REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
05-12-2013	Technical Report	

4. TITLE AND SUBTITLE

Language Enabled Airmen Program: Language Intensive Training Events 2011 Results

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Rebecca L. Oxford, Ph.D.
Jennifer Tucker, Ph.D.
Andrea Doyle, M.A.
Katie Gunther, M.S.
Rebecca McKenzie, M.A.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

USAF Culture and Language Center
600 Chenault Circle
Bldg 1405
Maxwell AFB, AL 36112-6010

8. PERFORMING ORGANIZATION REPORT NUMBER

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

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The Language Enabled Airman Program (LEAP) is a career-spanning program to sustain, enhance and utilize the existing language skills of general purpose force (GPF) Airmen. The objective of LEAP is to have cross-culturally competent leaders across all U.S. Air Force specialties with working level foreign language proficiency – leaders that can meet USAF global mission requirements. Language Intensive Training Events (LITEs) are a major element of LEAP in which participants engage in intensive language study in immersive environments over a four to six week timeframe. The report provides the findings from a program evaluation project with the goal of determining the changes in foreign language learning, intercultural communication, and motivation factors from pre to post-LITE. Seventy Airmen participated in LITEs in FY 2011 (54 were male and 16 were female and all were officers). The results demonstrated that the participants showed significant improvement in reading and listening proficiency on the Defense Language Proficiency Test (DLPT) and increases in self-assessed speaking proficiency. Additionally, self-reported intercultural communication competence and willingness to communicate both increased significantly from pre to post LITE. Recommended language learning strategies for LITEs are provided, and implications of the findings for the LEAP program are discussed.

15. SUBJECT TERMS

Foreign language learning, immersive training, Defense Language Proficiency Test, intercultural communication, interagency lan

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*

Language Enabled Airman Program

LANGUAGE INTENSIVE TRAINING EVENTS: 2011 RESULTS

AFCLC Technical Report 2013.1

Rebecca L. Oxford, Ph.D.
Jennifer Tucker, Ph.D.
Andrea Doyle, M.A.
Katie Gunther, M.S.
Rebecca McKenzie, M.A.

July 2013

LANGUAGE INTENSIVE TRAINING EVENTS: 2011 RESULTS

EXECUTIVE SUMMARY

Evaluation Requirement

This report documents the outcome evaluation related to the Language Enabled Airman Program's (LEAP) FY 2011 Language Intensive Training Events (LITEs). In addition to programmatic evaluation, the report includes information on survey item development and piloting for use in future LITE evaluations. The report addresses the requirements of both language and culture stakeholders at the Air Force Culture and Language Center (AFCLC) and the United States Air Force (USAF) at large.

Background

The Language Enabled Airman Program (LEAP) is a career-spanning program to sustain, enhance and utilize the existing language skills of general purpose force (GPF) Airmen. The objective of LEAP is to have cross-culturally competent leaders across all U.S. Air Force specialties with working level foreign language proficiency – leaders that can meet USAF global mission requirements. LEAP is a volunteer program and to become a participant in LEAP, Airmen must already possess moderate to high levels of proficiency in a foreign language specified on the Air Force Strategic Language List (SLL). LEAP is managed by the Air Force Culture and Language Center (AFCLC), part of Air University's Spaatz Center at Maxwell Air Force Base, Alabama.

Language Intensive Training Events (LITEs) are a major element of LEAP. LITEs provide opportunities for LEAP participants to engage in intensive language study in immersive environments. LITEs have only been offered since 2010, but the goal is for LEAP participants to take part in a LITE every three to four years. The dual goal of the LITEs is to increase the language proficiency and cultural competence of USAF personnel selected for LEAP, so that they are better able to meet the foreign language needs of current and future USAF operations around the world.

LITEs are typically four to six weeks in length and may be at either CONUS (continental U.S.) or OCONUS (outside of continental U.S.) locations, depending on a participant's language proficiency, AFSC, and site availability. Examples of CONUS LITEs are opportunities with government or military language training detachments, civilian language institutes, and courses at universities. Examples of OCONUS LITEs are commercial language training schools abroad, internships with civilian companies, embassies, and international governmental agencies and organizations, and short term deployments and military exchanges.

Procedure

The program evaluation addressed the evaluation questions reported below by employing a single-group pre-test / post-test design with three elements: pre-survey, program components (LITE experiences), and outcomes/post-surveys.

Seventy Airmen participated in LITEs in FY 2011 (54 were male and 16 were female and all were officers). The AFSCs of the participants varied. Approximately one-third (32.8%) of the participants represented the Intelligence career field while more than 1 out of every 10 was a Mobility Pilot, and 1 out of every 10 was an Engineer.

The measures used in the evaluation include those for language proficiency, affect (attitudes, emotion, and motivation), and cultural competence. In particular, AFCLC developed a speaking proficiency assessment by which participants could assess (both before and after the LITE) their own speaking proficiency based upon the Interagency Language Roundtable (ILR) proficiency level descriptions. For each ILR level, Speaking Can-Do items were written to measure individuals' perceptions of their ability to perform a specific speaking behavior at a particular level, and behaviors increase in difficulty with each subsequent level.

Additionally, based on the provisional ILR intercultural communication competence skill-level descriptions (finalized in 2012; ILR, 2012), we developed a new instrument for measuring participants' self-efficacy in performing intercultural communication behaviors, the Intercultural Communication Behavior Self-Efficacy Scale.

Due to the low number of CONUS LITE participants, comparisons between CONUS and OCONUS outcomes were not conducted for this report.

Findings

The major findings are presented according to questions developed by AFCLC language stakeholders for evaluating the 2011 LITEs.

Evaluation Question 1. What were the characteristics of the LITE participants?

The characteristics of the 2011 LITE participants showed that these individuals were not only well-educated but also linguistically and culturally experienced. When starting the LITEs, almost all participants rated themselves as having a minimum of one language (other than their native language) in which they could communicate at a survival level or beyond. Nine out of every 10 participants who responded to the survey indicated having experienced some cultural education or training, often as part of language instruction but sometimes in other forms.

Evaluation Question 2. How effective were the LITEs in developing language proficiency in listening, reading, and speaking?

The seasoned, highly motivated participants in the 2011 LITEs showed significant improvement in reading and listening proficiency on the Defense Language Proficiency Language Test (DLPT). Pre-LITE reading proficiency started slightly higher than pre-LITE listening proficiency, and reading proficiency levels remained higher than listening proficiency levels at the end of the LITEs. When averaging the degree of change across both listening and reading scores, the greatest increase in DLPT levels occurred for those who had a pre-LITE DLPT of level 1 in listening or reading. On the other hand, individuals with a pre-LITE DLPT of level 3 for both listening and reading showed very little change or a decrease in proficiency, indicating that it is more difficult to increase proficiency at higher levels than at level 1.

[However, we also discuss alternative explanations such that the DLPT does not provide upper-range tests (e.g., ILR levels 3+ and 4) for some languages thus causing a ceiling effect for some measurements. Further, we were unable to control for the timing of the pre and post DLPT assessments. Participants are allowed to take the DLPT only once every six months, and the scores are valid for one year. Thus, the timeframes varied across the participants for when they actually took these assessments.]

Self-assessed speaking proficiency improved significantly from pre-LITE to post-LITE. Nevertheless, both before and after the LITEs, speaking proficiency lagged behind reading and listening proficiency.

Evaluation Question 3. How effective were the LITEs in increasing intercultural communication competence, willingness to communicate, or perspective-taking?

Self-reported intercultural communication competence and willingness to communicate both increased significantly from pre-LITE to post-LITE, although participants reported that they were usually willing to communicate even at the start of the LITEs (i.e., the ratings were high for both pre and post measures). Further, participants who were more willing to communicate during their LITE experience achieved higher self-reported speaking and intercultural communication competence.

No significant differences between pre-test and post-test arose for the two perspective-taking variables, propensity to take others' perspectives or confidence in doing so. However, participants who had greater perspective-taking confidence during their LITE experience achieved higher self-reported speaking and intercultural communication competence.

Finally, self-reported speaking proficiency, willingness to communicate, and perspective-taking confidence were significantly correlated at the end of the LITEs. They seemed to form a highly useful nexus.

Evaluation Question 4. Did the LITEs have any influence on language motivation or language anxiety, and did motivation and anxiety relate to positive outcomes?

The Airmen participating in the LITEs were largely sparked by their own deep interest in the language they were learning (intrinsic motivation), although they also showed an interest in learning the language for external purposes (extrinsic motivation). As language motivation, both intrinsic and extrinsic, was rather strong before and after the LITE, no significant differences emerged from pre-test to post-test. This indicates that motivation was quite constant and the LITEs did not significantly increase or decrease motivation which is consistent with the idea that language motivation may be a less malleable, trait-like construct.

Similarly, no significant pre- to post-LITE difference emerged for language anxiety. Participants were neither particularly anxious nor particularly calm at either the pre-LITE or post-LITE time. However, language anxiety was negatively associated with many of the post-LITE outcomes: listening and self-reported speaking proficiency, intercultural communication competence, willingness to communicate, and perspective-taking confidence.

More rigorous testing showed that participants who had a higher level of motivation and less language anxiety during their LITE experience achieved a higher level of intercultural communication competence than those reporting lower levels of motivation and higher levels of anxiety. Further, participants who had less language anxiety achieved higher levels of speaking proficiency than those with higher levels of anxiety.

Evaluation Question 5. How did the participants evaluate the program features?

The respondents were rather satisfied with the nature of their classroom experiences, although the results possibly point to needed improvement in technology use and greater consistency in making participants accountable for speaking the target language.

Evaluation Question 6. Did any program features significantly relate to the outcomes?

The DLPT reading proficiency change score was marginally related with the perception of being held accountable for speaking the target language and more strongly related with the integration of technology in instruction. The DLPT listening proficiency change score was marginally related to the perception that the instructor utilized a variety of activities and teaching methods to deliver the material. Speaking competence was not related to any of the program factors.

In addition, two qualitative evaluation questions were asked.

Pre-LITE Qualitative Question 1. What did the participants expect from the LITEs?

Before the LITEs, participants said they hoped to make gains in the area of language proficiency, with speaking and listening being very powerful aims. Many participants hoped to improve their cross-cultural competence. Other expectations, such as having an opportunity to contribute to the USAF mission, also were mentioned by participants.

Post-LITE Qualitative Question 2. What types of cultural or communication challenges did the participants encounter during their LITEs?

Challenges experienced by participants, according to their own narrative comments, were often about cultural issues, followed by communication issues.

Recommendations

For greater language proficiency, LITE teachers should teach participants to use optimal learning strategies for reading, listening, and speaking. For example, reading comprehension only occurs when the individual actively reflects on what he or she is reading by using metacognitive strategies, such as planning, organizing, and evaluating their reading. It would be useful to find out which reading strategies the LITE participants use and to help these participants use the most optimal strategies for the reading tasks they face during the LITEs and thereafter.

We call for LITE instructors to work more on the listening skills of participants, so that listening can be as strong as reading. If teachers do not know how to teach listening, they require information and skills in this area. There are many existing resources that provide excellent suggestions for weaving listening strategy instruction into regular language teaching and provide assessment tools for foreign language listening strategies.

Speaking is often assumed to be more difficult than reading and listening, and the 2011 LITE results supported this assumption. It is particularly necessary to teach speaking to LITE participants. To help participants improve their speaking proficiency, LITE teachers must teach vocabulary-learning strategies and must recognize the importance of vocabulary access for students. Additionally, online or in-person teacher development could instruct LITE teachers how to weave such speaking-related strategy instruction into regular language instruction to spark faster progress in speaking proficiency.

As this evaluation was based solely on participant perspectives of the instructional and program features of the sites, future evaluation efforts should include data from LITE site administrators and teachers regarding the language learning pedagogy, curriculum, and assessments. The LITE sites need to be evaluated to identify what is working effectively and what needs to be changed.

The fact that the DLPT does not measure higher-range proficiency for most languages put some LITE participants at a disadvantage, since an increasing percentage of participants enter the LITEs with proficiency at 2+ or higher in reading and listening. We strongly encourage a systematic effort to create DLPTs in all strategic languages with higher proficiency "ceilings" up to level 4 or 5, so that individuals who start a LITE at level 2+ or 3 can have a proper assessment on the DLPT. There is a need for better language assessment at higher levels either in lieu of or to supplement the DLPT.

Finally, self-assessment of language proficiency, as accomplished in this evaluation, is not as desirable as an actual speaking performance assessment. LEAP staff should clearly mandate the use of Oral Proficiency Interviews (OPIs) or a computerized version of OPIs, because there are limitations associated with self-assessment of language proficiency. Greater effort and energy should be put into finding an appropriate speaking proficiency assessment tool.

Use and Dissemination of Findings

The findings of this report will be used to help the AFCLC staff make assignments of students to LITEs, design and deliver LITEs, and make decisions about the use of existing survey instruments in future LEAP LITEs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Section 1: Introduction

Purposes of the Report

The Language Enabled Airman Program (LEAP) is a career-spanning program to sustain, enhance and utilize the existing language skills of general purpose force (GPF) Airmen.[1] The objective of LEAP is to have cross-culturally competent leaders across all U.S. Air Force specialties with working level foreign language proficiency – leaders that can meet USAF global mission requirements. LEAP is a volunteer program originally only open to officers but available now to both officers as well as select enlisted career fields, or Air Force Specialty Codes (AFSCs) since September 2011. To become a participant in LEAP, Airmen must already possess moderate to high levels of proficiency in a foreign language specified on the Air Force Strategic Language List (SLL). LEAP is managed by the Air Force Culture and Language Center (AFCLC), part of Air University's Spaatz Center at Maxwell Air Force Base, Alabama.

Language Intensive Training Events (LITEs) are a major element of LEAP. LITEs provide opportunities for LEAP participants to engage in intensive language study in immersive environments. LITEs have only been offered since 2010, but the goal is for LEAP participants to take part in a LITE every three to four years. Further definition and details of LITEs are provided later in this section.

Evaluating the LITEs

Program evaluation applies systematic research methods to program design assessment, implementation, and effectiveness (Yarbrough, Shulha, Hopson & Caruthers, 2011). To the extent possible, the present study evaluates the effectiveness of the FY 2011 LITEs. By "effectiveness," we mean the extent to which the FY 2011 LITEs influenced the development of participants' language proficiency and cultural competence. (Because of limitations in sample size, we do not compare specific LITE sites.)

Reporting the Quality of New Measurement Scales

In addition to evaluation of the LITEs, this report also aims to assess the reliability and validity of program-related and student-attribute measures, so that decisions can be made about using these measures in the future. Several measures were purposefully developed and pilot-tested during the 2011 LITEs. This report describes the theoretical foundations and psychometric properties of these scales.

Background: The Need for Linguistically and Culturally Competent Airmen

The USAF must increase its inventory of Airmen who have strong language capabilities across all career fields to meet current and future demands of the global mission. This is a challenge for the following reasons. First, since 1947 the USAF has primarily focused on producing only small cadres of foreign language experts in specific AFSCs, such as "crypto-linguists" (intelligence collectors). Second, the USAF has experienced difficulty filling GPF,

[1] See Appendix A for a full listing of acronyms and definitions used in this report.

non-linguist language-coded billets with Airmen who speak the desired language, both officers and enlisted. Third, since the USAF has not traditionally been expeditionary in nature, few saw the need for the GPF to engage directly with native populations. However, the U.S. military's recent experience with combatants, allied partners, and civilian nationals in Iraq and Afghanistan has revealed the necessity of foreign language capability among all services as a critical skill and a force multiplier during deployments and contingency operations. The continued poor track record of filling language-coded billets with qualified Airmen further highlights such issues. Language skills should be integrated into operations at all levels to ensure those Airmen placed in posts requiring foreign language competence are actually able to accomplish the mission and those forces who deploy possess the essential ability to understand and effectively communicate with native populations, local and government officials, and coalition partners when in theater. Language proficiency can no longer be the domain only of a small group of Airmen, the assumption can no longer be made that other nations will unquestionably speak English, and we can no longer rely heavily on contracted interpreters to fulfill language needs.

Addressing the Challenge through LEAP

In 2009 LEAP was established so that the USAF could develop a large group of officers (and later, enlisted personnel) capable of operating effectively in foreign languages within culturally complex environments. LEAP provides training for language acquisition, development, and sustainment. Over a 10-year period, LEAP aims to support 4,000 active duty Airmen, officers and enlisted personnel, in the development of a minimum level of "2+/2+/2" proficiency, referring to Interagency Language Roundtable (ILR) level 2+ in reading and listening and level 2 in speaking on a 5-point scale. For the entering LEAP participants who already meet these levels (or higher) in their respective target languages (languages being learned), their goal is to reach general or advanced working proficiency in those languages. In some cases, high-proficiency learners in one language are offered the opportunity to learn another language on the SLL.

LEAP helps fill critical operational gaps for combatant commands with regard to foreign language communication in partnership-building, irregular warfare, and security force agreements. This helps meet USAF requirements for language-coded billets within the GPF and increases the "bench" of language-capable Airmen who can be quickly identified for contingency operations. No matter what parts of the world become strategically important, LEAP will continue to offer support in relevant languages to meet the global mission of the USAF.

LEAP's Developmental Approach

LEAP's "deliberate development" model consists of phases designed to target and retain "willing and able" participants who can best fill the foreign language needs of the Air Force. LEAP first selects qualified airmen into the program, it provides these participants training in their selected language(s) through the eMentor program (described below) and LITEs, and it also facilitates the use of these airmen for specific language-coded billets and other USAF language requirements.

Vetting and selection. Volunteer Airmen are vetted through a boarding process to ensure only quality candidates are selected into the program who not only possess language capability in a language, but are also viable investments for the Air Force. Selection is based on the Airmens' current language capabilities as substantiated by DLPT and/or OPI scores, potential to achieve higher levels of proficiency as indicated by past performance in language courses, specific Air Staff language requirements for their AFSCs, any in-country experience, performance in active duty or cadet duties, and Commanders' recommendations.

The LEAP Boards since 2009 have resulted in varied numbers of selectees, as shown in Table 1. The majority of those who participated in a 2011 LITE were from Board 2, and a few were from Board 1. Many LEAP selectees from Board 1 took part in a LITE in 2010. Demographics and other characteristics of the 2011 LITE participants appear in Section 2.

Table 1. Selectees for Each LEAP Board

LEAP Board	Date	Number of Applicants	Number of Selectees
1	23 Feb 2010	400 +/-	192 Officer
2	23-24 Sept 2010	404	290 Officer
3	14-15 Sept 2011	503	487 Total 480 Officer 7 Enlisted
4	April/May 2012	101	53 Total 31 Officer 22 Enlisted

eMentor Program. LEAP participants commit to sustaining their language skills as part of a structured training program throughout their career while pursuing their AFSC assignments. Through the eMentor Program that was launched in 2010, Airmen are paired with highly qualified language instructors and a cohort of other LEAP participants within their language and skill level for live (synchronous) distance learning sessions. Participants log into a virtual classroom and use a webcam, headset and microphone, and chat features to communicate with their eMentor instructor and the other participants in their cohort. The lessons include the use of PowerPoint presentations, video clips, audio clips, websites, news articles, whiteboard, out-of-class work and occasionally textbook use.

LITEs. LITEs are immersive language opportunities that can take several forms based on a participant's language proficiency, AFSC, and site availability. LITEs are typically four to six weeks in length and may be at either CONUS or OCONUS locations. CONUS LITEs are typically designed for those below a 2 level DLPT or OPI score, and OCONUS LITEs are typically designed for those at or above a 2 level DLPT or OPI score. Examples of CONUS LITEs are opportunities with government or military language training detachments, civilian language institutes, and courses at universities. Examples of OCONUS LITEs are commercial language training schools abroad, internships with civilian companies, embassies, and international governmental agencies and organizations, and short term deployments and military

exchanges. Ideally LEAP participants will take part in a LITE every three to four years. Frequency is based upon funding and participant availability.

Utilization. As an Air Education and Training Command (AETC) unit, the AFCLC is directed to provide Language, Region, and Culture (LRC)-qualified General Purpose Force (GPF) airmen to the USAF through training and tracking. LEAP is the program through which the AFCLC accomplishes this charge. As a training program, LEAP is not designed to place its participants into language-coded billets, however, it does facilitate this action. There are several ways that the LEAP office works with USAF and Department of Defense entities to place LEAP participants into appropriate positions. First, the LEAP office regularly briefs the functional managers at the Air Force Personnel Command (AFPC) about LEAP participants and how to use such information as a participant's LEAP Special Experience Identifier (SEI) to put him or her in an appropriate language-coded billet. The LEAP office also assists the Secretary of the Air Forces (SAF) International Affairs (IA) office vet their Regional Affairs Specialist (RAS) applicants, providing SAF/IA information on applicants who are LEAP participants. The LEAP office is also working closely with SAF/IA to align better with their training and SEI programs, with LEAP envisioned as a pipeline for future RAS officers. The LEAP office also collaborates and trains with the USAF's Mobility Support Advisory Squadrons (MSAS) and the United States Special Operations Command (USSOCOM) and envisions LEAP to also be a pipeline for their Airmen as well. Finally, LEAP is in the process of constructing an interactive database whereby Commanders and units in need of language-enabled Airmen can search for eligible Airmen to fill their language needs.

2011 LITE Evaluation Questions

AFCLC language stakeholders collaborated to outline the following quantitative questions for evaluating the 2011 LITEs (see Section 3 for these results):

Evaluation Question 1. What were the characteristics of the LITE participants?

Evaluation Question 2. How effective were the LITEs in developing language proficiency in listening, reading, and speaking?

Evaluation Question 3. How effective were the LITEs in increasing intercultural communication competence, willingness to communicate, or perspective-taking?

Evaluation Question 4. Did the LITEs have any influence on language motivation or language anxiety, and did motivation and anxiety relate to positive outcomes?

Evaluation Question 5. How did the participants evaluate the program features?

Evaluation Question 6. Did any program features significantly relate to the outcomes?

In addition, two qualitative evaluation questions were asked, one before the LITE and one after the LITE, with results provided in Section 4.

Pre-LITE Qualitative Question 1. What did the participants expect from the LITEs?

Post-LITE Qualitative Question 2. What types of cultural or communication challenges did the participants encounter during their LITEs?

Summary of Section 1

This section has provided introductory information about LEAP and about the LITEs, which are a major component of that program. We have also presented here the six questions guiding the 2011 LITE evaluation. The following section explains the methodology used in the evaluation.

Section 2. Methodology of the 2011 LITE Evaluation

This section presents the methodology for evaluating the 2011 LITEs. Sections include the overall evaluation design, settings and participants, measurement instruments, data collection procedures, and data analysis procedures.

Overall Evaluation Design

Figure 1 displays the evaluation design used to answer the evaluation questions. The design has three elements: pre-survey, program components (LITE experiences), and outcomes.

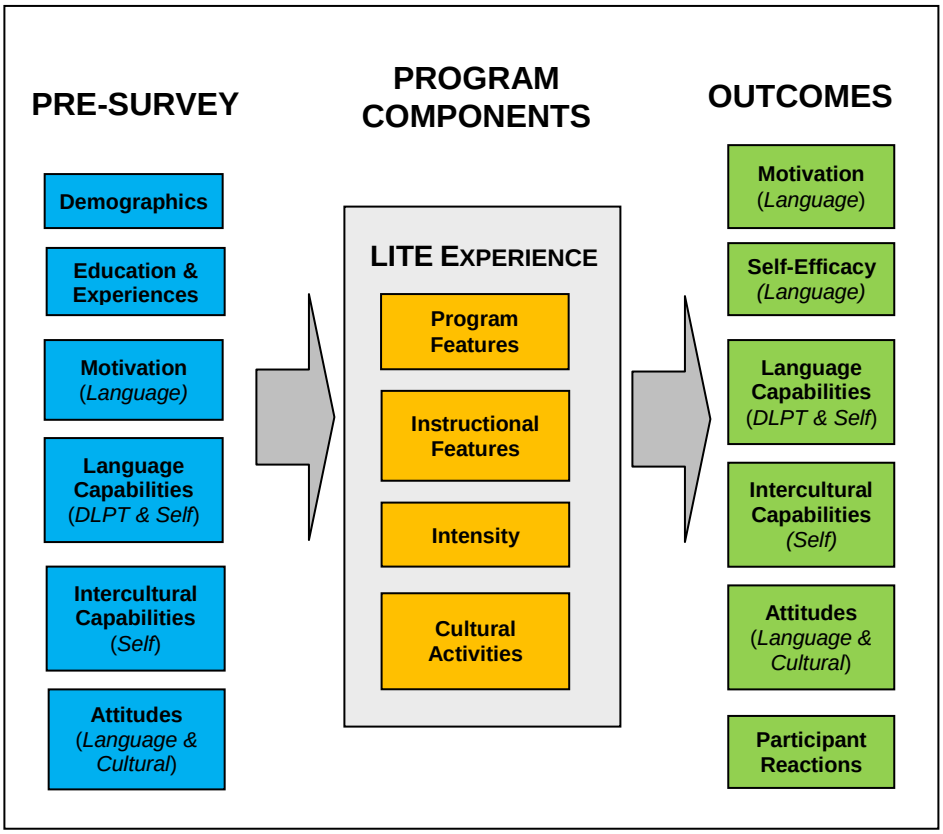

Figure 1. Overall Evaluation Design

Measurement of effectiveness (see the evaluation questions listed in Section 1) depends on having appropriate data at the right time. It is also useful to have an experimental design, in which randomization is used to place individuals into comparison and experimental groups, with only the latter group receiving the LITE. However, this was not possible due to practical limitations (only certain LEAP Airmen who are eligible for LITEs during a given timeframe due to DLPT levels, availability of OCONUS locations, etc.). This is therefore a single-group pre-test / post-test design (please see the discussion section for LEAP's 5-year assessment plan which will replace the current methodology). Without a comparison group, it is difficult or impossible to attribute the changes to the LITE in any definitive sense, but we can make strong speculations about

where the changes might have originated. Therefore, when we talk about "the effects of the LITEs," please understand that other unmeasured factors might have also had an effect (e.g., language learning activities performed by the participants independent of LEAP, eMentor assignments prior to the post DLPT).

Settings and Participants

The LEAP LITE settings for 2011 were as follows:

- A2Z Language Schools (various locations OCONUS)
- AFCLC (taught by instructors from the Defense Language Institute Foreign Language Center), Maxwell AFB, Alabama
- Royal Danish Air Force
- Defense Attaché Offices in Brazil, Thailand, China, and the Philippines
- Office of Defense Cooperation in Croatia and Portugal.

As shown in Table 2 below, of the 73 LITE participants, the majority (45) attended A2Z Language Schools in varied parts of the world. A2Z Language Schools are well known to AFCLC because they are the schools traditionally contracted to teach U.S. Air Force Academy study abroad students. Twelve highly proficient speakers were assigned to internships through the Defense Attaché Office, two had internships through the Office of Defense Cooperation, and one had an internship with the Royal Danish Air Force.

In the total group of 70, 54 were male and 16 were female. Twelve were second lieutenants, 21 were first lieutenants, 23 were captains, 10 were majors, three were lieutenant colonels, and one was a colonel. Commissioning sources differed: 35 of the participants graduated from USAFA, 22 received commissions through the Reserve Officer Training Corps (ROTC), and 13 were commissioned through Officer Training School (OTS).

The AFSCs of the participants varied. Approximately one-third (32.8%) of the participants represented 14N, Intelligence. More than 1 out of every 10 was a Mobility Pilot (11M), and 1 out of every 10 was an Engineer (62E). Smaller numbers of participants came from the following career fields:

- 38F, Force Support;
- 13S, Space and Missile;
- 17D, Cyberspace Operations;
- 21A, Aircraft Maintenance;
- 21R, Logistics Readiness;
- 11S, Special Operations Pilot;
- 12 M, Mobility Combat Systems;
- 12R, Recce/Surv/Elect Warfare Combat Systems;
- 16F Regional Affairs Strategist;
- 16P, Political-Military Affairs Strategist;
- 31P, Security Forces;
- 32E, Civil Engineer;
- 41A, Health Services Administrator;

- 43B, Biomedical Scientist;
- 43E, Bioenvironmental Engineer;
- 48G, General Medical Officer / Flight Surgeon;
- 63A, Acquisition Manager; and
- 64P, Contracting.

Thus, the LITEs were populated by very diverse individuals.

Table 2. 2011 LITEs: CONUS/OCONUS, Sites, Languages by Site, Number of Participants by Site, and Number in Sample by Site

Location	Site	Languages	# of Participants	# in Sample
CONUS				
	Maxwell AFB (DLI)	Chinese-Mandarin French , German, Russian, Spanish	13	11
OCONUS				
	A2Z Language Schools	Chinese-Cantonese, Chinese-Mandarin, Czech, French/Senegal, German, Hungarian, Italian, Arabic/Morocco, Portuguese-Brazilian Spanish	45	43
	Defense Attaché Office	Chinese-Mandarin, Portuguese-Brazilian, Tagalog, Thai	12	12
	Royal Danish AF (internship)	Danish	1	1
	Office of Defense Cooperation	Serbo-Croatian	2	2
	Total		**73**	**69***

Note. *This sample reflects all LITE participants who responded to the pre- or post-survey, not necessarily both surveys.

Measurement Instruments

The measures used in the study include those for language proficiency, affect (attitudes, emotion, and motivation), and cultural competence. Appendix D contains item-level descriptive data on all measures.

Reading and listening proficiency assessment instruments. The evaluation employed measures to assess reading, listening, and speaking proficiency. The first two were measured by the DLPT, and the third by the Speaking Can-Do Scale created for the present research. All three are based on ILR Skill-Level Descriptions. The skill levels are shown in Table 3.

Table 3. ILR Language Proficiency Skill Levels and Their Meanings

Level or Partial Level	Meaning	Comments
0	No Proficiency	Measured by all DLPTs
0+	Memorized Proficiency/ Survival Proficiency	
1	Elementary Proficiency / Minimally Functional Proficiency	
1+	Elementary Proficiency Plus / Minimally Functional Proficiency Plus	
2	Limited Working Proficiency	
2+	Limited Working Proficiency Plus	
3	General Professional Proficiency	
3+	General Professional Proficiency Plus	Measured by only a very small number of DLPTs (very few languages)
4	Advanced Professional Proficiency / Distinguished Proficiency	
4+	Advanced Professional Proficiency Plus / Distinguished Proficiency Plus	
5	Functionally Native Proficiency / Educated Native Proficiency	

The DLPT (Lower Range), Version 5 (DLPT 5), is a three-hour, computer-based test designed to measure reading and listening proficiency in a foreign language (DLI, n.d.). Reading is one subtest, and listening is the other. Results are reported separately for reading and listening. Raw scores, which are theoretically expressed as equal intervals, are translated into ILR proficiency levels and partial levels (e.g., 0, 0^+, 1, 1^+, 2, and so on) for reporting proficiency results. In order to create a numerical mean for listening and reading for parametric tests, "+" designations of the DLPT were converted to 0.5 for all Airmen.

The DLPT lower-range test assesses language proficiency at levels 0 to 3. As noted in Table 3, only a handful of languages have upper-range DLPTs, which are capable of assessing proficiency beyond the level of General Professional Proficiency (level 3). DLI FLC determines which languages have only lower-range tests, upper-range tests, or both (DLI (2012). In a practical sense the lack of readily accessible, high-level DLPT assessment of proficiency for most languages means that a LITE participant whose actual proficiency is General Working Proficiency Plus (3+), Advanced Professional Proficiency (level 4), or higher would not be able to demonstrate that level of proficiency on the DLPT in most languages. This is a serious limitation of the DLPT for the increasing number of LITE participants whose pre-LITE DLPT in listening or reading is already at level 3. For such individuals, the DLPT cannot demonstrate any pre-LITE to post-LITE progress due to the “ceiling effect” caused by lack of measurement past level 3 in most languages. (For the small number of languages that do have DLPT measurement at levels 3+, 4, 4+, and 5, there is no ceiling effect.)

Some other psychometric facts are also very relevant here.

1. It takes more time and effort to move up a whole or partial level on the DLPT (or on any other ILR-based measurement) when one's proficiency starting point is higher than when it is lower. For instance, when an individual starts at level 2 proficiency, it is much more difficult to move to level 3 proficiency than it would be if that same individual were starting at level 0+ and moving to level 1 (Clifford, 2010). Specifically, Clifford reported that it takes approximately four times the number of hours of intensive study to move from level 2 to level 3 than it takes to move from level 0+ to level 1. See Appendix E for details.

2. Language category (the categories are I, II, III, and IV), which reflects the difficulty to learn a language for native speakers of English, also makes a difference in the amount of time needed to move to a higher level (Clifford, 2010). Compared with Category I languages like French and Spanish, Category IV languages like Chinese and Arabic take approximately three times the number of hours to reach a given level of proficiency. See Appendix E for information.

3. Neither LITE participants nor LEAP staff members have access to DLPT raw scores. Instead, the DLPT results reported to LITE participants and to LEAP are expressed only as a single, compiled score by levels and partial levels (0, 0+, 1, 1+, 2, etc.), as noted earlier. Though the DLPT raw scores supposedly represent (or come closer to) equal-interval measurement, DLPT results expressed as levels and partial levels do not represent equal intervals of proficiency. Technically speaking, this restricts the types of statistical tests that are possible to use with DLPT data.

4. DLPT results viewed as levels and partial levels are ordinal (rank order) data. Because they do not reflect equal intervals, the only way we can consider them is to say that 2+ is greater than 2, and 2 is greater than 1+, but we cannot argue that these differences are in any way equal in terms of underlying language ability or the amount of time or energy expended to develop such ability. Therefore, for ordinal data, parametric statistics – the type used with equal-interval data – are not technically ideal, so non-parametric statistics are often used.

Speaking proficiency self-assessment instrument. The DLPT does not provide an assessment of speaking ability in the foreign language. The Oral Proficiency Interview (OPI), based on the ILR language proficiency skill-level descriptions for speaking, is a standardized test that assesses functional speaking ability in a foreign language. The USAF provides a means for LEAP participants to complete the OPI in their target languages. However, given that the OPI must be administered in-person by an interviewer who is proficient in the particular language, the logistics of scheduling and completing the OPI occasionally become difficult. Therefore, LEAP developed a speaking proficiency assessment by which participants could assess (both before and after the LITE) their own speaking proficiency based upon the ILR proficiency level descriptions. The format of the assessment is a "Can-Do" scale, in which each participant indicates whether he or she can perform specific speaking behaviors in the target language.

Items for the *Speaking Can-Do Scale* are aligned to the ILR proficiency levels 1 to 4. As the levels move upward, speaking proficiency becomes progressively richer, more complex, and more elaborate. The proficiency levels reflect both fluency and accuracy. Each higher level implies that the individual has control of the functions of the previous level(s). For instance, an individual at level 3 in speaking is expected to perform functions at levels 1 and 2 in speaking. Level 1 represents Elementary Proficiency. At level 1, individuals can express courtesies and maintain very simple face-to-face conversations, but their vocabulary is extremely limited. At level 2, entitled Limited Working Proficiency, individuals can successfully meet very basic work requirements, converse about their work or families, and give basic instructions, but they still speak awkwardly and have trouble with complicated grammatical structures and vocabulary. Individuals at level 3, known as General Professional Proficiency, can hold conversations in varied settings about numerous topics, can use structure and vocabulary accurately, and can employ a wide lexical range, including cultural references and nuances, but imperfections are still noticeable. At level 4, Advanced Professional Proficiency, individuals can speak fluently and accurately on all levels relevant to professional work and can very effectively control vocabulary, although with a few rare difficulties. Appendix B shows further descriptions and examples of behaviors for each level. Note that Level 0 is not included.

For each ILR level, Speaking Can-Do items were written to measure individuals' perceptions of their ability to perform a specific speaking behavior at a particular level.[2] The behaviors described by the items are representative of what an individual must be able to do at the designated level, and the behaviors increase in difficulty with each subsequent level. For levels 1 and 2, eight items (including one reverse-coded item) per level were created. Nine items were written for level 3 and nine more for level 4. In total, the Speaking Can-Do Scale consists of 34 items. These items were randomized in respect to the order they were presented to LITE participants on the pre- and post-surveys. We list below the Speaking Can-Do Scale items by proficiency level.

Level 1

1. I find it difficult to speak so that I can be understood by a native speaker. (reverse-coded)
2. I can arrange for a hotel room or taxi ride.
3. I can give or ask for simple directions (e.g., to a hotel, restaurant, or bank).
4. I can buy a needed item, such as a bus or train ticket, groceries, or clothing.
5. I can ask and answer simple questions about date or place of birth, job, and other demographics.
6. I can use appropriate expressions for greetings and farewells.
7. I can order a simple meal.
8. I can introduce myself and others.

Level 2

1. I can give straightforward orders or commands.
2. In casual social situations, I can discuss at length current events or activities and family-related topics.
3. I can talk about some work-related topics that are non-routine, such as requesting complicated changes in travel arrangements.
4. I can give complicated, detailed, and extensive directions.
5. I can ask and answer predictable questions on the job.
6. At work I can participate in ordinary discussions using familiar vocabulary and grammar.

[2] We did not use the ILR's own self-assessment scale for speaking (ILR, 2009a; also contained in Oxford, 2010) because we decided it was not sufficiently well written or detailed for our needs.

7. When trying to discuss complex topics, I make frequent errors that seem to disturb native speakers. (reverse-coded)
8. I can discuss my present or most recent job in some detail using ordinary, non-specialized vocabulary.

Level 3

1. Without having to search for words or phrases, I can speak clearly and freely so that native speakers easily understand me.
2. I can explain official (U.S. or Air Force) policies.
3. When giving a presentation or speech, I can use organizing phrases to make my speaking clearer (e.g., "The five key points are...," "Therefore...," "In conclusion...," or "To summarize ...").
4. I can participate in conversations about practical and social topics, whether in formal or informal settings.
5. I can effectively answer concerns about policy decisions.
6. I can readily give presentations about professional topics.
7. I can ask native speakers the appropriate questions to get the information I need, including details and informed opinions.
8. I can speak well enough that native speakers seem to understand me.
9. I can quickly repair a conversation if there is a cultural misunderstanding.

Level 4

1. I can serve as an informal interpreter in a range of unpredictable circumstances.
2. I can speak fluently and accurately about complex procedures related to my AFSC.
3. I can communicate successfully in virtually all professional situations, no matter how challenging, including giving detailed presentations, official speeches, or media interviews.
4. I can readily and appropriately use cultural references, shades of meaning, and figures of speech in conversation.
5. I can readily explain to a native speaker any cultural references, shades of meaning, and figures of speech from my own culture.
6. My errors are rare and do not hinder my work performance.
7. I can speak fluently and accurately about complex, non-work topics of interest to well-educated native speakers.
8. I can change my tone to make it appropriate for spoken interactions in virtually all settings, whether formal or informal.
9. I can appropriately do any of the following things: advise, command, argue, persuade, and negotiate.

On the Speaking Can-Do Scale, participants indicated their level of proficiency on each item by responding either "Yes, this describes me most of the time" or "No, this does not describe me most of the time." Respondents were encouraged to select the former response only if they were able to perform the behavior regardless of situation or context rather consistently and successfully.

To determine the extent of expected patterns of participants, a *scalogram analysis* was performed (Guttman, 1950). The scalogram analysis was intended to provide initial validation for the Speaking Can-Do Scale based on the fact that certain patterns of response are expected on a proficiency scale that includes multiple levels, and other response patterns are not expected. Appendix B shows the procedures and results of the scalogram.

In addition to language proficiency, we also assessed some important affective variables. These are described next.

Instruments for motivation, willingness to communicate, and anxiety. Training and education theorists have long considered that affective factors (e.g., motivation, attitudes, and emotions) influence the extent and manner in which students learn and develop. Therefore, during both pre- and post-surveys, participants were asked to respond to items measuring motivation, willingness to communicate, and language-related anxiety. We wanted to determine whether these variables were associated with changes in language proficiency and other variables. In this section we describe in detail the LITE evaluation's measurement of motivation, willingness to communicate, and anxiety.

For practical purposes the number of items on these scales was kept as small as possible without sacrificing the construct validity of the measure. Internal consistency reliability estimates (using Cronbach's alpha) and item-total correlations were calculated for all scales. We refined the scales by dropping or revising items that were unreliable (having low item-total correlations). The traditional rule of thumb is that reliabilities of .70 or above are considered acceptable (George & Mallery, 2003).

Language motivation. Motivation is the impetus or spark for goal-directed behavior (Petri & Govern, 2004), and it also involves persistence over time. Unlike some common views, motivation is highly dynamic or changeable. In language training contexts, motivation influences the degree to which students engage in the language instruction and with people of the target language culture, both of which influence the learning they will derive from the training. Language learning motivation, usually shortened to language motivation, was assessed.

A learner's motivation is shaped by a host of contextual factors (Gardner, 1982, 1985; Gardner, Tremblay, & Masgoret, 1997; Oxford & Shearin, 1994), such as proximity to the target-language community, language resources available, nature of interactions with language teachers and peers, prestige of the language, and cultural values and social power dynamics. Motivation to learn the language is also influenced by the learner's own emotions, beliefs, and attitudes. For this research, items to measure motivation were influenced by Noels, Pelletier, Clement, and Vallerand's (2003) Language Learning Orientations Scale - Intrinsic Motivation, Extrinsic Motivation, and Amotivation Subscales (LLOS-IEA) and the Self-Determination Theory of Ryan and Deci (2000).

The LITE motivation scale included 20 items overall, with an internal consistency reliability of .82. Some of the motivation items were designed to assess intrinsic motivation, while the rest were intended to measure extrinsic motivation. All had the following stem: "Learning the target language is important to me . . . ." A five-point Likert scale was used for responses with the following options: 1 = Very Untrue of Me, 2 = Untrue of Me, 3 = Sometimes True and Sometimes Untrue of me, 4 = True of Me, 5 = Very True of Me.

According to Ryan and Deci (2000), in *intrinsic motivation*, the activity is undertaken for its inherent interest and enjoyment. More specifically, intrinsic motivation can only occur when the activity holds intrinsic interest, novelty, challenge, and/or aesthetic value and when the person experiences autonomy, competence, and a sense of personal or social relatedness (Ryan & Deci, 2000). Research shows that both tangible rewards and negative impacts, such as threats, severe deadlines, and imposed rather than personally accepted goals, can suppress the

sense of autonomy that is necessary for intrinsic motivation (Ryan & Deci, 2000). The LITE *intrinsic motivation subscale* had an internal consistency reliability of .75 and contained the following seven items:

1. . . . for the satisfaction I experience in knowing the language better
2. . . . because I enjoy learning new information about the culture
3. . . . for the pleasure I get when I can do more with the language
4. . . . for the enjoyment I experience when I grasp something difficult in the language
5. . . . for the sense of achievement when I understand or say something new in the language
6. . . . for the "high" I experience when using the language
7. . . . for the mental stimulation of learning the language.

In *extrinsic motivation*, the activity is done to attain an outcome that is separable from the activity itself. The learner takes instrumental action to do something (receive rewards, avoid negative judgments, embody military values, or achieve other goals, such as helping others, building relationships, or becoming culturally effective) through learning the language. Extrinsic motivation in Ryan and Deci's model includes multiple aspects: external regulation, introjected regulation, identified regulation, and integrated regulation, discussed below. In the LITE evaluation, the extrinsic evaluation subscale contained 13 items and had an internal consistency reliability of .74. Its components were as follows.

External regulation involves doing a language activity to satisfy an external demand or an external reward contingency (like the "Ought To" self, Dörnyei, 2003). The three external regulation items in the LITE evaluation included:

1. . . . to get Air Force foreign language proficiency pay
2. . . . to increase my promotion opportunities later
3. . . . to have more opportunities for traveling in the Air Force.

Introjected regulation involves doing a language activity to avoid guilt or anxiety or to demonstrate ability (or avoid failure) in order to maintain self-worth and enhance ego. The LITE evaluation used two items measuring introjected regulation:

1. . . . to avoid feeling bad if I cannot communicate with people in their native language
2. . . . to avoid seeming like an "ugly American" who does not know other languages or cultures.

Identified regulation involves doing a language activity because of conscious valuing and acceptance of a behavioral goal that was originally specified by others (e.g., the language activity is valued because it is connected to a desire to be a good officer or to embody institutional values) (see also the "Ideal Self," Dörnyei, 2003). Four items measured identified regulation in the LITE evaluation:

1. . . . to influence other countries through knowledge of language and culture
2. . . . to influence other militaries as we conduct joint operations
3. . . . because my image of an ideal Air Force officer is one who speaks more than one language
4. . . . because professionally competent Airmen can get along in other cultures.

Integrated regulation involves doing a language activity because regulation from the outside is now fully and completely integrated into the self, after having been evaluated and brought into congruence with the person's other values and needs. Four items assessing integrated regulation in the LITE evaluation were:

1. . . . to be ready in case I can help my country by using the language
2. . . . to help the Air Force with cross-cultural crises or problems
3. . . . to build relationships with people in the culture
4. . . . because I want to be more effective in the culture.

In many language studies, language proficiency is much greater when motivation becomes increasingly self-regulated or self-determined, so intrinsic motivation/regulation or certain internal-type forms of extrinsic motivation – identified and integrated regulation – are beneficial. Lack of motivation (*amotivation*) works against development of language proficiency, because amotivation is nonintentional, nonvaluing, and shows lack of control and involvement. Amotivation was not assessed in this evaluation.

Willingness to communicate. Language motivation is inherent in *willingness to communicate* (WTC), i.e., the probability an individual will choose to initiate communication when free to do so (MacIntyre, 2003). Factors influencing WTC include target language role and vitality; intergroup motivation, attitudes, and climate; integrative motivation, or the desire to connect closely with the culture where the target language is spoken; personality, communicative competence, target-language confidence, situational self-confidence, low anxiety, and prior experiences (Clément, Baker, & MacIntyre, 2003; MacIntyre, Clément, Dörnyei & Noels, 1998). Since foreign language learning depends upon such interaction, this variable was seen as a potential predictor of the degree to which language proficiency developed during LITE, and also as a beneficial language intensive experience outcome. Since willingness to communicate has been shown to increase as exposure to a language increases, it is sometimes useful as a proxy in demonstrating the effectiveness of a language program.

The WTC has subscales representing willingness to communicate publicly, in groups, at meetings, with strangers, with acquaintances, and with friends. McCroskey's (1992) Willingness to Communicate (WTC) measure employed the following item stem, which the LITE evaluation also used: "I am willing to . . . ." Of the 20 original WTC items, eight were considered filler items (e.g., "Talk with a garbage collector," "Talk with a police officer"). For the 2011 LITE evaluation, most of the WTC filler items were removed because of their lack of social and cultural authenticity in many cultures in which LITEs were held. For instance, it would be unusual for a LITE participant to have the opportunity to chat with a garbage collector in France or Taiwan, and it might make the garbage collector feel socially uncomfortable as well. In addition, certain of the original WTC items regarding communicating with friends were removed because the nature of the relationship already indicates that individuals would be highly willing to communicate with those persons. The final WTC scale in the 2011 LITE evaluation included nine items and had an internal consistency reliability of .94. The items used the prompt, "I am willing to . . .," as noted earlier. Items referred to talking with friends, strangers, acquaintances, police officers, salespersons, and others; and talking in various size groups.

Language anxiety. Language anxiety has been defined as "the feeling of tension and apprehension specifically associated with second language contexts, including speaking, listening, and learning" (MacIntyre & Gardner, 1994, p. 284). Language anxiety may reflect both stable trait anxiety and situational factors such as the structure of the learning environment and performance demands (Horwitz, Horwitz, & Cope, 1986). High language anxiety is related to negative attitudes towards language learning, confused cognition, lower self-confidence, reduced oral proficiency, and the avoidance of foreign language opportunities. Language anxiety also has physical symptoms, such as blushing or stammering when needing to perform in a target language. Some experts argue mild "facilitative anxiety" aids cognition, but in reality anxiety is a state of fear rarely helpful to anyone. "Facilitative anxiety" might be considered helpful tension (Oxford, 2011). Effective language learners find ways to reduce anxiety through deep breathing, humor, positive self-talk, or making light of a difficult situation, while less effective language learners are sometimes unable to extricate themselves from a tight web of anxiety and negativity (Horwitz, 2007). For the LITE evaluation, several items were adapted from Gardner's (2004) Attitude/Motivation Test Battery for classroom settings, and we added an item of our own. In the LITE 2011 administration of this scale, the internal consistency reliability was .80. The five items in our anxiety scale were as follows:

1. It upsets me when I don't understand what someone is saying in the language.
2. I worry about making mistakes in the language.
3. I get nervous when I am speaking in my target language.
4. I always feel that the others speak my target language better than I do.
5. I never feel quite sure of myself when I am speaking in the language.

The anxiety scale employed a four-point Likert scale for responses. The options included: 1 = Very Untrue of Me, 2 = Untrue of Me, 3 = Sometimes True and Sometimes Untrue of me, 4 = True of Me.

Intercultural Communication Behavior Self-Efficacy Scale. So far we have presented the language proficiency measurement instruments and the instruments used to assess motivation, willingness to communicate, and anxiety. Now we describe two key measurement tools that involve culture. Drawing from the important arena in which affective factors and intercultural behavior overlap, we adapted or created items that resulted in two scales: the Intercultural Communication Behavior Self-Efficacy Scale and the Perspective-Taking Scale.

ILR intercultural communication competence skill-level descriptions were finalized in 2012 (ILR, 2012), but the provisional skill-level descriptions were available before that (ILR, 2009b) and were used in the 2011 LITE evaluation. Based on the provisional descriptions, we developed a new instrument for measuring participants' self-efficacy in performing intercultural communication behaviors. We called this measure the Intercultural Communication Behavior Self-Efficacy Scale, or the ICSE for short. The ICSE employs a five-point Likert type scale (1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Usually, 5 = Almost Always) to indicate how often respondents believe they can perform the intercultural communication behavior described in each item.

Items for the ICSE were written to be aligned with levels 1-4 of the ILR Intercultural Communication Competence levels. [3] Each ILR level describes the intercultural communication behaviors expected of someone with that competence level. Each subsequent level implies that the individual can perform the intercultural communication behaviors of the previous levels. For instance, an individual at level 3 is expected to be able to perform behaviors at both levels 1 and 2. Appendix C describes intercultural communication expectations at levels 1-4 (ILR, 2009b). Note that level 0 is not included. ICSE items include the following:

Level 1

1. I can use appropriate body language (poster, eye contact, and personal space) in the target culture.
2. I can avoid taboos (e.g., certain gestures or topics).
3. I can follow cultural rules for speaking to people of a different age, gender, or status.
4. I find it hard to be understood by a native speaker except when discussing very simple topics. (reversed – later removed)

Level 2

1. I can follow cultural politeness rules (for instance, about invitations, gifts, and asking for help).
2. If I behave incorrectly, I can recognize the problem and fix it.
3. In a conversation, I can refer to cultural topics or historical events.
4. I can address job-related problems in a culturally appropriate way.
5. I have difficulty in cultural situations that are not familiar to me. (reversed – later removed)

Level 3

1. I can adapt my speaking to a range of different cultural settings and still be understood.
2. I can shift appropriately into informal or formal communication styles without thinking.
3. I can handle unfamiliar cultural situations appropriately, even if they involve tense or emotionally-charged topics.
4. I can change my gestures to fit different situations that are not familiar to me.
5. I can discuss cultural issues and subjects, such as history, politics, literature, or the arts.
6. I can repair any cultural errors before they become a problem.

Level 4

1. I can reply appropriately when native speakers use cultural references, shades of meaning, and figures of speech.
2. I can use words and body language to show empathy.
3. I can appropriately explain literary works, classic films, or political documents.
4. I can discuss and explain in depth the culture's traditions, beliefs, and history.
5. I can explain in detail the country's most pressing public issues.
6. I hardly ever make a cultural error anymore.

[3] Because we did not employ dichotomous, yes/no items for the ICSE, it was impossible to conduct a scalogram to validate the way the ICSE operates, level by level. After the fact we tried to split the Likert scale into two parts, thus creating a dichotomous scale, but this did not work effectively because so many of the LITE participants had responded toward the high end of the Likert range for large numbers of items. We then attempted to create a dichotomy by contrasting all responses falling past the first standard deviation above the mean and those falling below that point, there was still a preponderance of responses past the first standard deviation. Therefore, we had to forsake the idea of running a scalogram analysis for this instrument.

Perspective-Taking Scale. Perspective-taking, sometimes referred to as cognitive empathy (Rentsch, Gunderson, Goodwin, & Abbe, 2007), is "the cognitive process of understanding how another person thinks and feels about the situation and why they are behaving as they are" (Sessa, 1996, p. 105). The underlying competencies associated with intercultural perspective-taking are self-awareness, interpersonal skills, and regional expertise. Intercultural perspective-taking is not isomorphic with intercultural communication competence (see above), but both are related to an understanding of the local culture (regional expertise) and to some degree of skill in interpersonal interaction across cultures. The Association of American Colleges and Universities (AACU) considers perspective-taking a key skill to target in educational interventions for intercultural competence development, including it in the AACU's VALUE Rubric for assessing student success (Rhodes, 2010).

For this research, perspective-taking was measured with a scale adapted from Gehlbach's research (2004), who considers perspective-taking to be multi-dimensional, composed of motivational and cognitive elements. Twelve items were adapted from the confidence and propensity subscales of Gelbach's scale. These items specifically dealt with attempting to take another's point of view when trying to understand them, attending to cultural beliefs and values underlying behavior, predicting others' motivations and behaviors, and seeking the cause for another's differing point of view.

For perspective-taking propensity, defined as a disposition toward taking perspectives of others (Gehlbach, 2004), six items were included with in an internal consistency reliability of .85. The propensity subscale used a 5-point Likert-type rating scale measuring the participants' frequency of performing various perspective-taking behaviors (i.e., 1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Usually, 5 = Almost Always). The prompt was "How often do you try to . . .?," and the items were as follows:

- . . . understand people from another culture better by trying to figure out what they are thinking?
- . . . think of more than one reason for a person's behavior?
- . . . "put yourself in someone else's shoes" when you are angry with that person?
- . . . figure out what motivates others to behave as they do?
- . . . figure out what emotions people are feeling when you meet them for the first time?
- . . . understand how other people view a particular situation?

Perspective-taking confidence refers to individuals' sense of efficacy in being able to "read" or understand others' views in day-to-day interactions (Gehlbach, 2004) and was measured with six items (Cronbach's alpha = .90). The confidence subscale used a 5-point Likert-type rating scale which measured participants' degree of confidence in performing certain perspective-taking behaviors (i.e., 1 = Not at all Confident, 2 = Not Very Confident, 3 = Somewhat Confident, 4 = Quite Confident, 5 = Extremely Confident). The item stem was "How confident are you that you can . . .?," and the items were as follows:

- . . . understand the point of view of people from different backgrounds?
- . . . figure out why a person is acting strangely?
- . . . figure out what a person is thinking when you disagree with him or her?
- . . . accurately guess what motivates another person?
- . . . understand what strangers are thinking when you are talking with them?
- . . . figure out what other people are feeling?

Data Collection and Analysis Procedures

Data for this report came from several sources. First, demographic variables, such as gender, commissioning source, DLPT scores (required to be dated within one year of the LEAP Selection Board), and cultural experience indicators were collected from the LITE participants' applications to their LEAP Selection Board. Second, in August 2011, prior to departing for LITEs, participants were emailed links to online surveys, which collected self-report data. Participants completed a somewhat shorter survey again after the LITEs in October of 2011. Post-LITE DLPTs were scheduled by participants, typically within 60 days, after returning from their respective LITEs.

One of the most important program features was placement in CONUS vs. OCONUS LITE locations. However, in the sample of LITE participants who completed the survey, there were only 11 CONUS participants compared with 58 OCONUS participants. We decided that it did not make sense to report any quantitative findings comparing CONUS versus OCONUS this year. Thus, we combined the samples in subsequent analyses.

As mentioned earlier, we also asked two qualitative questions, one before and one after the LITEs (see results in Section 4). For both of these questions, we used standard thematic interpretation techniques to discern the major categories of comments. Then, to obtain the big picture, we roughly quantified the categories of narrative comments and showed frequencies and percentages.

Summary of Section 2

This section has presented the overall evaluation design, the sample, the measurement instruments, and the procedures for data collection and data analysis. Our methodological emphasis has been on both language and culture. The assumption is that the LITEs are expected to produce growth in language proficiency and cultural competence.

Section 3. Quantitative Results

This section presents the results according to the order of the evaluation questions and offers tables and graphs to help readers understand the results.

Evaluation Question 1: What are the characteristics of the LITE participants?

We examined the following characteristics of LITE participants based on self-report: highest education levels, number of languages with prior language competence at survival level (0+) or higher, prior international experience, and prior cultural education or training.

Highest Education Level

All LITE survey respondents possessed at least a bachelor's degree. Of the 63 respondents to the pre-survey, 38 (60%) listed their highest degree as a bachelor's, 23 (37%) listed a master's degree, and two (3%) listed a Ph.D.

Number of Languages at Level 0+ or Higher

In response to the question asking how many languages (other than English) they spoke with at least an 0+ level proficiency (survival proficiency), 49% (n = 31) of the 63 respondents checked one additional language, 35% (n = 22) checked two additional languages, and 13% (n = 8) checked three or more additional languages. In summary, 97% believed they had at least a 0+ in one or more languages other than English (see Figure 2).[4]

[4]Although two participants (3%) selected that they did not have at least a 0+ level proficiency, we believe that they likely misunderstood the question as all Airmen selected for LEAP possess this minimum proficiency level.

Figure 2. Languages Other than English Spoken at Survival Proficiency or Better

Prior International Experience

Many of the LITE participants had international experience, as revealed in Figure 3. Each participant was asked how much cumulative time he or she had spent living or working outside of the home country. Of the 63 responses, the greatest percentage (34.9%, *n* = 22) checked the response more than 3 years, while about one quarter each checked more than 1 year but less than 3 years (27%, *n* = 17) and 0 to 6 months (25.4%, *n* = 16). About one out of 10 (9.5%, *n* = 6) checked between 6 months and 1 year, while 3.2% (*n* = 2) checked the response indicating that they had no experience living or working outside of their home country.

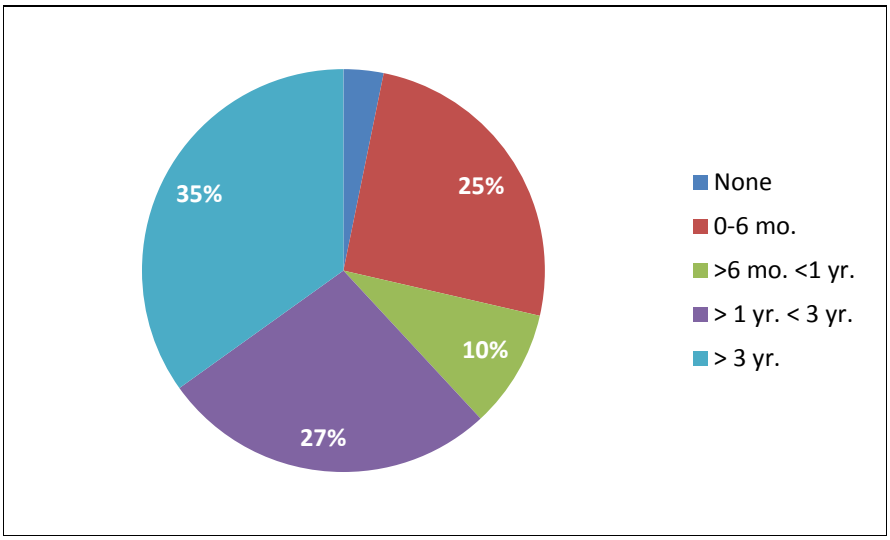

Figure 3. Prior International Living or Working Experience

Prior Cultural Education or Training

In the pre-LITE survey, participants were asked to indicate the types of cultural training or education they had prior to participating in a LITE. Participants were allowed to choose more than one type of cultural training and experiences if appropriate. They reported a fairly high level of cultural and language experience. Of the 70 respondents, only 8.6% (*n* = 6) checked that they had no previous cultural training or education, the rest (91.4%) had such experience. Most had experienced (a) culture training as part of language training (65.7%, *n* = 46), (b) other cultural courses that were not language related (30.0%, *n* = 21), or (c) other forms of education and training in culture (67.1%, *n* = 47). In this last category, participants referred to a variety of other forms of education and training and life experiences such as professional work in other countries, study abroad programs, living in other countries as part of an USAF family, mission trips, and language immersion experiences.

Evaluation Question 2: How effective were the LITEs in developing language proficiency in listening, reading, and speaking?

Before moving into the findings for each of the skill areas, it is helpful to note that these skill areas are significantly intercorrelated. Pre-LITE DLPT listening and reading were moderately correlated with each other (Spearman's rho, ρ =.67, $p < .01$, n=59). After the LITEs, the correlation between listening and reading on the DLPT had decreased somewhat (ρ=.51, $p < .01$, n=48). Post-LITE DLPT listening and reading were significantly correlated with the Speaking Can-Do scale (ρ =.38, $p < .05$, $n = 38$; ρ =.38, $p < .05$, $n = 38$, respectively).

For each of these skill areas, listening, reading, and speaking, we present the pre-test findings, the post-test findings, and the change results. It should be remembered that the DLPTs, which are the measures for listening and reading, usually do not measure above level 3 (General Professional Proficiency), so the DLPT is an inaccurate measure of language proficiency for many individuals who, if tested by a different measurement tool, would actually exceed that level of proficiency in listening and reading.

Listening Proficiency Results

The DLPT scores for the first language listed by the participants were used for selection into LEAP and were also employed as pre-LITE DLPT scores for listening. Table 4 displays the frequencies and percentages found at each pre-LITE DLPT listening level. All DLPT proficiency levels (i.e., from 0 to 3) were represented in the pre-LITE listening results, but the majority were at level 2 and above.[5]

Table 4. Frequencies and Percentages for Each Pre-LITE DLPT Listening Level

ILR Level in Listening on DLPT	Frequency	Percentage
0+	2	3.4
1	2	3.4
1+	5	8.5
2	16	27.1
2+	13	22.0
3	21	35.6
3+	0	0.0
Total	59	100.0

[5] Non-respondents to the LITE survey had pre-LITE DLPT listening and reading scores similar to the ones shown for respondents.

Table 5 reveals the frequencies and percentages for each post-LITE DLPT listening level for those participants who reported a pre-LITE DLPT listening score. For most of the participants, post-LITE listening was at level 2 and above, with one of the participants reading level 4. As noted earlier, since most of the DLPTs do not go as high as level 4, we cannot say for sure whether other participants would have reached level 4 if the tests had allowed measurement as high as that level for their respective language.

Table 5. Frequencies and Percentages for Each Post-LITE DLPT Listening Level

ILR Level in Listening on DLPT	Frequency	Percentage
0	0	0.0
0+	0	0.0
1	1	2.4
1+	3	7.0
2	11	26.2
2+	13	31.0
3	13	31.0
4	1	2.4
Total	42	100.0

Table 6 shows that the change in listening proficiency from pre-LITE to post-LITE was significant. The pre-LITE DLPT mean was 2.34 (*SD* = .66), compared to the post-LITE mean of 2.45(*SD* = .57). Both of these are in the range of Limited Working Proficiency, but one is significantly greater than the other. Practically speaking, effect size (Cohen's *d* = .24) is small. Cohen's *d* was calculated using the sample means (i.e., not the paired t-test means).

Table 6. Pre-Post Change in Listening Proficiency

	Mean (SD)					
	Pre	*Post*	*df*	*t*	*p*	*d*
DLPT Listening	2.20(.61)	2.45 (.57)	41	3.43	0.001	0.24

Reading Proficiency Results

The DLPT scores used to select participants into LEAP were used as pre-LITE DLPT scores for reading. Table 7 presents frequencies and percentages for the pre-LITE DLPT reading levels. The extremes of pre-LITE reading were 0 and 3+ (1 individual each). The majority of participants showed pre-LITE reading at levels 2+ and 3.

Table 7. Frequencies and Percentages for Each Pre-LITE DLPT Reading Level

ILR Level in Reading on DLPT	Frequency	Percent
0	1	1.7
0+	0	0.0
1	4	6.8
1+	8	13.6
2	7	11.8
2+	16	27.1
3	23	39.0
Total	59	100.0

Post-LITE DLPT reading levels are captured in Table 8. Note that only 42 participants had post-LITE DLPT reading level data. Of these, the majority were at levels 2+ and 3.

Table 8. Frequencies and Percentages for Each Post-LITE DLPT Reading Level

ILR Level in Reading on DLPT	Frequency	Percent
0	0	0.0
0+	0	0.0
1	0	0.0
1+	4	9.5
2	8	19.0
2+	13	31.0
3	17	40.5
3+	0	0.0
Total	42	100.0

Table 9 shows that the change in reading proficiency from pre-LITE to post-LITE was significant. The pre-LITE DLPT mean was 2.36 (*SD* = .71; Limited Working Proficiency), while the post-LITE mean was 2.51 (*SD* = .50; or slightly over the imaginary half-way point toward General Professional Proficiency). Cohen's d, the effect size, was 0.24, or small.

Table 9. Pre-Post Changes in Reading Proficiency

	Mean(SD)					
	Pre	*Post*	*df*	*t*	*p*	*d*
DLPT Reading	2.29 (.65)	2.51(.50)	41	3.03	0.004	0.24

Summary of Pre-Post Results for Listening and Reading

The changes were nearly parallel for listening and reading (Figure 4). However, LITE participants started with a slightly higher reading than listening score at pre-test and ended with a slightly higher reading than listening score at post-test.

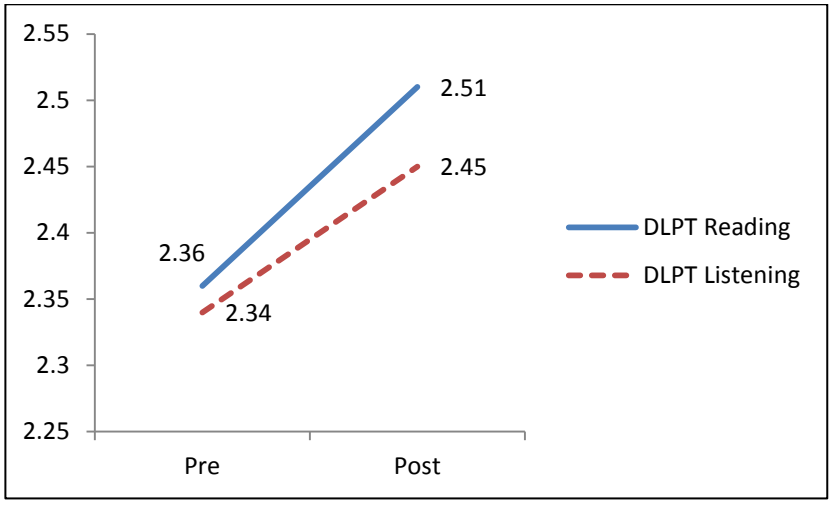

Figure 4. Pre-Post Changes in DLPT Listening and Reading

Since the DLPT does not report raw scores to the participants or to LEAP, it is possible for an individual's DLPT level to remain the same from pre- to post-testing even though the individual experienced some gain in reading or listening proficiency. Table 10 shows the percentage of participants who declined to a lower ILR level, advanced to a higher ILR level, or stayed within the same ILR level from pre- to post-LITE. A cumulative 50% of participants increased in DLPT-measured listening proficiency, greater than the percentage that stayed the same (38.1%) or declined (11.9%). A cumulative 40.2% increased in DLPT-measured reading, which was less than the percentage that experienced no change (52.4%); however, greater than the percentage of those who declined (9.5%).

Table 10. Frequency of Change by DLPT Listening and Reading Proficiency

	DLPT Listening		DLPT Reading	
	Frequency	Percent	Frequency	Percent
Up by 2 Levels	-	-	1	4.5
Up by 1^{+} Levels	-	-	-	-
Up by 1 Level	6	14.3	4	9.5
Up by +	15	35.7	11	26.2
Stayed the Same	16	38.1	22	52.4
Down by +	4	9.5	4	9.5
Down by 1 Level	1	2.4	-	-
Down by 1^{+} Levels	-	-	-	-
Total	42	100	42	100

Note: Wilcoxon Signed Ranks Tests were conducted. The results were as follows: DLPT Listening ($Z = -3.01$, $p = .01$, with 16 ties); DLPT Reading ($Z = -2.87$, $p = .01$, with 22 ties).

Additional analyses were also conducted to determine whether there were differences in the amount of change in language proficiency from pre- to post-LITE depending upon the incoming proficiency level of the participant. Figures 5 and 6 below reflect the mean amount of *change* in DLPT levels experienced by individuals at different incoming listening and reading levels (e.g., their baseline, pre-LITE DLPT scores). The degree of change ranged from -1.00 to 1.00 for the listening DLPT scores and -.50 to 2.00 for the reading DLPT scores.

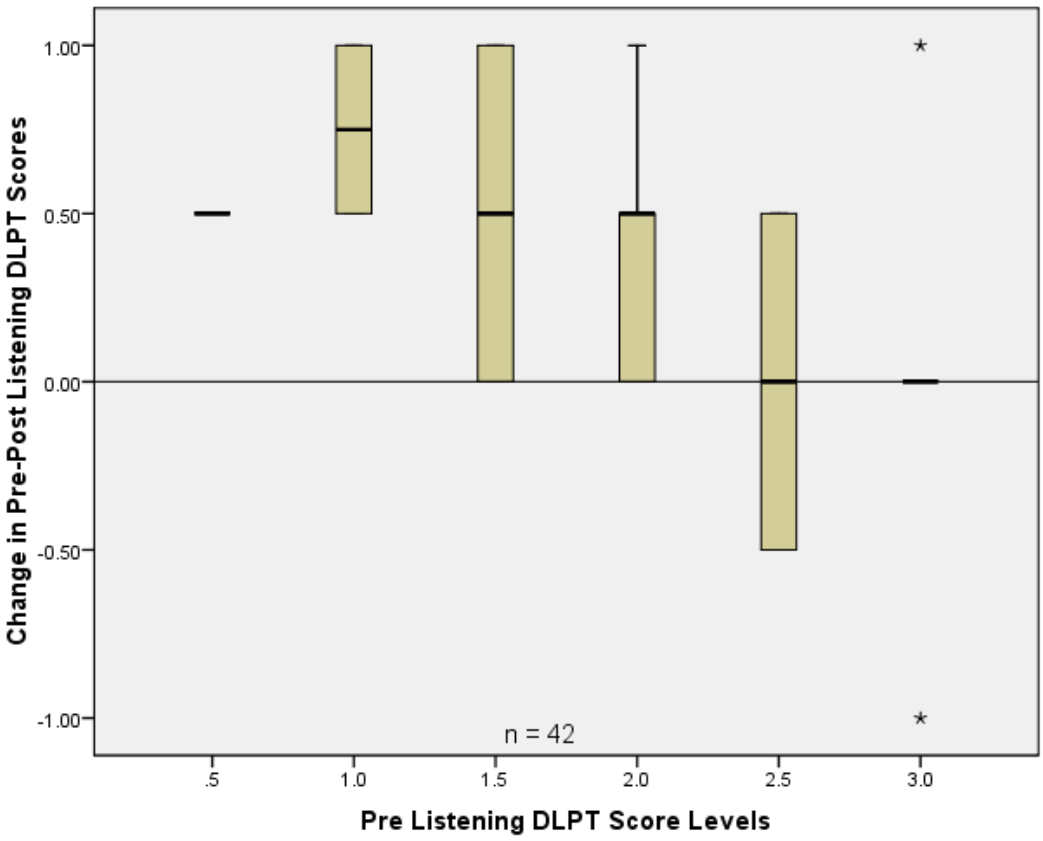

Figure 5. Mean Amount of Change in Listening Levels Experienced by Individuals at Different Incoming Listening Levels

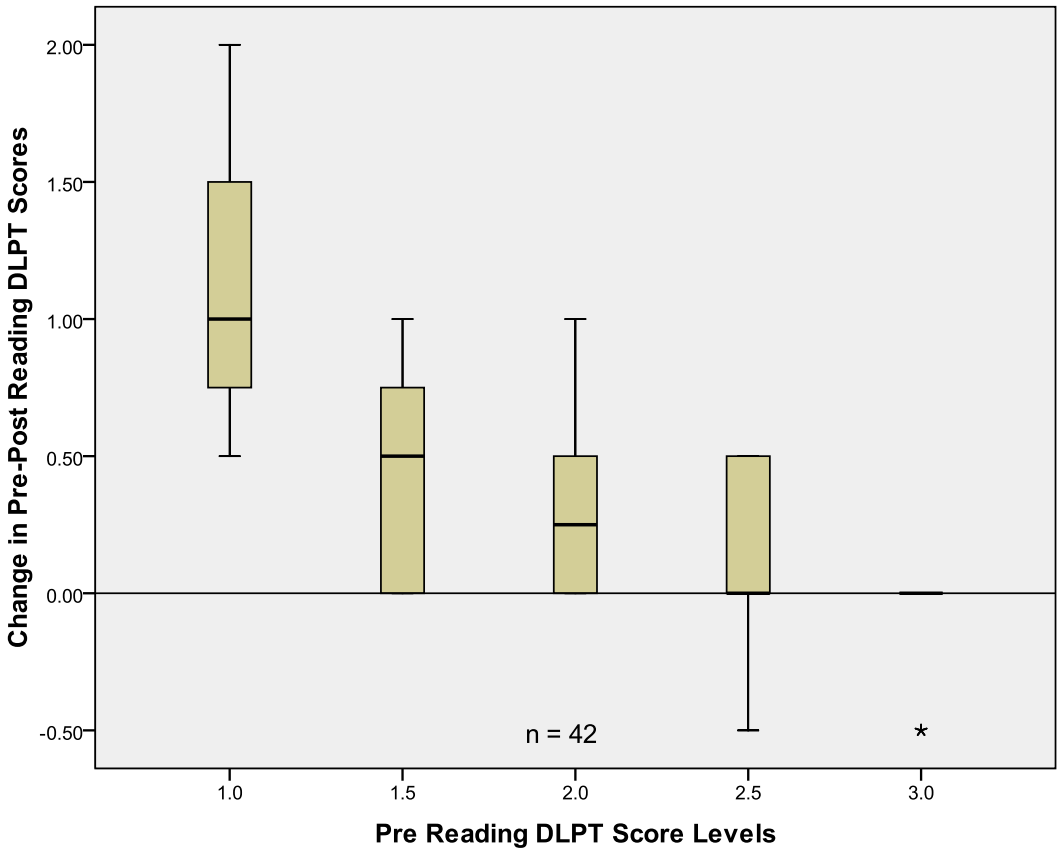

Figure 6. Mean Amount of Change in Reading Levels Experienced by Individuals at Different Incoming Listening Levels

When averaging the degree of change across both listening and reading scores, the greatest increase in DLPT levels occurred for those who had a pre-LITE DLPT of level 1 in listening or reading. These individuals increased an average of .75 if they had a pre DLPT score of a 1 in listening and an average of .67 if they had a pre-LITE DLPT score of 1 in reading. Improvement of half of a level was shown for those whose pre-LITE DLPT in listening was 0+. All other pre-LITE DLPT levels in which there were data increased slightly less than half of a level or remained unchanged. Finally, the results for individuals with a pre-LITE DLPT level of 3 for both listening and reading showed very little change or a decrease in proficiency. Figure 7 graphically shows that it is more difficult to increase proficiency at higher levels than at level 1.

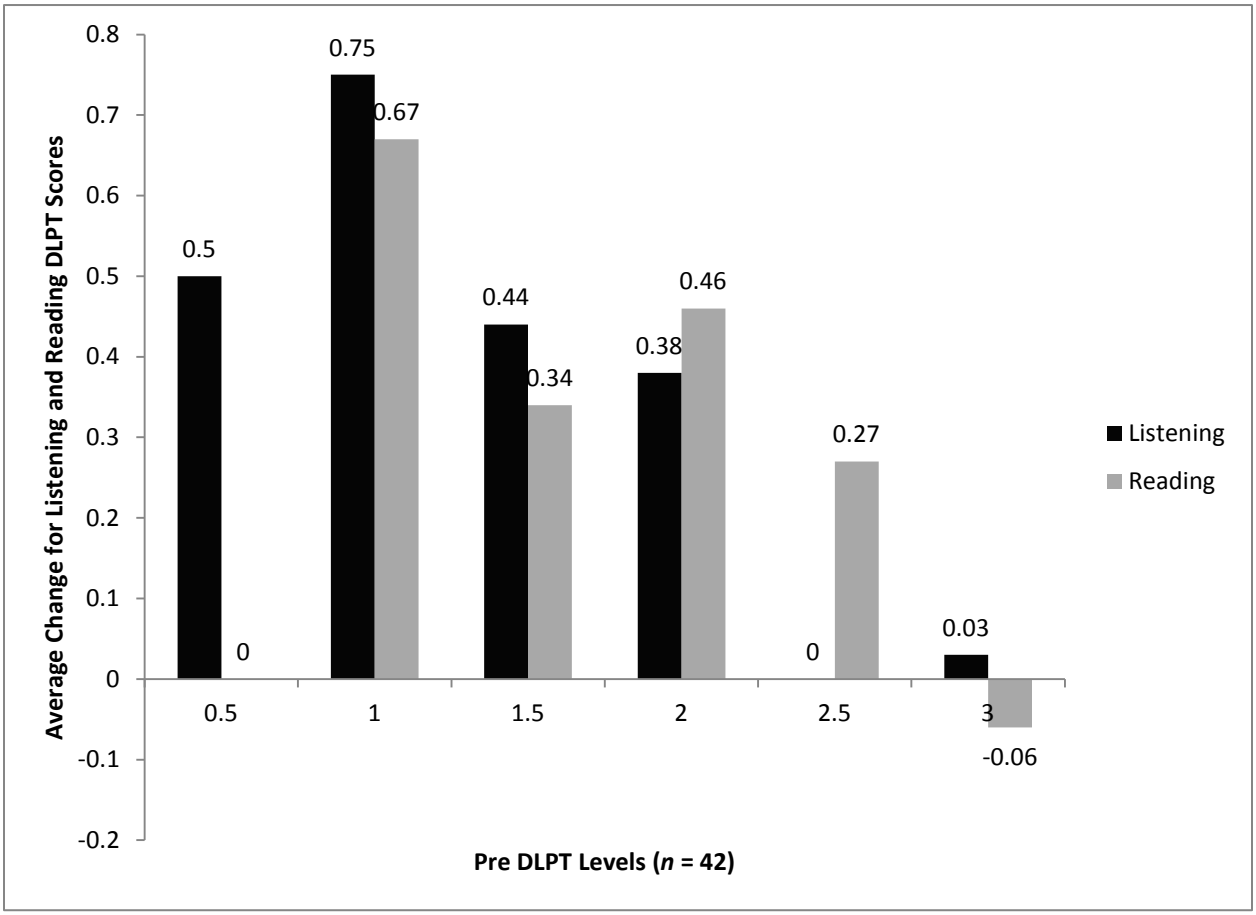

Figure 7. Mean Amount of Change in Listening and Reading Levels Experienced by Individuals at Different Incoming Listening and Reading Levels

Speaking Proficiency Results

In Section 2 of this report we explained the scalogram analysis, which helped to validate the Speaking Can-Do Scale. Here we present the results of the Speaking Can-Do Scale. Table 11 presents pre-post changes in speaking proficiency based on the Speaking Can-Do scale. A total of 40 people had usable pre-post Speaking Can-Do data, and the first analysis was based on summing the number of items selected by an individual across all of the levels. Each LITE participant then had a total score for speaking based on the number of speaking items selected overall. This was called the Speaking Can-Do sum. The pre-LITE Speaking Can-Do sum of 20.73 (*SD* = 6.92) was compared with the post-LITE Speaking Can-Do sum of 23.24 (*SD*=6.34), resulting in a significant difference ($t = 4.55(df) = 39$, $p < .01$). The effect size (Cohen's *d*) was .38, indicating a small practical effect.

Table 11. Pre-Post Change in Speaking Proficiency

	Mean(SD)					
	Pre	*Post*	*df*	*t*	*p*	*D*
Speaking Can-Do	20.33(7.17)	24.20(5.84)	39	4.55	0.000	0.38

Next, the percentages of items selected within each level of the Speaking Can-do scale (Levels 1-4) were compared pre and post LITE. The results indicated a significant difference. The pre-LITE speaking mean for the percent selected within each level was 1.67 (*SD*=1.044), while the post-LITE percent selected with each level was 2.26 (*SD*=1.063), with $t=4.52(df) = 37$, $p < .01$. This means that participants rated themselves as significantly increasing in speaking proficiency level during the course of the LITE. The effect size (Cohen's *d*) was .56, indicating a moderate practical effect.

Evaluation Question 3: How effective were the LITEs in increasing intercultural communication competence, willingness to communicate, or perspective-taking?

ICSE Results

Pre-post measures of the ICSE scale were created by averaging the items. The comparison of the pre-LITE ICSE measure (M = 3.66, SD = .60) with the post-LITE measure (M = 3.85, SD = .61) was significant (t = 2.72(df) = 40, p < .05, Table 12). The effect size (Cohen's d) was .31, indicating a small practical effect.

Table 12. Pre-Post Change in Intercultural Communication Competence

	Mean(*SD*)					
	Pre	*Post*	*Df*	*t*	*p*	*D*
ICSE	3.61(.60)	3.84(.63)	40	2.72	.010	.31

Willingness to Communicate Results

Pre-LITE and post-LITE results indicated that LITE participants were usually willing to communicate in the target language. The pre-LITE mean was 4.10 (SD = 0.86), meaning usually willing. The post-LITE mean of 4.30 (SD=0.71) was indicated that the participants were slightly more emphatically willing after the LITE was over. The pre-LITE to post-LITE comparison was significant (t = 2.21(df) = 40, p < .05; see Table 13). Thus, participants were significantly more eager to communicate in the language after the LITE than before, although they were usually willing to communicate even at the start of the LITEs. The effect size of this difference was (Cohen's d) was .35, indicating a small practical effect.

Table 13. Pre-Post Change in Willingness to Communicate

	Mean					
	Pre	*Post*	*df*	*t*	*p*	*D*
Willingness to Communicate	4.10	4.30	40	2.21	0.03	0.35

Willingness to communicate seems to be a very important variable, partly because of its linkage to other variables of interest. For example, post-LITE willingness to communicate (n = 48) was correlated with the post-LITE Speaking Can-Do sum (r = .69, n = 48) and with post-LITE perspective-taking confidence (r = .58, n = 48), with both relationships significant at p < .01.

To determine whether willingness to communicate predicted post-test language proficiency (post-LITE), the pre-LITE measure was controlled for in a hierarchical regression analyses (pre willingness to communicate was entered into the first step of the regression equation). The results indicated that the post willingness to communicate score predicted the post Speaking Can Do measure and the ICSE, controlling for the pre-LITE measure ($\beta = 4.28$, $t(38) = 2.72$, $p < .05$; $R^2 = .47$, $F(2,38) = 17.09$, $p < .01$; $\beta = .66$, $t(38) = 3.39$, $p < .01$; $R^2 = .35$, $F(2,38) = 10.01$, $p < .01$, respectively). Thus, participants who were more willing to communicate during their LITE experience achieved higher speaking and intercultural communication competence.

Perspective-taking Results

No significant differences between pre-test and post-test arose for the two perspective-taking variables, propensity to take others' perspectives or confidence in doing so. Participants started their LITEs with a mean of 4.02 ("usually") regarding propensity to take others' perspectives, and their confidence level of 3.67 was toward the upper range of "somewhat confident" (see Table 14). Starting at such high levels, it is understandable that the participants did not increase significantly in perspective-taking propensity or confidence.

Table 14. Pre-Post Change in Perspective-Taking

	Mean(SD)					
	Pre	*Post*	*df*	*t*	*p*	*D*
Perspective Taking-Propensity	4.00(.62)	4.05(.61)	41	0.83	.41	0.10
Perspective Taking-Confidence	3.62(.75)	3.78(.76)	40	1.68	.10	0.15

Perspective-taking confidence and perspective-taking propensity are clearly not the same, although they are significantly related. The correlations between the two variables is moderate (Pre-Lite: $r = .65$, $p < .01$, $n = 63$; Post-Lite: $r = .47$, $p < .01$, $n = 48$).

A complex relationship exists between perspective-taking confidence and language proficiency. Before the LITEs, perspective-taking confidence was significantly related to DLPT listening proficiency ($\rho = .37$, $p < .01$, $n = 52$) and the Speaking Can-Do sum ($r = .29$, $p < .05$, $n = 61$) but only marginally related to DLPT reading proficiency ($\rho = .27$, $p < .10$, $n = 52$). However, after the LITEs, perspective-taking confidence was significantly related to the Speaking Can-Do sum ($r = .68$, $p < .01$, $n = 48$) and listening proficiency ($\rho = .37$, $p < .05$, $n = 42$), and unrelated to reading proficiency ($\rho = .22$, $p = .16$, $n = 42$).

To determine whether perspective-taking confidence predicted post-test language proficiency (post-LITE), the pre-LITE measure was controlled for in a hierarchical regression analysis (pre perspective-taking confidence was entered into the first step of the regression equation). The results indicated that the post perspective-taking confidence score predicted the post Speaking Can Do measure and the post ICSE score, controlling for the pre-LITE measure ($\beta = 4.83$, $t(38) = 3.54$, $p < .05$; $R^2 = .48$, $F(2,38) = 17.54$, $p < .01$; $\beta = .48$, $t(38) = 3.56$, $p < .01$; $R^2 = .45$, $F(2,38) = 15.41$, $p < .01$, respectively). Thus, participants who had greater perspective-

taking confidence during their LITE experience achieved higher speaking and intercultural communication competence.

Interestingly, perspective-taking propensity had no significant relationship with language proficiency (reading, listening, or speaking). Thus, the proclivity to take other perspectives is unrelated to the ability to communicate in the target language. On the other hand, controlling for pre-LITE perspective-taking propensity, the post-LITE score predicted intercultural communication competence ($\beta = .60$, $t(38) = 2.70$, $p < .05$; $R^2 = .21$, $F(2,38) = 4.91$, $p < .05$).

Evaluation Question 4: Did the LITEs have any influence on language motivation or language anxiety, and did motivation and anxiety relate to positive outcomes?

We checked to see whether the LITEs affected language motivation or language anxiety. In both cases, the results indicated that the pre-LITE and post-LITE scores were not significantly different for either language motivation or language anxiety.

Language Motivation and Anxiety Results

Language motivation, both intrinsic and extrinsic, was rather strong before and after the LITE for the 42 who responded to the motivation items at both times. No significant differences emerged from pre-test to post-test. This indicates that motivation was quite constant and the LITEs did not significantly increase or decrease motivation (see Table 15).

The mean language anxiety score at the beginning of the LITEs was 2.83 (*SD*=0.79), indicating that the anxious behavior described was sometimes true of the participant but sometimes not true. The post-LITE mean for language anxiety was 2.99 (*SD* = 0.74), again indicating that anxious behavior was sometimes true but sometimes not. No significant pre- to post-LITE difference emerged for language anxiety. Participants were neither particularly anxious nor particularly calm at either the pre-LITE or post-LITE time (see Table 15).

Table 15. Pre-Post Change in Language Motivation and Anxiety

	Mean(SD)					
	Pre	*Post*	*df*	*T*	*p*	*D*
Language Intrinsic Motivation	4.25(.49)	4.27(.49)	41	0.27	.79	.02
Language Extrinsic Motivation	4.09(.40)	4.09(.52)	41	0.09	.93	.02
Language Anxiety	2.83(.76)	2.92(.69)	41	1.04	.31	.21

Language anxiety was significantly related to language proficiency. Post-LITE language anxiety was negatively related to post-LITE speaking ($r = -.51$, $p < .01$, $n = 49$) and to post-LITE listening proficiency ($r = -.33$, $p < .05$, $n = 38$) in the expected directions.

Furthermore, pre-LITE language anxiety ratings were negatively correlated with most pre-LITE attitude measures [willingness to communicate (r = -.26, p < .05, n = 63); confidence in perspective taking (r = -.39, p < .01, n = 63); ICSE (r = -.48, p < .01, n = 63)]; post-LITE confidence in perspective taking (r = -.36, p < .05, n = 41); and ICSE (r = -.33, p < .05, n = 41)].

Similarly, post-LITE language anxiety ratings were negatively associated with most post-LITE attitude measures: willingness to communicate (r = -.40, p < .01, n = 48); confidence in perspective taking (r = -.52, p < .01, n = 48); and ICSE (r = -.52, p < .01, n = 48).

To determine whether language anxiety and language motivation predicted post-test language proficiency (post-LITE), the pre-LITE measures were controlled for in separate hierarchical regression analyses. For motivation, a combined measure (extrinsic and intrinsic combined) was used. The results indicated that the post motivation score predicted post ICSE (β = .82, t(38) = 3.35, p < .01; R^2 Δ = .28, F(2,38) = 7.39, p < .01), controlling for prior motivation, while post language anxiety predicted post-LITE speaking and post ICSE (β = -6.16, t(38) = -3.63, p < .01; R^2 = .31, F(2,38) = 8.93, p <.01; β = -.55, t(38) = -3.26, p < .01; R^2 = .31, F(2,38) = 8.34, p <.01, respectively). Thus, participants who had a higher level of motivation and less language anxiety during their LITE experience achieved a higher level of intercultural communication competence than those reporting lower levels of motivation and higher levels of anxiety. Further, participants who had less language anxiety achieved higher levels of speaking proficiency than those with higher levels of anxiety.

Evaluation Question 5: How did the participants evaluate the program features?

After the LITEs, when participants were asked to rate various aspects of their LITE classroom experiences on a four-point scale ranging from 1 (strongly disagree) to 4 (strongly agree), 33 responded. No statements about classroom variables garnered consistently strong agreement, but within the "agree" range," the highest ratings emerged for the statements found in Table 16.

Table 16. Highest Rated Program Evaluation Items

Program Evaluation Items	Mean	Standard Deviation
The instructor utilized a variety of activities and teaching methods to deliver material.	3.55	0.56
Cultural topics were integrated into the lessons in a way that enhanced my cultural awareness.	3.52	0.57
Course materials were helpful and contributed to my language learning.	3.39	0.56
The instructor appropriately adapted the materials and activities to meet the range of skill levels in the classroom.	3.39	0.61
I received frequent and appropriate feedback from the instructor on my language skills.	3.24	0.71
The course instruction was organized in a logical and coherent manner.	3.12	0.77
I was held accountable for speaking the target language at all times during my immersive experience.	3.09	0.98

The results above indicated that the participants agreed that their instructors used a range of activities and teaching methods and integrated culture. They agreed, but somewhat less strongly, that the course materials facilitated learning, that the instructor adapted the instruction to existing skill levels, that feedback was frequent and appropriate, that instruction was logically and coherently organized, and that participants were held accountable for speaking the target language at all times.

On the other hand, they disagreed with two statements. For the statement "Technology was appropriately and effectively integrated into language instruction," responses were slightly on the side of disagreement (M =2.88, SD = 0.86). Respondents disagreed with the statement "The pace of the learning was too fast given the skill level of the audience" (M = 1.70, SD = 0.69).

Although the respondents were rather satisfied with the nature of their classroom experiences, the results indicated a need to improve technology use and a need to make participants accountable for speaking the target language. The latter classroom factor had the highest variability of all program variables.

Many classroom variables were interrelated for the 33 who responded to these items. Because of the dearth of other forms of classroom-related data (due to lack of on-site observations or examination of course materials for this evaluation), we can rely only on participants' responses about classroom features. Therefore, to obtain a better picture of the overall classroom situation, we provide all the significant correlations, both large and small, among classroom variables.

For instance, perceptions about logical and coherent course organization were moderately correlated with perceptions about helpful course materials (r = .68, p < .01) and frequent and appropriate feedback (r = .57, p < .01). Perceptions about such course materials were correlated with perceptions about appropriate use of technology (r = .49, p < .01) and the integration of cultural topics (r = .35, p < .05).

Participants' perceptions about frequent and appropriate feedback were moderately related to their perceptions of logical course organization (r = .57, p < .01) and adaptation of materials and methods to meet student skill levels (r = .57, p < .01), while perceptions of such feedback were related to perceptions about appropriate technology use (r = .41, p < .05), helpful course materials (r = .39, p < .05), and integration of cultural topics (r = .38, p < .05).

Perceptions about accountability for speaking the target language were moderately correlated with perceptions about the use of a variety of activities (r = .53, p < .01). Speaking-accountability perceptions were related to perceptions about adaptation of materials and methods to meet skill-level needs (r = .41, p < .05). Perceptions about adaptation of materials were associated with use of a variety of activities (r = .45, p < .01), while perceptions about helpful course materials were linked to perceptions about integration of cultural topics (r = .43, p < .05) and use of appropriate technology (r = .37, p < .05).

Unexpectedly, the perception that technology integration was appropriate was correlated with a perception that the pace of learning was too fast ($r = .36$, $p < .05$); perhaps appropriate technology was used to balance a time-pressured learning situation, although we cannot be sure of this interpretation.

Evaluation Question 6: Did any program features significantly relate to the outcomes?

Naturally we wanted to find out whether language proficiency outcomes related to program characteristics. Because there were so few who completed the items about program factors and also had DLPT scores ($n = 22$), we can only provide limited information. Since the number is so small, we are using a *p*-value of .10 to indicate significance for this particular question.

The DLPT reading proficiency change score was marginally related to the perception of being held accountable for speaking the target language ($\rho = .39$, $p < .10$, $n=22$) and somewhat more related to the integration of technology in instruction ($\rho = .49$, $p < .05$, $n = 22$). The DLPT listening proficiency change score was marginally related to the perception that the instructor utilized a variety of activities and teaching methods to deliver the material ($\rho = .39$, $p < .10$, $n = 22$). Speaking competence was not related to any of the program factors.

Summary of Section 3

This section has shown significant pre-LITE to post-LITE changes in listening, reading, and speaking proficiency. In addition, changes occurred in intercultural communication behavior self-efficacy and willingness to communicate. No changes occurred in the strong motivation or the moderate anxiety experienced by LITE participants.

Section 4. Qualitative Results

We have presented quantitative results, and now we turn to qualitative results. We focus here on two questions asking for narrative responses. One question was answered before the LITEs and one answered afterwards.

Pre-LITE Qualitative Question: What did the participants expect from the LITEs?

The pre-LITE survey asked Airmen to explain in their own words their expectations, hopes, and aims regarding the LITEs. All of the 63 Airmen who responded wrote more than one expectation, hope, or aim. The total number of expectations, hopes, or aims was 234. Three categories of expectations emerged. Almost 8 out of every 10 participants (77.4%) cited expectations related to language proficiency, while 13.7% mentioned expectations about cross-cultural competence, and 9.0% referred to other expectations, including contributing to the USAF.

Expectations Related to Language Proficiency

This theme garnered 181 responses from Airmen participating in the LITEs. The greatest overall desire was to improve speaking proficiency (30% of all responses in this theme) followed by improvement of listening proficiency (21%) and cross-cutting language abilities (e.g., grammar, vocabulary, and formality/informality, 20%), then by enhancement of literacy (14%), renewal of prior language knowledge or starting a new language (11%), complete immersion in the target language (2%), and maintenance of language skill after the LITEs (1%; see Table 17 for common responses).

Expectations Related to Cross-Cultural Competence

This theme is comprised of 32 responses from Airmen participating in the LITEs. Participants mentioned wanting to use the LITEs to improve their understanding of culture, history, economics, customs, traditions, and sociopolitical issues. One participant stated the desire to dispel the “ugly American” image while one participant mentioned a *lack* of an expectation for learning culture (see Table 18).

Table 17. Expectations Related to Language Proficiency

Expectation for Improvements	# of Specific Responses	% of All Responses for Theme	Common Responses
Speaking proficiency	55	30%	• Improving speaking skills, pronunciation or accent for spoken communication • Desire to become more confident, relaxed, or proactive in speaking the language • Desire to interact in the target language with native speakers during the LITEs both socially and professionally
Listening proficiency	38	21%	• Enhancing listening skills • Improving listening DLPT scores • Interacting in the target language with native speakers
Cross-cutting language abilities (grammar, vocabulary, communication)	37	20%	• Improving professional, military, and technical vocabulary • Learning how to communicate formally or informally
Literacy (reading/writing proficiency)	25	14%	• Improving reading • Improving reading on the DLPT • Enhancing writing in the target language.
Renewal of prior language knowledge or encountering an entirely new language	20	11%	• Intensively reviewing or relearning language information they previously knew, thus reviving their language proficiency. • Learning an entirely new language • Comments were from Airmen who had been selected to be “cross-trained” in a new language. Normally such Airmen already had strong proficiency in at least one foreign language already.
Being completely immersed in the language	4	2%	• Become completely immersed in the specific culture
Maintenance of post-LITE language skill	2	1%	• Learning ways to maintain language skills after the LITEs
Total	181	100.0%	

Table 18. Expectations Related to Cross-Cultural Competence

Expectation for Improvements	# of Specific Responses	% of All Responses for Theme	Common Responses
Greater understanding of the culture in general, history, economics, customs, traditions, and sociopolitical issues	30	93.8%	• Achieve greater understanding and appreciation for a culture about which I have very little background understanding • Be able to discuss their history, the defining values of their culture, and their aspirations for the future • Understanding language by understanding culture
Do not expect to learn anything about culture	1	3.3%	• Do not expect to learn about culture while studying the target language in the U.S
Dispel the idea of the "ugly American" (culturally ignorant)	1	3.3%	• Create mutual understanding across cultures
Total	32	100.0%	

Other Expectations

Airmen provided 21 other expectations not cited elsewhere such as contributing to the USAF (42.9%), encountering contrasts in language or culture and facing a positive challenge (19.1% for each of these two expectations), and interacting with LITE colleagues and gaining knowledge in new areas (9.5% for each of these; see Table 19).

Table 19. Other Expectations of the LITEs

Expectation for Improvements	# of Specific Responses	% of All Responses for Theme	Common Responses
Contributions to the Air Force or other services	11	52.4%	• Improving their USAF careers • Finding opportunities for using the language to help the USAF • Supporting USAF duties or doctrine • Understanding the USAF or Defense Department mission in the countries where the target language is used
Contrasts among cultures or languages	4	19.1%	• Seeing cultural or linguistic contrasts
Positive challenge	4	19.1%	• Looking forward to the challenge of the LITEs • Challenging listening and speaking abilities • Gaining experience and cultural understanding
Interactions with LITE colleagues	2	9.5%	• Interacting with their own colleagues within the LITE cohorts
Total	21	100.0%	

Post-LITE Qualitative Question: What types of cultural or communication challenges did the participants encounter during their LITEs?

Thirty seven participants responded to this question for a total of 47 comments. A large percentage concerned cultural challenges (45%), followed by communication challenges (38%), other challenges (9%), and no major challenges (9%).

Cultural Challenges

As shown in Table 20, LITE participants mentioned 21 cultural challenges dealing with restaurants and vendors (38.1%), cultural differences or problems (28.6%), stereotyping and other attitudes (23.8%), and host families (9.5%, though one of the comments was actually positive rather than citing a challenge).

Communication Challenges

Of the 18 responses reflecting communication challenges, most (61%) dealt with general or specific limitations in speaking ability while the rest (39%) concerned cross-training in a different language from the one originally learned.

Military Challenges

Two responses concerned military challenges such as understanding acronyms used in military and government related topics and being in the role of a student vice a military trainer.

Program Administration Issues on Site

Two comments reflected administrative issues at the local site such as a lack of understanding of the role and purpose of the program coordinator and placement of students at the appropriate level of instruction.

No Major Cultural or Communication Challenges

Four comments indicated that there were few cultural or communication challenges and mentioned that the teacher spoke good English or was supportive or that they themselves felt competent.

Table 20. Cultural Challenges Encountered During the LITES

Expectation for Improvements	# of Specific Responses	% of All Responses for Theme	Common Responses
Restaurants and vendors	8	38.1%	• Sociocultural confusions regarding restaurants and vendors • Perceived harassment by or unpleasantness with street vendors • Ability to argue, persuade, and negotiate in the target language
Cultural differences or problems	6	28.6%	• Differences in timeliness and specific cultural norms • Learning about authentic behavior regarding religious events and health care systems • Both challenges and opportunities in becoming friends with the locals, outside of class
Cultural stereotypes and attitudes	5	23.8%	• Challenges posed by stereotypes and attitudes about Westerners or Americans held by ordinary people or those in authority in the local culture
Host family	2	9.5	• Cited restrictions regarding meals posed by host family • Praised the help given by the host family
Total	21	100.0%	

Table 21. Communication Challenges Encountered During the LITEs

Expectation for Improvements	# of Specific Responses	% of All Responses for Theme	Common Responses
General or specific limitations in speaking ability	11	61.1%	• Vocabulary limitations • Accent difficulties • Grammar mistakes while speaking • Speaking level limited ability to communicate with people unused to speaking with foreigners • Ability greatly improved over the four-week period • Challenging to speak with native speakers outside of the classroom environment
Issues about cross-training in a different language	7	38.9%	• In the course of four weeks, can now communicate in social situations with some locals • Difficulty in communicating with locals or during daily activities • With more practice will be able to switch back and forth more easily • Learned new ways of saying things and grasped new ideas about how things were in this new environment • Getting comfortable with the accent and slang took some getting used to
Total	18	100.0%	

Summary of Section 4

Expectations held by LITE participants focused largely on improving language proficiency. Cultural learning and growth was a weak second focus. However, the challenges that the participants most often mentioned after the LITEs dealt with cultural issues, followed by communication. It appears that for some participants, expectations and challenges do not align very well. Perhaps participants do not recognize that communication through language directly and intimately involves culture.

Section 5. Discussion and Recommendations

The findings from the current research show that the 2011 LITEs succeeded in multiple ways in terms of improving language proficiency and cultural competence. In this section, we discuss what those results mean and situate them in the context of existing research in the language learning field.

Seasoned Participants

The characteristics of the 2011 LITE participants showed that these individuals were not only well-educated but also linguistically and culturally experienced. When starting the LITEs, almost all participants rated themselves as having a minimum of one language (other than their native language) in which they could communicate at a survival level or beyond. Nine out of every 10 participants who responded to the survey indicated having experienced some cultural education or training, often as part of language instruction but sometimes in other forms. Participants were also strongly motivated, as we discuss next.

High and Unswerving Motivation

As noted in the definitions of types of motivation (see earlier), drawn largely from Ryan and Deci's (2000) Self-Determination Theory, intrinsic motivation is theoretically the key to great autonomy and self-direction. The 2011 LITE participants had intrinsic and extrinsic motivation, though the former was slightly stronger than the latter at pre-LITE and post-LITE times. Thus, the Airmen participating in the LITEs were largely sparked by their own deep interest in the language they were learning (intrinsic motivation), although they also showed an interest in learning the language for external purposes (extrinsic motivation).

Motivation was rather strong both before and after the LITEs. This would be expected, because LITE participants – and LEAP members in general – are usually hard-charging and determined. Their selection into the program requires pre-existing evidence of high, ongoing motivation. However, these findings also support the idea that language motivation is a less malleable, trait-like construct and as such is less likely to be affected by training and education interventions.

Negative Effects of Language Anxiety

The negative effects of anxiety on many of the outcomes assessed in this evaluation were expected and support prior research (e.g., MacIntyre & Gardner, 1994). The results do *not* suggest that anxiety served as a facilitative role, and we recommend that steps be taken where feasible and practical to minimize stressful, environmental factors that may contribute to higher anxiety levels. For example, scaffolds could be used so that the language learners build confidence and proficiency through exercises which provide contextual cues versus those which demand strict recall. Thus, the learning environment can be structured in a way that maximizes learning and at the same time minimizes anxiety.

Reading and Listening Proficiency Outcomes

Listening improvement was a high expectation of LITE participants, and reading improvement had adherents as well. Both of these skills significantly improved pre-LITE to post-LITE. In reading, the pre-LITE DLPT mean was 2.29, and it increased to 2.51 at post-LITE testing. The pre-LITE mean could be seen as Limited Working Proficiency-Plus, while the post-LITE mean, while still at Limited Working Proficiency-Plus, moved toward General Professional Proficiency. This change showed an effect size that was close to medium. As we mentioned in the 2010 LITE evaluation report, foreign and second language experts view reading as an active, nonlinear process in which readers (a) test mental hypotheses and (b) use background and linguistic knowledge to determine meaning (e.g., Oxford, 1990, 2011; Scarcella & Oxford, 1992). Reading comprehension only occurs when the individual actively reflects on what he or she is reading. Research shows that highly proficient second language readers use more – or better targeted – reading strategies, often combining bottom-up (analytic) strategies and top-down (holistic) strategies for reading textbooks and taking tests (e.g., Afflerbach, Pearson, & Paris, 2008; Anderson, 1991; Hosenfeld, 1977; Lee, 2007; Winograd & Hare, 1988). Competent readers also use metacognitive strategies, such as planning, organizing, and evaluating their reading (Mokhtari & Reichard, 2002; Mokhtari & Sheorey, 2002). It would be useful to find out which reading strategies the LITE participants use and to help these participants use the most optimal strategies for the reading tasks they face during the LITEs and thereafter.

In listening, LITE participants started with an average DLPT of 2.20 and ended at 2.44, showing a medium effect size. Thus, reading started and ended as a slightly stronger proficiency than listening for the LITE participants. Although these means were both in the range of Limited Working Proficiency–Plus, we must pay attention to the significance of the change in listening proficiency that occurred during the short period of the LITEs. That such an improvement occurred is a testament to the hard work of the participants. However, it does appear that there is room for improvement in terms of bringing listening up to par with reading. Most language teaching experts agree that one of the roles of language teachers is to help students become more strategic in listening and thus enhance listening proficiency (Mendelsohn & Rubin, 1995). We call for LITE instructors to focus more on the listening skills of participants, so that listening can be as strong as reading. If teachers do not know how to teach listening, they require information and skills in this area. Macaro, Graham, and Vanderplank (2007), Rost (2010), and Oxford (2011) provide excellent suggestions for weaving listening strategy instruction into regular language teaching. Vandergrift, Goh, Mareschal, and Tafaghodatari (2006) provide an exceptional assessment tool for foreign language listening strategies.

A Nexus: Speaking Proficiency, Willingness to Communicate, and Perspective-Taking Confidence

Speaking improvement was the most frequently mentioned aim and expectation of most LITE participants. Significant pre-LITE to post-LITE improvement occurred for self-assessed speaking proficiency. The mean self-reported speaking proficiency level in the pre-LITE survey was 1.58, or Elementary Proficiency-Plus (leaning toward Limited Working Proficiency). In the post-LITE survey, the mean endorsed level was 2.32, or Limited Working Proficiency-Plus. The

Speaking Can-Do sum changed from an average of 20.33 at pre-LITE to 24.20 at post-LITE. All of these results point toward an important gain in speaking proficiency.

Nevertheless, speaking is often assumed to be more difficult than reading and listening, and the 2011 LITE results supported this assumption. It is particularly necessary to teach speaking to LITE participants. To help participants improve their speaking proficiency, LITE teachers must teach vocabulary-learning strategies (Gu & Johnson, 1996; Nation, 1990; Nation & Waring, 1997; Nyikos & Fan, 2007; Oxford & Scarcella, 1994; Scarcella & Oxford, 1992) and must recognize the importance of vocabulary access in speaking a foreign language. Communication strategies for coping with vocabulary breakdowns in conversations or presentations (Nakatani & Goh, 2007; Oxford, 2011) and speech-act strategies for sociolinguistic appropriateness (Cohen & Ishihara, 2005) also must be taught. Thus, strategy instruction should be integrated into regular language instruction with ease and simplicity (see Oxford, 2011). Online or in-person teacher development could instruct LITE teachers how to weave such speaking-related strategy instruction into regular language instruction to spark faster progress in speaking proficiency. Moreover, some speech-act strategy instruction directly tailored for students is online (see Cohen & Ishihara, 2005; Olshtain & Cohen, 1989).

Self-reported speaking proficiency and willingness to communicate were significantly and strongly linked at the end of the LITE (r = .69). This suggests that when one's speaking proficiency increases, one is more willing to communicate in the language, and when one is more willing to communicate, one's speaking skill becomes more powerful. Just as self-reports of speaking proficiency increased during the LITEs, willingness to communicate also significantly increased (though with a small effect size). Post-LITE willingness to communicate was also significantly tied to post-LITE perspective-taking confidence. Such confidence was, not surprisingly, also strongly related to speaking proficiency. The following seem to form a wonderfully positive post-LITE nexus: speaking skill, willingness the use the language, and confidence in taking another perspective.

We did not contrast OCONUS and CONUS results in this report because of the small number of CONUS LITE participants. However, we need to emphatically encourage OCONUS and CONUS LITE sites to promote speaking proficiency in and outside of class. OCONUS programs should ensure that LITE participants have abundant in-class and extracurricular opportunities to get to know and communicate orally with native speakers of target languages. This is relatively easy to arrange, given that OCONUS LITEs occur in target language locations. CONUS LITE participants should not lack speaking opportunities, either. As mentioned in last year's evaluation report (Gunther et al., 2011a), CONUS LITE site managers or teachers should find native speakers of the target language for evening or weekend social events involving LITE participants. A strong expectation of CONUS programs should be to ensure that LITE participants have as many outside-of-class, informal contacts as possible with target language speakers throughout the LITE. CONUS programs can often be nearly as rich in speaking opportunities as OCONUS programs if the proper planning occurs.

Expectations and Challenges

As Section 4 indicated, before the LITEs the participants wrote down their expectations, hopes, and aims in their own words. Seventy-seven percent had expectations about gaining language proficiency, while 14% were hoping for greater cross-cultural competence, and the rest had other expectations. Regarding expectations for language proficiency, speaking and listening were in the lead. Improvement of literacy (reading and writing) was not as profound a desire for LITE participants. Improving cross-cultural competence was important to many participants. Other expectations, such as contributing to the USAF also clearly emerged.

The challenges that the participants wrote about after the LITEs were often about cultural issues (45% of all the challenges cited). Many of the challenges reported by participants should be recorded by instructors and used as teaching points, especially opportunities to learn and practice unfamiliar forms of discourse such as persuading and negotiating in the target language.

Communication challenges (38.3% of all the challenges mentioned) included, among other things, accent, vocabulary, and cross-training in a new language rather than continuing in the originally studied foreign language. One challenge that emerged was that some LITE participants had difficulty finding native speakers who spoke an educated form rather than a broken, less competent form of the target language. Some LITE participants did not know where to look for native speakers, and others found that native speakers did not always want to communicate with LITE participants in the target language. Staying with a host family usually helps, but it is not the total answer, nor is a host family always available. It should not be totally up to the LITE participants to make all their own out-of-class contacts for communication, particularly in certain countries, although we recognize that participants do share (with site teachers and staff) the responsibility to locate those communication partners. Program managers and teachers at the LITE locations should help to ensure – through planned gatherings and informal sharing of connections – that LITE participants actually make contacts with at least some educated, interested native speakers outside of the classroom. Clearly, this is an issue that needs to be addressed at some of the LITE sites and by the LITE headquarters staff.

Perhaps surprisingly, nearly one out of 10 of the post-LITE narrative responses said there were no major challenges during the LITEs. We believe they meant no unexpected or unwarranted challenges, rather than none at all. We should also note that not all of the LITE participants wrote narrative responses after the LITEs were over, so we do not know their perspectives.

Discussion of the Assessment Tools Used in the Evaluation

Remaining assessment issues for this evaluation involved the DLPT's restricted range, the need for raw scores for the DLPT, self-assessment issues, assessment of self-efficacy in intercultural communication, and inadequate site-level data.

Restricted range of the DLPT. The fact that the DLPT does not measure higher-range proficiency for most languages put some LITE participants at a disadvantage, since an increasing percentage of participants enter the LITEs with proficiency at 2+ or higher in reading and

listening. We strongly encourage a systematic effort to create DLPTs in all strategic languages with higher proficiency "ceilings" up to level 4 or 5, so that individuals who start a LITE at level 2+ or 3 can have a proper assessment on the DLPT.

Lack of raw scores for and timing of the DLPT. In addition, without raw scores for the DLPT it is very difficult to track small gradations of progress. Given that the LITEs are generally only four to six weeks long, it is unfortunate that we do not have access to the raw scores. Moreover, raw scores might help alleviate the issue of whether to use parametric versus nonparametric statistics, an issue that always dogs researchers when dealing with ILR-level data.

Further, we were unable to control for the timing of the pre and post DLPT assessments. Participants are allowed to take the DLPT only once every six months, and the scores are valid for one year. Thus, the timeframes varied across the participants for when they actually took these assessments.

Self-assessment issues. Developing the Speaking Can-Do Scale and validating it with the scalogram technique were two important goals of the present research. The Speaking Can-Do Scale proved to be a useful measure. However, as noted in last year's LITE evaluation report (Gunther et al., 2011a), it would clearly help if all LITE participants had a formal oral proficiency interview (OPI) both before and after the LITEs. No speaking self-assessment tool can take the place of an OPI. Even a computer-based, simulated OPI would be more useful than a self-assessment tool for speaking proficiency.

Self-assessments of language proficiency, such as the Speaking Can-Do Scale, are advantageous for several reasons: simplicity of administration, helpfulness in spurring students to reflect on their performance, and adaptability for peer or instructor assessment of student performance. The National Language Service Corps Self-Assessments demonstrated significant, moderate correlations (between .41 and .54) with Oral Proficiency Interviews (Stansfield, Gao, & Rivers, 2009). Ross (1998) found significant meta-analytic average correlations between language self-assessments and performance measures (r = .55, .61, and .65 for speaking, reading and listening proficiency, respectively). A much stronger relationship was identified by the Association of Language Testers in Europe, which reported a correlation of .91 between language self-assessments and Common European Framework Level exams (ALTE, 2002).

Despite the advantages of language proficiency self-assessment, research has found many disadvantages, which would support the need to use a formal OPI instead of the Speaking Can-Do Scale. For example, self-assessments of ability are often inflated as compared to objective measures. Kruger and Dunning (1999) found that people commonly overestimate their own abilities, with those least able overestimating the most (a finding consistent in language testing as well; Blanche & Merino, 1989). We found this to be the case with the Speaking Can-Do Scale and with the ICSE.

Moreover, sometimes self-assessments are more strongly related to affective variables, such as motivation, than to ability measures, leading some to believe that self-assessments – even of cognition or performance – should be considered measures of affect instead of ability (Sitzmann et al., 2010). Using 2010 LITE data, Gunther et al. (2011b) found that correlations

between self-assessment of language proficiency and affective variables were significant, a finding that supported Sitzman et al. With the 2011 LITE data, the Speaking Can-Do Scale was correlated moderately with an affective variable, willingness to communicate (r = .52, p < .01, n = 60), so the "affective load" of a self-assessment of language proficiency should be further studied.

Self-assessments also are difficult to construct so that specific items are indicators of proficiency at particular levels. To be valid measures, language self-assessments should relate to objectively measured language proficiency data (such as the DLPT) with each item corresponding to a known level of underlying ability. The items should be endorsed in order, with items hypothesized as "basic proficiency" endorsed by all respondents at or above that proficiency level (i.e., a scalogram response pattern; Guttman, 1950). The researchers involved in the Self-Assessment Project for the Association of Language Testers in Europe (ALTE, 2002) found that several features of self-assessment statements cause them to perform poorly from a statistical standpoint. Statements that were negatively worded, that had overly complex sentence structures, or that referred to technology use, for instance, were not endorsed in accordance with their hypothesized level of language ability.

In the present research, we found that the wording – particularly for the self-assessed ICSE – was problematic, even though we employed the ILR's own particular wording as much as possible. In conducting a validity exercise (blind categorization of items to levels) to determine whether the scale items actually seemed to fit logically in given levels, we questioned not only the ILR's wording in some instances but also the ILR's logic in placing some behaviors at a given level and not another. For instance, communicating empathy is not necessarily limited to Level 4, and shifting with ease from informal to formal communication styles can be done at many levels.

Many of the above comments about self-assessment underscore (a) the value of using objective measures for language proficiency rather than a self-report measure, as we used in speaking; and (b) the necessity of helping to clear up the logic and the wording of the ILR Intercultural Communication Competence skill-level descriptions.

Another limitation of the evaluation was that the lack of dichotomous scoring for the ICSE prevented us from validating that scale by means of a scalogram analysis.

Inadequate LITE site data. For several years the LEAP staff has relied solely on participants to give their perspectives on instructional and program variables at the sites. There has been no formal, systematic data collection of the LITEs using any other sources. No systematic effort has yet been made to gather data from LITE site administrators or teachers about what exactly they do during the LITEs and their perspectives on their own performance.

Recommendations

Overall, the 2011 LITEs met the objective of increasing language proficiency and cultural competence by offering short periods of intensive engagement with the target language. Strongly motivated participants were the other crucial part of the equation. Without their drive and stamina, no amount of instruction would have been helpful. Below are our recommendations regarding the LITEs.

1. Continue the LITEs as an integral part of the USAF LEAP.

2. Build and systematize the site-level part of the LITE evaluation to identify what is working effectively and what needs to be changed.
   a. LEAP staff members must visit sites and systematically gather program data from instructors and site administrators. This can involve use of existing site observation forms and interview forms. Student focus groups can also be used on site for evaluation, deepening the narrative strand and giving a clearer picture of site-level instruction and perspectives.
   b. In the event that on-site observations and visits cannot be accomplished, a process should be in place that allows site instructors and administrators to collect and send the relevant information and data to the LEAP staff for analysis.

3. For greater language proficiency, LITE teachers should teach participants to use optimal learning strategies for reading, listening, and speaking. Given that learning strategies are often significantly related to proficiency and are highly teachable (Oxford, 2011), it would be useful to encourage LITE teachers to focus more intentionally on those learning strategies. This is especially important because after the LITEs, participants will need to be able to use self-directed learning strategies to continue their progress, even if they have the help of eMentoring.

4. LITE participants at some sites commented that they do not have adequate access to educated native speakers of the target language outside of class. Site administrators and teachers at those specific sites should go out of their way to create formal gatherings or informal connections to provide the needed access outside of class. In addition, LEAP staff can write a brochure or guidebook for how LITE participants can find conversation partners and make friends outside of class.

5. According to narrative findings, some LITE participants need more guidance on how to use difficult cultural situations, such as pressure from vendors, as cultural and communicative learning experiences. This type of learning can be best handled at the site level, but LEAP staff might want to give some written guidelines on how to interpret and use difficult cultural encounters in the most productive way.

6. Each of the LITE sites currently goes its own way in terms of curriculum and instruction, with little guidance from the LEAP staff beyond the abstract expectations of “push the participants to a higher DLPT level” or “focus on communication.” This situation might work adequately or well as long as all the LITE participants are highly intelligent, well-educated, and highly motivated and have a language background, but if any of those personal criteria does not exist

(particularly if LEAP expands its base of applicants to those who have less education and less intrinsic motivation), the hands-off, non-guidance approach will not work any longer. One site visitor remarked that site administrators said they would appreciate additional guidance from LEAP staff.

7. LEAP staff should clearly mandate the use of OPIs or a computerized version of OPIs, because there are limitations associated with self-assessment of language proficiency. Greater effort and energy should be put into finding an appropriate speaking proficiency assessment tool.

References

Afflerbach, P., Pearson, P., & Paris, S. (2008) Clarifying differences between reading skills and reading strategies. *The Reading Teacher* 61(5), 364-373.

Association of Language Testers in Europe (ALTE). *The ALTE Can-Do Project 1992-2002.* Retrieved on April 9, 2011 at http://www.alte.org/cando/index.php.

Anderson, N. J. (1991) Individual differences in strategy use in second language reading and testing. *Modern Language Journal, 75* (4): 460-472.

Blanche, P. & Merino, B. (1989). Self-assessment of foreign language skills: Implications for teachers and researchers. *RELC Journal, 19*, 75-93.

Clément, R.C., Baker, S.C., & MacIntyre, P.D. (2003). Willingness to communicate in a second language: The effects of context, norms, and vitality. *Journal of Language and Social Psychology* 22(2), 190-209.

Clifford, R. (2010, July). *Hours to attain specific proficiency levels in different language categories, based on Defense Language Institute Foreign Language Center (DLIFLC) data*. Personal communication. Brigham Young University, Provo, UT.

Cohen, A.D. & Ishihara, N. (2005). *A web-based approach to strategic learning of speech acts: Research report.* Center for Advanced Research on Language Acquisition, University of Minnesota, Minneapolis, MN.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, Ohio: Cengage.

Defense Language Institute. (n.d.). *Defense Language Proficiency Test 5 System: Familiarization Guide for Multiple Choice Format.* Retrieved on May 8, 2013 from http://www.dliflc.edu/archive/documents/Generic-Fam%20Guide-MC-CBT.pdf

Dörnyei, Z. (2003). *Attitudes, orientations, and motivations in language learning: Advances in theory, research and applications.* Malden, MA: Wiley-Blackwell.

Gardner, R.C. (1982). Language attitudes and language learning. In E. Bouchard Ryan & H. Giles, *Attitudes towards language variation* (pp. 132-147). Edward Arnold.

Gardner, R.C. (1985). *Social psychology and second language learning: The role of attitudes and motivation.* London, Ontario: Edward Arnold.

Gardner, R.C. (2004). *Attitude/Motivation Test Battery: International AMTB research project.* London, Ontario: University of Western Ontario. Retrieved on April 4, 2010 from http://publish.uwo.ca/~gardner/docs/englishamtb.pdf.

Gardner, R., Tremblay, P., & Masgoret, A-M. (1997). Towards a full model of second language learning: an empirical investigation. *The Modern Language Journal*, 81, 344-362.

Gehlbach, H. (2004). A new perspective on perspective taking: A multidimensional approach to conceptualizing an aptitude. *Educational Psychology Review, 16*(3), 207-234.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step*. Boston: Pearson Education.

Gu, P.Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning 46*, 643-679.

Gunther, K., Doyle, A., Oxford, R., & Brittain, M. (2011a). *Language Enabled Airman Program: Language Intensive Training Events: 2010 results.* Technical report. Montgomery, AL: Air Force Culture and Language Center, Maxwell Air Force Base.

Gunther, K., Oxford, R., Brittain, M., & Doyle, A. (2011b). *Language self-assessments: A test of scale ordering and validity*. Paper presented at the annual meeting of the International Military Testing Association, Bali, Indonesia.

Guttman, L. (1950). The basis for scalogram analysis. In S.A. Stouffer et al. (Eds.), *Measurement and prediction. The American soldier, Vol. IV*. Princeton, N.J.: Princeton University Press.

Horwitz, E. (2007). *Words fail me: foreign language anxiety crippling for some students.* E. Horwitz interviewed by K. Randall. University of Texas at Austin feature story. Retrieved on June 10, 2010 from http://www.utexas.edu/features/2007/language/.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125-132.

Hosenfeld, C. (1977). A preliminary investigation of the reading strategies of successful and nonsuccessful second language learners. *System* 5(2), 110-123.

Interagency Language Roundtable (ILR). (2009). *Skill level descriptions for intercultural communication*, Draft version 3.2. Washington, DC: Author.

Interagency Language Roundtable (ILR). (2012). *The new ILR skill-level descriptions for intercultural communication.* Washington, DC: Author.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121-1134.

Lee, K.R. (2007). *Strategy awareness-raising for success: Reading strategy instruction in the EFL context*. Unpublished doctoral dissertation, University of Maryland, College Park, MD.

Macaro, E., Graham, S., & Vanderplank, R. (2007). A review of listening strategies: Focus on sources of knowledge and on success. In A.D. Cohen & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice* (pp. 165-185). Oxford, UK: Oxford University Press.

MacIntyre, P.D. (2003, June 1). *Willingness to communicate in the second language: Proximal and distal influences.* Paper presented at the annual conference of the Canadian Association of Applied Linguistics, Halifax, NS, Canada.

MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K.A. (1998). Conceptualising willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal*, 82, 545-562.

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44*, 283-305.

Mendelsohn, D. J., & Rubin, J. (1995). *A guide for the teaching of second language listening.* San Diego, CA: Dominie Press.

McCroskey, J. (1992). Reliability and validity of the Willingness to Communicate scale. *Communication Quarterly, 40*(1), 16-25.

Mokhtari, K. & Reichard, C.A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology* 94(2), 249-259.

Mokhtari, K. & Sheorey, R. (2002). Measuring ESL students' reading strategies. *Journal of Developmental Education* 25(3), 2-10.

Nakatani, Y. & Goh, C. (2007). A review of oral communication strategies: Focus on interactionist and psycholinguistic perspectives. In A. Cohen & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice.* Oxford, UK: Oxford University Press.

Nation, I.S.P. (1990). *Teaching and learning vocabulary.* New York: Newbury House.

Nation, I.S.P. & Waring, R. (1997). Vocabulary size, text coverage, and word lists. Retrieved April 6, 2011 from http://www1.harenet.ne.jp/~waring/papers/cup.html

Noels, K. A., Pelletier, L. G., Clement, R., & Vallerand, R. J. (2003). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning, 53*, 33-64.

Nyikos, M. & Fan, M. (2007). A review of vocabulary learning strategies: Focus on language proficiency and learner voice. In A.D. Cohen & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice* (pp. 251-273). Oxford, UK: Oxford University Press.

Olshtain, E. & Cohen, A. (1989). Speech act behavior across languages. In H. Dechert & R. Manfred (Eds.). *Transfer in language production* (pp. 53-67). Norwood, NJ: Ablex.

Oxford, R.L. (1990). *Language learning strategies: What every teacher should know*. Boston: Heinle/Cengage.

Oxford, R.L. (2010, Nov.). Instructional objectives for each ILR language skill level. *Toward a coherent, successful curriculum for the Language Enabled Airman Program.* Montgomery, AL: Air Force Culture and Language Center, Maxwell Air Force Base.

Oxford, R.L. (2011). *Teaching and researching language learning strategies*. Harlow, Essex, UK: Pearson Longman.

Oxford, R.L. & Scarcella, R.C. (1994). Second language vocabulary learning among adults: State of the art in vocabulary instruction. *System* 22(2), 231-243.

Oxford, R. L. & Shearin, J. (1994). Language learning motivation: Expanding the theoretical framework. *Modern Language Journal, 73,* 291-300.

Petri, H., & Govern, J. (2004). *Motivation: Theory, research and applications*. Belmont, CA: Thomson-Wadsworth.

Rentsch, J., Gunderson, A., Goodwin, G., & Abbe, A. (2007). *Conceptualizing multicultural perspective-taking, Technical Report 1216.* Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Rhodes, T. (Ed.). (2010). *Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics*. Washington, DC: Association of American Colleges and Universities.

Ross, S. (1998). Self assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing,* 15, 1-20.

Rost, M. (2010). *Teaching and researching listening.* 2nd ed. London: Longman.

Ryan, R.M & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78. http://home.ubalt.edu/tmitch/641/deci_ryan_2000.pdf.

Scarcella, R.C. & Oxford, R.L. (1992). *The tapestry of language learning: The individual in the communicative classroom.* Boston: Heinle / Thomson Learning.

Sessa, V. I. (1996). Using perspective-taking to manage conflict and affect in teams. *Journal of Applied Behavioral Science, 32*, 101-115.

Sitzmann, T., Ely, K., Brown, K., & Bauer, K. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education, 9*(2), p. 169-191.

Stansfield, C., Gao, J., & Rivers, B. (2009). Evaluating the validity of NLSC self-assessment scores. Retrieved on April 4, 2011 from http://www.govtilr.org/Publications/Evaluating%20the%20Validity%20of%20NLSC%20Self-Assessment%20Scores_Oct27%20(2).ppt.

Vandergrift, L.**,** Goh, C., Mareschal, C. & Tafaghodatari, M.H. (2006). The Metacognitive Awareness Listening Questionnaire (MALQ): Development and validation. *Language Learning* 56, 431-462.

Winograd, P. & Hare, V.C. (1988). Direct instruction of reading comprehension strategies: The nature of teacher explanation. In C.E. Weinstein, E.T. Goetz, E.T., & P.A. Alexander (Eds.), *Learning and study strategies: Issues in assessment, instruction, and evaluation* (pp. 121-139). San Diego, CA: Academic Press.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., and Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.

Appendix A
Acronyms, Definitions and Program Types for LEAP LITEs

Table A1. List of Acronyms

Term	Definition
AACU	Association of American Colleges and Universities
AFCLC	Air Force Culture and Language Center
AFSC	Air Force Specialty Codes
CONUS	Continental U.S. (contiguous 48 states)
DLI	Defense Language Institute
DLPT	Defense Language Proficiency Test
GPF	General Purpose Force
ILR	Interagency Language Roundtable
ICSE	Intercultural Communication Behavior Self-efficacy Scale
LEAP	Language Enabled Airman Program
LITE	Language Intensive Training Event
LLOS-IEA	Language Learning Orientations Scale – Intrinsic Motivation, Extrinsic Motivation, and Amotivation Subscales
OCONUS	Outside the Continental U.S (outside of contiguous 48 states)
OPI	Oral Proficiency Interview
OTS	Officer Training School
ROTC	Reserve Officer Training Corps
SLL	Strategic Language List
USAF	United States Air Force
USAFA	United States Air Force Academy
WTC	Willingness to Communicate

Definitions and Program Types for LEAP LITEs

Target Language – Language being learned

Foreign Language Setting – An environment in which the target language (e.g., French in Alabama) *is not* the language of daily communication of the majority of people in the geographic region where the learner is learning it; thus, an environment in which learners do not have ready-made opportunities for using the target language beyond the classroom and have to work very hard to find them (through special media opportunities, books, or an occasional tourist or immigrant) (Oxford, 1990).

Second Language Setting – An environment in which the target language (e.g., Chinese in China or French in France) *is* the language of daily communication of the majority of people in the geographic region where learner is learning it; thus, an environment in which there are abundant target language and target culture resources, most people are native speakers of the target language, and learners can have almost constant opportunities to use the language for communication (Oxford, 1990).

Intensive Program – In an *intensive program*, participants have at least 3-4 classroom contact hours per day, 5 days a week, with an anticipated 2-3 additional hours of language study per day, but out-of-class and non-homework hours can be spent in the native language unless the program intentionally provides further target language exposure. There is no pledge to use the target language full-time. These programs are often very good at increasing reading and listening proficiency. Unless the teachers and programs intentionally integrate culture and language during classes and organize some meaningful target-language communication opportunities outside of class (as they should be expected to do), intensive programs might not be as good for gaining cultural understanding, increasing speaking proficiency, or improving knowledge of how to use the language in context, i.e., pragmatics. An intensive program can be conducted CONUS (e.g., Korean in a language training detachment at Maxwell AFB) or OCONUS (Swahili at PLTCE, Russian in Kiev).

Immersion Program– In an *immersion program*, the participant is surrounded by the target language 24 hours per day, and daily survival and all communication *are expected to involve the target language*.[6] In CONUS immersion programs, a target-language-only pledge is generally used to confirm this expectation. Special aims of immersion programs are often to gain cultural understanding, to increase speaking proficiency, and to improve knowledge of how to use the language in context, i.e., pragmatics, but reading and listening remain important in immersion programs. In immersion programs, language classes of at least 3-4 hours per day are taught in the target language, with culture smoothly integrated (though field trips might take the place of classes on certain days, especially OCONUS). Homework, which can range from one to three hours, builds on the class activities. Additional evening and weekend programming provides activities in the target language and culture.

Special LITE Internship – For a special LITE internship, an Airman who is at an ILR level 3 or above has been selected to use his or her foreign language skills in performing an actual job at an OCONUS location. There are generally no language classes involved.

[6]In K-12 schools, the definition of immersion is different and involves only the school day or even part of a school day. That definition is not relevant to adult language programs.

Appendix B
Procedures and Results of the Scalogram for the Speaking Can-Do Scale

Table B1. Speaking Descriptions and Examples for Each ILR Language Proficiency Level

ILR Level	Description	Examples
1 - Elementary Proficiency	Ability to satisfy minimum courtesy requirements and maintain very simple face-to-face conversations on familiar topics	• Can exchange greetings • Can provide skeletal bio information • Can give information about business hours • Can explain routine procedures in a very limited way • Can make travel and accommodation needs known • Vocabulary is extremely limited • Speech is slow • Repetition of phrases is necessary
2 - Limited Working Proficiency	Able to satisfy most work requirements with language usage that is often, but not always, acceptable and effective	• Can handle routine work-related interactions • Can participate in casual conversations about work, family, autobiographical information, and current events • Can give simple, straightforward instructions to subordinates • Speech is perceived by native speakers as awkward and inaccurate • Cannot always easily produce general vocabulary
3 - General Professional Proficiency	Able to speak language with sufficient structural accuracy and vocabulary to participate effectively in most formal and informal conversations in practical, social, and professional topics	• Can communicate acceptably • Can effectively combine structure and vocabulary to convey meaning accurately • Can easily repair a conversation • Can clarify points, justify decisions, state and defend policy, conduct meetings, and deliver briefings • Strength is in breath of vocabulary, including sociolinguistic/cultural references and nuances of close synonyms • Some noticeable imperfections continue
4 - Advanced Professional Proficiency	Able to use the language fluently and accurately on all levels normally pertinent to professional needs	• Language usage and ability to function are fully successful • Can organize discourse well • Can use appropriate rhetorical speech devices and native cultural references • Can serve as an informal interpreter • Can advocate a position at length, both formally and in chance encounters • Can control vocabulary and phrasing in a sophisticated way that is rarely imprecise • Shows only occasional weaknesses in idioms, colloquialisms, pronunciation, and cultural references

Here are the steps used in the scalogram analysis:

Step 1. Assigning 1 or 0 to an item for each individual. A positive response to an item was assigned a value of 1, and each negative response a value of 0.

Step 2. Calculating positive items within a level for each individual. For each level, the number and percentage of positive responses was calculated.

Step 3. Identifying the individual's proficiency (or non-proficiency) at each level. If a participant positively endorsed (gave a positive response to) 80% of the items for a level, he or she was considered to be "proficient" at that level. For each level at which at least 80% of the items were endorsed, the participant was assigned a value of 1 for that level. If fewer than 80% of the items were endorsed at a given level, the participant was assigned a value of 0 for that level. Therefore, each participant had a 1 or 0 assigned for all four levels.

Step 4. Identifying expected (allowable) response patterns across levels for each individual. Given the premise of the ILR, individuals should increase in proficiency in order from level 1 to level 4. They should *not* be able to perform 80% of level 3 speaking behaviors if they do not also have level 1 proficiency and level 2 proficiency (i.e., positive responses to at least 80% of level 1 items and level 2 items). Thus, on the Speaking Can-Do scale, certain patterns of response are expected, and others are not expected.

In a scalogram analysis, participants' response patterns are visually inspected and coded for allowable (i.e., expected) response patterns. An allowable response pattern (see examples in Set 1 in Table B2 below) is one in which all lower speaking levels are endorsed, in this case at 80% or more of the items, and higher levels are *not* endorsed at 80% or more of the items. For instance, Respondent C in Set 1 is considered as having an allowable response pattern because he or she endorsed at least 80% of items at level 1 and level 2 but did not endorse at least 80% of items at level 3 or level 4. All the rest of the response patterns in Table B2 are also allowable (expected).

Table B2. Set 1, Examples of Allowable Response Patterns

	Levels			
	L1	L2	L3	L4
Respondent A	1	1	1	1
Respondent B	1	1	1	0
Respondent C	1	1	0	0
Respondent D	1	0	0	0
Respondent E	0	0	0	0

Any response pattern that fails to show endorsement of all lower levels while endorsing an upper level is coded as an "error" (see examples in Set 2 in Table B3 below). For example, Respondent C in Set 2 did not have an acceptable response pattern, because he or she endorsed at least 80% of items at level 2 but failed to endorse at least 80% of items at level 1 (indicating he or she is not proficient at that level). All the other response patterns in Table B3 are erroneous (disallowed).

Table B3. Set 2, Examples of Disallowed Response Patterns (Errors)

	Levels			
	L1	L2	L3	L4
Respondent A	1	1	0	1
Respondent B	1	0	1	0
Respondent C	0	1	0	0
Respondent D	0	0	1	1
Respondent E	1	0	0	1

Step 5. Determining the coefficient of reproducibility. The degree to which the Speaking Can-Do scale demonstrates the allowable response patterns is assessed by the coefficient of reproducibility (CR). The formula is as follows:

CR = 1 – (total number of errors/total number of responses)

A .90 CR is often used as the rule of thumb for evidence of a well-ordered scale (Crocker & Algina, 2008). Once evidence of an ordered scale is established, support for the comparability of the Speaking Can-Do Scale and DLPT scores for listening and reading would be significant correlations among these three variables (speaking, listening, and reading). In the scalogram analysis of the Speaking Can-Do Scale, only 2 of the participants' response patterns were disallowed (coded as errors) on the pre-LITE scale. See Table B4.

Table B4. Pre-LITE Speaking Can-Do Levels - Disallowed Response Patterns

	Levels			
Pre-test	L1	L2	L3	L4
Error 1	✓	✓		✓
Error 2	✓		✓	

Key: (✓) = Levels Endorsed on Disallowed Response Patterns

On the post-LITE Speaking Can-Do Scale, only 3 of the participants' response patterns were coded as errors. See Table B5.

Table B5. Post-LITE Speaking Can-Do Levels - Disallowed Response Patterns

	Levels			
Post-test	L1	L2	L3	L4
Error 1	✓	✓		✓
Error 2	✓		✓	
Error 3	✓		✓	

Key: (✓) = Levels Endorsed on Disallowed Response Patterns

Since the scores of respondents with disallowed response patterns do not accurately reflect their standing at an ILR level, these scores of these individuals were removed from the dataset when conducting analyses using the Speaking Can-Do Scale.

An item analysis was then conducted based on the scalogram results from the pre-LITE Speaking Can-Do data. The item analysis showed that of the 34 items in the Speaking Can-Do Scale, 8 did not perform exactly as might be expected when compared to performance of individuals on the DLPT, which was, like the Speaking Can-Do Scale, based on the ILR Language Proficiency Skill-Level Descriptions. These items might be viewed as potentially problematic, and further examination is warranted in the future. We retained these items in the analysis this year.

- Level 2, item 8. *I can give complicated, detailed, and extensive directions*. The percentage of participants who endorsed this item (i.e., who said they could give such directions) was very similar across all DLPT levels, ranging from 45% to 57%. Given that this was originally built from the ILR level 2 speaking skill-level description and was therefore viewed as a level 2 item, it would be expected that the proportion endorsed should be higher for those who had a level 2, level 3, and level 4 on DLPT listening. One possible problem is that the meaning of the term "directions" might have been ambiguous. Perhaps giving geographic directions is easier (or harder) than giving directions for a task.

- Level 3, item 4. *I can readily give presentations about professional topics*. For this item, 36% of level 1 participants gave an endorsement (said they could do it), yet only 15% of level 2 DLPT scorers and 29% of level 3 DLPT scorers endorsed this item.

- Level 3, item 8. *I can quickly repair a conversation if there is a cultural misunderstanding*. Over 80% of the DLPT group Level 1 endorsed this item, yet only 69% of the DLPT group level 2 endorsed the item.

- Level 3, item 9. *I can speak well enough that native speakers seem to understand me.* This item did not perform as expected because almost all participants across all DLPT levels endorsed the item. Level 1 participants endorsed the item at 100%, while level 2 participants and level 3 participants endorsed the item at 96% and 95%, respectively. Thus, almost all LITE participants expect native speakers of the target language to understand them, even if this belief does not reflect reality. This might be an interesting insight about human nature, or at least about the LITE sample.

- Level 4, item 3. *I can appropriately do any of the following things: advise, command, argue, persuade, and negotiate.* This item did not perform as expected because a greater proportion of level 1 participants (55%) endorsed the item than the level 2 participants (35%). Advising, commanding, arguing, and persuading are often high-level communication skills in any language, and it is unclear why level 1 participants believed they could do this in the target language. Alternatively, perhaps the level 1 participants who endorsed the item believed they could perform these language functions with very minimal vocabulary with the help of nonverbal communication.

- Level 4, item 7. *I can change my tone to make it appropriate for spoken interactions in virtually all settings, whether formal or informal*. This item did not perform as expected because a greater proportion of level 1 participants (73%) endorsed the item than level 2 participants (62%). Level 1 participants might have thought that this referred merely to tone of voice instead of tone as conveyed through vocabulary, grammar, and sociolinguistic register.

How can we explain the unexpected item responses, particularly the ones in which learners at lower levels of proficiency on the DLPT claimed greater speaking ability on the Speaking Can-Do than did learners at higher DLPT levels? There are several possibilities. First, we might guess that speaking operates very differently from reading and listening. However, the DLPT levels and Speaking Can-Do levels are positively and significantly correlated (see Section 3). Second, in some cases, the wording of new items might not have conveyed the meaning to all participants (see level 2, item 8 and level 4, item 7 above). Third, individuals who are not far along in learning a language might not realize how much they do not yet know and how much they are not yet able to do. Fourth, sometimes respondents at lower levels of proficiency on the DLPT might think they can communicate at higher levels even with minimal vocabulary merely through force of will and the help of nonverbal communication. Fifth, those at level 1 on the DLPT might intentionally try to make themselves (or the researchers) feel better by exaggerating their capability.

Appendix C
ILR Intercultural Communication Competence Levels

Table C1. Descriptions and Examples for Each ILR Intercultural Communication Competence Level

ILR Level	Description	Examples
1 - Elementary Competence	Able to participate in some everyday interactions, though not always acceptably	• Can recognize cultural differences but does not show understanding of significance of differences • Can conform to cultural practices during interactions, such as posture, eye contact, and distance from others • Avoids well-known taboo topics and behaviors • May exhibit confusion when faced with unfamiliar cultural cues
2 - Limited Working Competence	Able to participate acceptably in many everyday social and work-related interactions	• Shows awareness of significant differences between cultures • Attempts to adjust behavior, although not always successfully • Can typically avoid taboos and adhere to norms & rules of etiquette, such as offering and receiving gifts • Can give straightforward directions and instructions in work environment • May be able to address some job-related issues • May sometimes misinterpret cultural cues, but usually able to repair misunderstandings
3 - Professional Competence	Able to participate successfully in most social, practical, and professional interactions, including those that may require a range of formal and informal language and behavior	• Can transition smoothly from informal to formal styles of communication • Can control nonverbal responses, such as gestures • Can understand and make appropriate use of cultural references and expressions • Can interact appropriately during meetings • Rarely misreads cultural cues • Reflects significant knowledge and understanding of cultural expectations during interactions
4 - Advanced Professional Competence	Able to participate successfully in virtually all social, professional, and official interactions, including those where leadership is required	• Can respond effectively to verbal and nonverbal forms of communication • Can participate successfully in public discourse, such as presentations, conferences, speeches, and media interviews • Almost always correctly interprets visual cues, cultural allusions, nuance, tone, and values • Makes frequent and appropriate use of cultural references, literary allusions, literary quotes, and other significant documents

Appendix D
Item and Scale Descriptive Statistics for Pre- and Post-Survey Measures

Table D1. Item and Scale Descriptive Statistics for Pre- and Post-Survey Measures

			Pre-test (n~63)		**Post-test (n~48)**	
	Cronbach's Alpha	*Item-Scale r*	*M*	*SD*	*M*	*SD*
Willingness to Communicate *(I am willing to…)*	***0.943***		*4.22*	*0.794*	*4.25*	*0.726*
…present a talk to a group of friends.		*0.830*	*4.46*	*0.930*	*4.50*	*0.875*
…talk with a service station attendant.		*0.799*	*4.56*	*0.713*	*4.58*	*0.767*
…talk in a small group of strangers.		*0.833*	*4.32*	*0.913*	*4.40*	*0.818*
…talk in a large meeting of strangers.		*0.800*	*3.83*	*1.144*	*3.79*	*1.091*
…talk with a stranger while standing in line.		*0.743*	*4.11*	*0.994*	*4.15*	*0.899*
…talk with a police officer.		*0.758*	*4.10*	*1.067*	*4.29*	*0.922*
…present a talk to a group of strangers.		*0.769*	*3.78*	*1.099*	*3.79*	*1.129*
…present a talk to a group of acquaintances.		*0.798*	*4.32*	*0.877*	*4.23*	*0.890*
…talk with a salesperson in a store.		*0.760*	*4.51*	*0.780*	*4.50*	*0.899*
Language Anxiety	***0.803***		*2.83*	*0.787*	*2.99*	*0.740*
It upsets me when I don't understand what someone is saying in the language.		*0.311*	*3.46*	*1.060*	*3.59*	*0.888*
I worry about making mistakes in the language.		*0.588*	*3.10*	*0.962*	*3.29*	*0.979*
I get nervous when I am speaking in my target language.		*0.734*	*2.46*	*1.075*	*2.59*	*1.019*
I always feel that the others speak my target language better than I do.		*0.653*	*2.62*	*1.113*	*2.88*	*1.053*
I never feel quite sure of myself when I am speaking in the language.		*0.677*	*2.48*	*1.052*	*2.59*	*0.998*

		Pre-test (n~63)			Post-test (n~48)	
	Cronbach's Alpha	*Item-Scale r*	*M*	*SD*	*M*	*SD*
Perspective-Taking Propensity Subscale *(How often do you try to…)*	***0.850***		*4.02*	*0.624*	*4.08*	*0.581*
…understand people from another culture better by trying to figure out what they are thinking?		*0.528*	*4.00*	*0.747*	*4.14*	*0.707*
…think of more than one reason for a person's behavior?		*0.688*	*4.18*	*0.779*	*4.06*	*0.719*
…"put yourself in someone else's shoes" when you are angry with that person?		*0.691*	*4.06*	*0.780*	*4.08*	*0.731*
…figure out what motivates others to behave as they do?		*0.717*	*4.17*	*0.814*	*4.22*	*0.654*
…figure out what emotions people are feeling when you meet them for the first time?		*0.580*	*3.68*	*1.029*	*3.78*	*0.919*
…understand how other people view a particular situation?		*0.647*	*4.03*	*0.740*	*4.18*	*0.667*
Perspective-Taking Confidence Subscale (*How confident are you that you can…)*	***0.897***		*3.67*	*0.684*	*3.78*	*0.748*
…understand the point of view of people from different backgrounds?		*0.618*	*3.98*	*0.729*	*3.98*	*0.729*
...figure out why a person is acting strangely?		*0.614*	*3.60*	*0.814*	*3.66*	*0.815*
...figure out what a person is thinking when you disagree with him or her?		*0.818*	*3.63*	*0.848*	*3.85*	*0.875*
...accurately guess what motivates another person?		*0.733*	*3.62*	*0.923*	*3.70*	*0.858*
...understand what strangers are thinking when you are talking with them?		*0.756*	*3.59*	*0.854*	*3.63*	*0.937*
...figure out what other people are feeling?		*0.759*	*3.62*	*0.854*	*3.85*	*0.899*

		Pre-test (n~63)			Post-test (n~48)	
	Cronbach's Alpha	*Item-Scale r*	*M*	*SD*	*M*	*SD*
Intrinsic Motivation Subscale	***0.753***					
…for the satisfaction I experience in knowing the language better.		*0.422*	*4.56*	*0.646*	*4.56*	*0.542*
…because I enjoy learning new information about the culture.		*0.313*	*4.69*	*0.501*	*4.53*	*0.616*
…for the pleasure I get when I can do more with the language.		*0.652*	*4.44*	*0.646*	*4.41*	*0.610*
…for the enjoyment I experience when I grasp something difficult in the language.		*0.534*	*4.30*	*0.641*	*4.14*	*0.736*
…for the sense of achievement when I understand or say something new in the language.		*0.577*	*4.28*	*0.756*	*4.24*	*0.723*
…for the "high" I experience when using the language.		*0.473*	*3.70*	*1.160*	*3.98*	*0.924*
…for the mental stimulation of learning the language.		*0.449*	*4.21*	*0.686*	*4.20*	*0.645*
Extrinsic Motivation Subscale	***0.736***					
…to get Air Force foreign language proficiency pay.		*0.185*	*3.33*	*0.978*	*3.49*	*1.120*
…to increase my promotion opportunities later.		*0.494*	*3.33*	*0.995*	*3.35*	*1.011*
…to have more opportunities for traveling in the Air Force.		*0.429*	*4.18*	*0.940*	*4.35*	*0.805*
…to be ready in case I can help my country by using the language.		*0.511*	*4.61*	*0.556*	*4.59*	*0.610*
…to help the Air Force with cross-cultural crises or problems.		*0.531*	*4.41*	*0.668*	*4.37*	*0.761*
…to build relationships with people in the culture.		*0.267*	*4.49*	*0.744*	*4.45*	*0.647*
…because I want to more effective in the culture.		*0.387*	*4.61*	*0.525*	*4.57*	*0.577*
…to influence other countries through knowledge of language and culture.		*0.383*	*4.15*	*0.771*	*4.16*	*0.874*
…to influence other militaries as we conduct joint operations.		*0.502*	*4.08*	*0.802*	*3.84*	*0.898*
…to avoid feeling bad if I cannot communicate with people in their native language.		*0.261*	*3.48*	*1.043*	*3.84*	*0.898*

	Pre-test (n~63)			Post-test (n~48)	
Extrinsic Motivation Subscale	*Item-Scale r*	*M*	*SD*	*M*	*SD*
…to avoid seeming like an "ugly American" who does not know other languages or cultures.	*0.281*	*3.77*	*1.160*	*3.90*	*0.984*
…because my image of an ideal Air Force officer is one who speaks more than one language.	*0.300*	*4.05*	*0.956*	*3.92*	*0.954*
…because professionally competent Airmen can get along in other cultures.	*0.404*	*4.38*	*0.734*	*4.20*	*0.763*

		Pre-test			Post-test	
	KR-20	*Item-Scale r*	*Yes*	*No*	*Yes*	*No*
Speaking Can-Do	***0.924***					
Speaking Can-Do - Level 1	***0.751***					
I find it difficult to speak so that I can be understood by a native speaker. *(reverse-coded)*		*0.327*	*83.3%*	*16.7%*	*81.6%*	*18.4%*
I can arrange for a hotel room or taxi ride.		*0.822*	*98.%*	*1.7%*	*100%*	*0%*
I can give or ask for simple directions (e.g., to a hotel, restaurant, or bank).		*0.822*	*98.3%*	*1.7%*	*100%*	*0%*
I can buy a needed item, such as a bus or train ticket, groceries, or clothing.		*0.000*	*100%*	*0%*	*100%*	*0%*
I can ask and answer simple questions about date or place of birth, job, and other demographics.		*0.561*	*96.7%*	*3.3%*	*100%*	*0%*
I can use appropriate expressions for greetings and farewells.		*0.708*	*96.7%*	*3.3%*	*100%*	*0%*
I can order a simple meal.		*0.000*	*100%*	*0%*	*100%*	*0%*
I can introduce myself and others.		*0.822*	*98.3%*	*1.7%*	*100%*	*0%*
Speaking Can-Do – Level 2	***0.776***					
I can give straightforward orders or commands.		*0.396*	*86.7%*	*13.3%*	*87.8%*	*12.2%*
In casual social situations, I can discuss at length current events or activities and family-related topics.		*0.699*	*70.0%*	*30.0%*	*85.4%*	*14.6%*
I can talk about some work-related topics that are non-routine, such as requesting complicated changes in travel arrangements.		*0.587*	*70.0%*	*30.0%*	*71.4%*	*28.6%*

		Pre-test			Post-test	
	KR-20	*Item-Scale r*	*Yes*	*No*	*Yes*	*No*
I can give complicated, detailed, and extensive directions.		*0.537*	*50.0%*	*50.0%*	*63.3%*	*36.7%*
I can ask and answer predictable questions on the job.		*0.569*	*71.7%*	*28.3%*	*95.9%*	*4.1%*
At work I can participate in ordinary discussions using familiar vocabulary and grammar.		*0.423*	*73.3%*	*26.7%*	*93.9%*	*6.1%*
When trying to discuss complex topics, I make frequent errors that seem to disturb native speakers. *(reverse-coded)*		*0.152*	*63.3%*	*36.7%*	*30.6%*	*69.4%*
I can discuss my present or most recent job in some detail using ordinary, non-specialized vocabulary.		*0.530*	*83.3%*	*16.7%*	*85.7%*	*14.3%*
Speaking Can-Do Level 3	***0.710***					
When giving a presentation or speech, I can use organizing phrases to make my speaking clearer (e.g., "The five key points are...," "Therefore...," "In conclusion...," or "To summarize ...").		*0.407*	*47.5%*	*52.5%*	*69.4%*	*30.6%*
I can participate in conversations about practical and social topics, whether in formal or informal settings.		*0.457*	*81.7%*	*18.3%*	*87.8%*	*12.2%*
I can effectively answer concerns about policy decisions.		*0.397*	*20.0%*	*80.0%*	*32.7%*	*67.3%*
I can readily give presentations about professional topics.		*0.364*	*23.3%*	*76.7%*	*45.8%*	*54.2%*
I can ask native speakers the appropriate questions to get the information I need, including details and informed opinions.		*0.633*	*71.7%*	*28.3%*	*81.6%*	*18.4%*
I can speak well enough that native speakers seem to understand me.		*0.312*	*95.0%*	*5.0%*	*95.9%*	*4.1%*
I can quickly repair a conversation if there is a cultural misunderstanding.		*0.406*	*78.3%*	*21.7%*	*87.8%*	*12.2%*
Speaking Can-Do – Level 4	***0.874***					
I can serve as an informal interpreter in a range of unpredictable circumstances.		*0.494*	*60.0%*	*40.0%*	*69.4%*	*30.6%*
I can speak fluently and accurately about complex procedures related to my AFSC.		*0.390*	*18.3%*	*81.7%*	*22.4%*	*77.6%*
I can communicate successfully in virtually all professional situations, no matter how challenging, including giving detailed presentations, official speeches, or media interviews.		*0.343*	*13.3%*	*86.7%*	*18.4%*	*81.6%*

		Pre-test		Post-test	
	Item-Scale r	*Yes*	*No*	*Yes*	*No*
I can readily and appropriately use cultural references, shades of meaning, and figures of speech in conversation.	*0.502*	*38.6%*	*47.1%*	*55.1%*	*44.9%*
I can readily explain to a native speaker any cultural references, shades of meaning, and figures of speech from my own culture.	*0.503*	*53.3%*	*46.7%*	*69.4%*	*30.6%*
My errors are rare and do not hinder my work performance.	*0.462*	*30.0%*	*70.0%*	*34.7%*	*65.3%*
I can speak fluently and accurately about complex, non-work topics of interest to well-educated native speakers.	*0.469*	*31.7%*	*68.3%*	*40.8%*	*59.2%*
I can change my tone to make it appropriate for spoken interactions in virtually all settings, whether formal or informal.	*0.454*	*71.7%*	*28.3%*	*77.6%*	*22.4%*
I can appropriately do any of the following things: advise, command, argue, persuade, and negotiate .	*0.504*	*51.7%*	*48.3%*	*59.2%*	*40.8%*

			Pre-test		Post-test	
	KR-20	*Item-Scale r*	*M*	*SD*	*M*	*SD*
Intercultural Communication Behavior Self-Efficacy (ICSE) Scale N=60	***0.947***					
In a conversation, I can refer to cultural topics or historical events.		*.828*	*3.44*	*0.947*	*3.73*	*0.917*
I can adapt my speaking to a range of different cultural settings and still be understood.		*.741*	*3.59*	*0.854*	*3.75*	*0.882*
I can shift appropriately into informal or formal communication styles without thinking.		*.739*	*3.49*	*1.091*	*3.73*	*0.962*
I can discuss cultural issues and subjects, such as history, politics, literature, and the arts.		*.862*	*3.34*	*1.063*	*3.73*	*1.026*
I can reply appropriately when native speakers use cultural references, shades of meaning, and figures of speech.		*.712*	*3.65*	*0.953*	*3.73*	*0.984*
I can appropriately explain literary works, classic films, or political documents.		*.882*	*2.95*	*0.923*	*3.15*	*1.010*

			Pre-test		Post-test	
	KR-20	*Item-Scale r*	*M*	*SD*	*M*	*SD*
I can discuss and explain in depth the culture's traditions, beliefs, and history.		*.705*	*3.35*	*1.003*	*3.54*	*0.898*
I can explain in detail the country's most pressing public issues.		*.777*	*3.31*	*1.065*	*3.52*	*0.922*
I can use appropriate body language (poster, eye contact, and personal space) in the target culture.		*.495*	*4.02*	*0.684*	*4.23*	*0.722*
I can avoid taboos (e.g., certain gestures or topics).		*.595*	*4.08*	*0.768*	*4.33*	*0.663*
I can follow cultural rules for speaking to people of a different age, gender, or status.		*.652*	*3.95*	*0.728*	*4.29*	*0.771*
I can follow cultural politeness rules (for instance, about invitations, gifts, and asking for help.)		*.639*	*4.23*	*0.663*	*4.35*	*0.668*
If I behave incorrectly, I can recognize the problem and fix it.		*.572*	*4.11*	*0.625*	*4.25*	*0.668*
I can address job-related problems in a culturally appropriate way.		*.717*	*3.44*	*0.969*	*3.77*	*0.857*
I can handle unfamiliar cultural situations appropriately, even if they involve tense or emotionally-charged topics.		*.627*	*3.89*	*0.764*	*3.88*	*0.815*
I can change my gestures to fit different situations that are not familiar to me.		*.761*	*3.71*	*0.750*	*3.89*	*0.729*
I can repair any cultural errors before they become a problem.		*.726*	*3.83*	*0.834*	*4.02*	*0.838*
I can use words and body language to show empathy.		*.589*	*4.18*	*0.587*	*4.29*	*0.798*
I hardly ever make a cultural error anymore.		*.580*	*3.54*	*0.913*	*3.73*	*0.962*

Appendix E

Time Needed for a Given Proficiency Level

Table E1. Foreign Language Program Characteristics and Likely Exit Proficiency of Motivated Students

(Based on DLI Data (Clifford, 2010), adapted by Oxford. NOTE: This table does not show proficiency levels 4 and 5.)
Directions: In each row, find the column that best describes the characteristics of the program you are considering. The level at the top of that column represents the maximum proficiency level attainable in that program by *motivated students with above average aptitude.* Note: It is important to remember that a program's characteristics serve as both *enabling and blocking factors*. Thus, a low rating on a single characteristic may block learners from progressing beyond the level assigned for that characteristic -- even if the ratings for other factors would theoretically enable a higher level proficiency.© Dr. Ray Clifford, 14 July 2010 (used with permission)

Program Characteristic (Enabling and Blocking Factors)	**Maximum Proficiency Level Attainable by Motivated Students**			
	ILR: 0+ ILR: Memorized/ Survival Proficiency ACTFL: Novice	**ILR: 1 ILR: Minimal Functional/Elementary Proficiency ACTFL: Intermediate**	**ILR: 2 ILR: Limited Working/ Functional Proficiency ACTFL: Advanced**	**ILR: 3 ILR: General Functional/ Professional Proficiency ACTFL: Superior**
Topics and cultural content taught	Typical survival topics and situations	Everyday, high frequency daily activities	News, current events, general work activities, and popular media	Societal, political, academic, and complex issues, plus history and the fine arts
Typical utterances presented/taught	≈ 200 words and phrases	Simple sentences & ≈ 1,000 word families	Multiple paragraphs &≈ 5,000 word families	Multiple pages & ≈ 25,000 word families
Supporting knowledge	Word definitions	The grammar, vocabulary, pronunciation, orthography, and culture needed to understand and produce sentences	The grammar, vocabulary, pronunciation, orthography, and culture needed to accurately understand and produce extended discourse such as news reports and broadcasts	The grammar, vocabulary, pronunciation, orthography, and culture needed to accurately understand and produce formal speeches, essays, editorials, and other professional communications
Teaching method	Presentation and drills	Open-ended questions and answers	Use of authentic publications and media in discussions	Practice using foreign language in real-world formal and informal settings
Learner activities and communication tasks	Repeat model utterances	Create sentences and responses to meet one's needs within familiar settings	Relate lengthy stories in past, present, and future time frames & give extensive descriptions of people, places, and things	Accurately defend opinions, explain complex relationships, and discuss abstract topics with precision and depth
Level of learning required	Direct application of memorized knowledge	Near transfer of one's language skill repertoire in familiar settings	Analysis of communication needs and far transfer of language skills to solve real-world problems	Synthesis of professional and linguistic expertise for the far transfer of language skills in professional communications
Type of feedback provided to the learner	Right or wrong	Right or wrong plus feedback about the nature of the error	Multi-level, scaled assessment judgments with tailored feedback	Individualized diagnosis and personalized learning plans for remediation and enhancement
Type of progress checks and tests	Multiple choice, matching, fill-in-the blank, etc	Short answer responses	Open-ended answers, constructed responses, and role play scenarios	Respond to a wide range of real-world communication tasks, e.g. presentations, essays, debates, speeches, etc.
Hours of balanced instruction and study time required to attain automaticity	Cat I: 200 hours Cat II: 250 hours Cat III: 400 hours Cat IV: 600 hours	Cat I: 400 hours Cat II: 500 hours Cat III: 800 hours Cat IV: 1,200 hours	Cat I: 800 hours Cat II: 1,000 hours Cat III: 1,600 hours Cat IV: 2,400 hours	Cat I: 1,600 hours Cat II: 2,000 hours Cat III: 3,200 hours Cat IV: 4,800 hours

Based on the DLI figures in Table E1 above, Table E2 summarizes the number of hours (class and additional study hours combined) needed to progress from one proficiency level to the next.

Table E2. Time Needed to Progress in Proficiency

Based on DLI Figures in Table C1 (Rebecca Oxford). NOTE: This table does not show proficiency levels 4 and 5.

Category of Language Based on Difficulty	Time in hours to move from one level to the next			
	From 0 to 0+ ("a")	From 0+ to 1 ("b")	From 1 to 2 ("c")	From 2 to 3 ("d")
Cat. I	200	200 more	400 more	800 more
Cat. II	250	250 more	500 more	1,000 more
Cat. III	400	400 more	800 more	1,600 more
Cat. IV	600	600 more	1,200 more	2,400 more

The amount of time that it takes for an individual to move up on the ILR scale depends on: (a) initial starting point, (b) intrinsic motivation, (c) extrinsic motivation (i.e., reward, such as language pay), (d) language aptitude, (e) learning strategies employed (are they appropriate and used well?), (f) time on task, (g) nature of learning support (classroom, e-mentor, etc.), (h) willingness to communicate in the foreign language, (i) attitude toward the language and culture, and (j) other factors.