language teachers' professional development. *Dialog on Language Instruction, 24*(1), 23-29.

Johnson, J. L., & Berrett, M. T. (2011). Cultural topography: A new research tool for intelligence analysis. *Studies in Intelligence, 55*(2), 1-22.

Karamanoli, V., & Papachristopoulos, K. (2007). Military personnel and psychological support: An analysis based on social representation theory. *Journal of Military & Veterans' Health, 25*(1), 10-18.

Kim, D. (2013). Foreign area officer: Strategic assets for future success of U.S. forces. *Foreign Area Officer Association Journal of International Affairs, 16*(2), 34-36.

Kim, H., & Cha, K. (2017). Effects of experience abroad and language proficiency on self-efficacy beliefs in language learning. *Psychological Reports, 120*(4), 670-694.

Knight, B. (2012, November). Sustainment preparation of the operational environment planning and assesment tool. *Army Sustainment*, 36-39.

Kurum, Y. Y. (2011). The effect of motivational factors on the foreign language success of students at the Turkish military academy. *Novitas-Royal, 5*(2), 299-307.

Lakoff, R. T. (1975). *Language and woman's place*. Berkeley, Ca: Department of Linguistics University of California.

Marchi, G. Lucertini, G., & Tsoukias, A. (2016). From evidence-based making to policy analytics. *Annuals of Operation Research, 236*(1), 15-38.

Markova, I. (2013). Ethics in the theory of social representation. *Papers on Social Representation, 22*(1), 1-8.

Miller, L. (2017). The relationship between language proficiency and language attitudes evidence from young Spanish-English bilinguals. *Spanish in Context, 14*(1), 99-123.

Miller, Z. F. (2016). Military and civilian L2 instructors: Decoding perceptions of U.S. service academy cadets. *Applied Language Learning, 201*(2), 25- 52.

Mills, D. R. (2011). Signals intelligence and the coder special branch of the Royal Navy in the 1950s. *Intelligence and National Securty , 26*(5), 639-655.

Moloney, G., & Walker, I. (2007). *Social reprensentations and identity.* New York, NY: Palgrave Macmillan.

Morgan, B., & Vandrick, S. (2009). Imagining a peace curriculum: What second-language education brings to the table. *Peace & Change, 34*(4), 510-532.

Nolan, A. (2014). Assessing professional military English language skills in Sweden and its neighboring states. *Baltic Security & Defense Review*, 16(1), 234-253.

Obama, B. (2015). *The national securty strategy.* Washington , D.C., USA: The Office of the President of the United States

O'Dwyer, E., Lyons, E., & Cohrs, J. C. (2016). How Irish citizens negotiate foreign policy: A social representation approach to neutrality. *Political Psychology, 37*(2), 165-181.

O'Sullivan, E., Rassel, G. R., Berner, M., & Taliaferro, J. D. (2017). *Research methods for public administrators* (6th ed.). New York, NY: Routledge.

Oxford, R. L. (2013). *Teaching & researching: Language learning strategies.* New York, NY: Routledge.

Salloum, M. (2016). The social representation of teachers between reality and aspiration. *Journal Plus Education / Educatia Plus, (14),* 204.

Santizo, I. P. (2017). Correlating aptitude with oral proficiency: A quantitative study of DLAB and OPI scores across four language categories. *Applied Language Learning, 27*(1&2), 73-91.

Shih, C. (2013). Cultural competency as a central component in immersion. *Dialog on Language Instruction, 23*, 62-64.

Shirey, W. T., Sullivan, K. T., Lines, B., & Smithwick, J., (2017). Application of lean six sigma to improve service in healthcare facilities management: A case study. *Journal of Facility Management Education and Research*, *1*(1), 9-18

Slabakova, R. (2013). Adult second language acquisition: A selective overview with a focus on the learner linguistic system. *Linguistic Approaches to Bilingualism, 3*(1), 48-72.

Stepanoviene, A. (2016). The importance of selecting a teaching/learning strategy in studies of professional lanuguage. *Public Security and Public Order, 17*, 281-289.

Song, M. (2014). Optimizing interactivity in online course design. *Dialog on Language Instruction, 24*(2), 13-22.

Sparks, R. L., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2006). Native language predictors of foreign language proficiency and foreign language aptitude. *Annals of Dyslexia, 56*(1), 129-160.

Tyler, R. J. (2012). Interpersonal trust and willingness to communicate in the adult foreign language classroom. *Dialog on Language Instruction, 22*(2), 1-21.

U.S. Army. (2016). *Army Regulation 11-6 Total Army Language Program*. Washington D.C.: Headquarter DA G-2.

Waldvogel, D. A., Youtz, B., & Laser, E. (2013). The importance of foreign language & culture training: Should SOF increase training opportunities for this crucial combat skill. *Journal of International Affairs, 16*(3), 9-54.

Web Center for Social Research Methods. (n.d.). *Research methods knowledge base.* Retrieved from http://www.socialresearchmethods.net/kb/index.php

Won, U., & MacDonald, K. (2014). Enhancing teaching through diagnostic assessment. *Dialog on Language Instruction, 24*(2), 53-54.

Wu, S. (2013). Applications of elicited imitation to second language education. *Dialog on Language Instruction, 23*, 51-54.

Xu, M., Fralick, D., Zheng, J. Z., Wang, B., Tu, X. M., & Feng, C. (2017). The differences and similarities between two-sample t-test and paired t-test. *Shanghai Archives of Psychiatry, 29*(3), 184-188.

Zhang, Y., & Zhou, X. (2014). Immersing culture in daily teaching. *Dialog on Language Instruction, 24*(2), 58-59.

Zhou, X., & Ma, R. (2014). Enhancing delivery: A key to improving oral proficiency. *Dialog on Language Instruction, 24*(2), 73-76.

Appendix A: Acronym List

CALL: Center for Army Lessons Learned

DLPT: Defense Language Proficiency Test

DLAB: Defense Language Aptitude Battery

DOD: Department of Defense

FLCAS: Foreign Language Classroom Anxiety Model

FOIA: Freedom of Information Act

ILR: Interagency Language Roundtable

L2: Second Language

OE: Operational Environment

S2R: Strategic Self-Regulation

X: Independent Variable

Y: Dependent Variable