DRAFT

FRAMEWORK FOR THE VERY LOW RANGE DEFENSE LANGUAGE PROFICIENCY TESTS

(VLR DLPT)

March 2010

Version 5

NOTE for version 5: Changes from versions 3 and 4 include the correction of typos and some clarifications and substantive changes, all of which are in response to stakeholder comments as conveyed by DLIFLC. All clarifications and substantive changes from version 3 are highlighted.

– W. Lorié of EduMetrica, Inc.

TABLE OF CONTENTS

1.0 PURPOSE AND SCOPE OF THE VLR DLPT FRAMEWORK

The Very Low Range Defense Language Proficiency Test (VLR DLPT) is part of the Defense Language Proficiency Testing System 5 (DLPT5), which is designed to assess the global language proficiency in reading and listening of native speakers of English who have learned a foreign language as a second language, and speakers of other languages with very strong English skills.

The purpose of this document is to provide stakeholders and prospective test users with information regarding the rationale for the VLR DLPT, its intended users and uses, the inferences to be made on the basis of test scores, the interpretation and operationalization of Interagency Language Roundtable (ILR) levels covered in the VLR DLPT, the implications of the VLR DLPT rationale and construct definition for test design, and the research required for supporting the intended VLR DLPT score-based inferences and uses.

Section 2 provides an overview of the VLR DLPT, in the context of the existing DLPT5. The rationale for the VLR testing program, and anticipated users and uses, are outlined. Section 2 presents a list of stakeholder groups and addresses stakeholder benefits and concerns, as well as the intended consequences of test use. This section lays the groundwork for the interpretation and operationalization of VLR DLPT-relevant ILR levels (in Section 3).

The final section of the VLR DLPT framework draws upon a test-validation framework to propose a research agenda for the VLR DLPT testing program. The agenda outlines the investigations needed to support specific anticipated test uses and in particular the VLR-score-based inferences on which such decisions are based. Section 4 will address research that will have been completed to support test uses and research that will likely be needed.

EduMetrica has developed this document at the request of the Test Development Division at the Defense Language Institute Foreign Language Center (DLIFLC). Draft versions of this document are forwarded to the DLIFLC for its subsequent distribution to stakeholder groups for review.

2.0 OVERVIEW OF THE VLR DLPT

The VLR DLPT is an extension of the DLTPT5 testing system, designed as a measure of language proficiency in the reading and listening skills of foreign language learners, as set forth in the Interagency Language Roundtable (ILR) descriptors of language proficiency for levels 0+, 1, and 1+.

Like the DLPT5,[1] the VLR DLPT will be used to document and make operational readiness, incentive and pay, and assignment and training decisions for military and civilian personnel of the United States government.

In response to the current needs of government agencies that make use of language proficiency testing, the VLR DLPT is explicitly designed to make distinctions among personnel with language proficiency of 0+, 1, and 1+. Although the current DLPT5 tests do in fact make distinctions below level 1, the VLR DLPT is explicitly designed to *focus* on distinctions below level 2.

This section of the VLR DLPT framework addresses that set of purposes, the stakeholders of the VLR DLPT and the uses to be made of the test, VLR DLPT score-based decisions, and the intended consequences of test use.

2.1 RATIONALE FOR THE VLR DLPT

Stakeholders have voiced concerns about three specific aspects of the current lower-range DLPT5:

1. The test is too long;
2. There are no items targeted at level 0+; and
3. Discrimination of levels below level 2 may be inadequate.[2]

Of these, test length has been identified as the most important driving factor, with the leadership of the Defense Language Institute Foreign Language Center (DLIFLC) and the Defense Language Office (DLO) have accordingly indicated that the VLR DLPT must have a maximum test length of 1.5 hours for testing examinees with a target level of 1 or 1+.[3] This is half the test length of the current DLPT5 test. Adaptive testing is planned for the future, but it is not a capacity of the current DLPT test delivery platform.

With respect to the second area of concern listed above, we note that, according to the DLIFLC and the DLO, at the time of DLPT5 design, there was no documented need to accurately classify examinees into 0+ or levels surrounding 0+. Since that time, however, a need for low-level language proficiency testing has emerged.

In a small-scale survey of senior command language program managers administered in late 2007, sixteen respondents (eleven military and five civilian) were asked questions relevant to low-level testing

---

[1] Defense Language Institute Foreign Language Center (DLIFLC). Defense Language Proficiency Testing System 5 Framework (DLPT5-F). p.3.
[2] Defense Language Institute Foreign Language Center ATFL, Evaluation and Standardization (ATFL-ES). Concept of Operations, Very Low Range Defense Language Proficiency Test (DLPT-VLR), Meeting Field Needs for the Lowest Proficiency Ranges, version 1.1 (CONOPS-VLR 1.1). 22 May 2009. p.2.
[3] CONOPS-VLR 1.1. p.2.

needs across government agencies.[4] Fifteen of the respondents indicated that they use the DLPT5, and twelve of those surveyed stated that the test does not meet their needs.[5] Half of those surveyed stated that the test does not meet their needs because it does not identify low levels, and only four said that level 0+ was not needed for any of the major Department of Defense (DoD) purposes of language testing.[6] Thus, at least among a segment of DoD stakeholders, there is a documented need for assessment at level 0+ which cannot be satisfied by the current DLPT5 tests, and the proficiency language testing needs in 2009 and beyond are different from when the DLPT5 system was designed.

With additional input from stakeholders, the current needs were defined to include the following:

1. Provide robust measurement at ILR levels 1 and 1+;
2. Provide 0+ items for "context and warm-up";
3. Allow for the use of test scores for foreign language proficiency bonus (FLPB) pay; and
4. Allow for an administration time of 90 minutes or less.[7]

Subsequently, point 2 was further refined by the DLIFLC and the DLO. The emerging concept regarding 0+ items is that their inclusion in the VLR DLPT will also serve to identify test takers whose language proficiency in the target language is below level 1, but above 0. That is to say, the VLR DLPT should accurately classify an individual as 0+, if this is indeed their true target language proficiency level.

Nine languages – Arabic (Iraqi), Persian-Farsi, Urdu, French, Spanish, Korean, Chinese (Mandarin), Pashto, and Dari – have been prioritized; the development, field testing, and standard setting for these have been funded.[8]

The rationale and justification for developing VLR DLPT tests can be fairly well documented. The logical next step is to consider how to operationalize the ILR levels for these tests. This is especially the case for level 0+ as models for level 1 and 1+ exist in the current DLPT5.

Inclusion of level 0+ items in a DLPT5 test adds a new dimension to the DLPT5 construct. The DLIFLC has observed that since ILR level 0+ consists essentially of "memorized words and phrases, at that level the ILR is less a "proficiency" scale and more an "achievement" scale.[9]

What separates the level 0+ from level 1 is the ability, at level 1, to understand basic grammatical relations in the receptive skills of listening and reading. For example, the ILR description for level 0+ reading states that individuals at this level are able to "read […] numbers, isolated words and phrases, personal and place names, street signs, office and shop designations"[10] but cannot "read *connected*

---

[4] DLIFLC. Survey on Low-Level Testing Descriptive Analysis (SLLTDA).
[5] SLLTDA. p.3.
[6] SLLTDA. pp.3,5. The uses include determining eligibility for hiring, pay, job or mission assignment, completion of training, and further training.
[7] CONOPS-VLR 1.1. p.5. Note that the original rationale for 90 minutes or less arose from the desire to complete both listening and reading tests in a single half-day (morning or afternoon). Stakeholders have been consistent in requesting as short an administration time as possible.
[8] See Solicitation W9124N-09-R-0060 and Contract W9124N-09-C-0072.
[9] CONOPS-VLR 1.1. p.3.
[10] Interagency Language Roundtable Language Skill Level Description: Reading. Cited in DLPT5-F. p.91.

prose"[11] [emphasis added]. In contrast, the ILR description for level 1 reading begins "Sufficient comprehension to read very simple *connected* written material"[12] [emphasis added] and continues to note that level 1 individuals are able to read and understand "known language elements that have been *recombined* in *new* ways to achieve *different* meanings at a similar level of simplicity"[13] [emphasis added]. This notes a level of understanding of the grammatical relations of a language not found in level 0+.

The distinction between these two terms "achievement" and "proficiency" has been drawn in several different ways, none of them definitive. Broadly, and when limiting the discussion to *language* proficiency and *language* achievement, the difference between proficiency and achievement is that the first designates an underlying or general ability to understand and use written and spoken language, while the second refers more to a set of pre-specified skills, knowledge, or abilities that have been acquired in a prescribed course of instruction, self-directed or otherwise.

Because it is not always possible to distinguish usefully when language arises from inherent general language-generation ability or as a consequence of specific language instruction properly applied, the proficiency / achievement distinction is best viewed as one of degree rather than category. However, it can also be noted that language ability at a low level does typically reflect an authentic need for language use. The beginning adult learner may have a purpose for learning a language or has interacted with the language in the course of daily activities.

A second important distinction is the distinction between proficiency and performance testing. A given learner's purpose for learning may be that he or she needs to perform a particular task, such as manning a checkpoint or building rapport with local people. Particular tasks may require only a limited range of content vocabulary, and a performance test would focus on the content of interest. A proficiency test, on the other hand, as it attempts to measure general ability, will typically cover a wider range of content than a performance test will, and will typically fail to cover specialized vocabulary that may be important for particular performance situations. Thus, stakeholders who need to assess the ability of personnel to perform specific language tasks are best served by a performance test tailored to their tasks rather than a general proficiency test.

The *range* of language settings in which 0+ "proficiency" applies is clearly more limited than for the higher levels, but like the higher levels it is also delineated as a practical ~~assessment~~ testing matter. Thus, as we move from the definition of the levels per ILR to the operationalization of the levels for the purposes of language testing ~~assessment~~, the DLPT5 framework and item development guidelines delineate specific "final learning objective" (FLO) language settings, such as military and security; economic and political; scientific and technological; cultural and social; and geography: physical,

---

[11] Interagency Language Roundtable Language Skill Level Description: Reading. Cited in DLPT5-F. p.91.
[12] Interagency Language Roundtable Language Skill Level Description: Reading. Cited in DLPT5-F. p.91.
[13] Interagency Language Roundtable Language Skill Level Description: Reading. Cited in DLPT5-F. p.91.

political, and economic.[14] The acceptable forms of passages and language settings within these FLOs likewise are explicitly more limited than for higher levels.

Words and phrases are fundamental to level 0+ in a way that they are not for the higher levels. To be sure, as the ILR level increases, expectations with respect to the mastery of increasingly sophisticated or uncommon vocabulary also increases. But, at level 0+, not only are vocabulary expectations the lowest among the levels, they are also much more delineated. If words and phrases are memorized, they are memorized from a list, within a learning setting, or as a consequence of limited exposure to the language. Thus, level 0+ is more closely tied to a specific lexicon than the other levels are. Specific words and phrases play a useful role in defining it, and in defining acceptable opportunities for assessing it. In this way, lists of words and phrases are important in operationalizing this level of the VLR DLPT. Precisely *which* lists (and why *those* lists) is discussed in Section 3.

In the case of 0+, the ILR source documents clearly single this level out as significantly different than the rest. The paradoxically-termed "Memorized Proficiency" level states that an individual with 0+ "proficiency" in reading, for example, can recognize printed letters, high frequency words, and numbers, and can read (and presumably understand) isolated words and phrases.

The anticipated uses and users of the VLR DLPT affect the choice of situations and vocabulary that become part of it. This aspect of defining the test framework is most important when considering the relationship between *Headstart*, a language learning program developed by the DLIFLC and intended to prepare individuals in the DoD for deployment in foreign language environments, and the VLR DLPT. (The relationship between the VLR DLPT and *Headstart* is discussed in greater detail in Section 3.2, below.)

2.2 STAKEHOLDERS -- POLICY MAKERS, DESIGNERS, AND TEST USERS

The stakeholders of the VLR DLPT are those individuals who are significantly affected, directly or indirectly, by one or more aspect of the conceptualization, design, implementation, and use of the VLR DLPT. Stakeholders can be divided into three main groups: policy makers, designers, and test users.

Policy makers propose and put in place the policies that should be adhered to with respect to the use of the VLR DLPT. Policy makers operate throughout the DoD; their interest is that the VLR DLPT is designed according to identified needs and that the resulting instrument indeed fulfills those needs. The identified overall need is to accurately ascertain when individuals are at certain levels of proficiency (specifically, 0+, 1, and 1+) with respect to key foreign languages, for purposes related to placement, promotion, or pay decisions in matters related to proficiency in the target languages. Subsequent decisions based on the achievement of these levels are also the domain of the policy makers.

The designers include key individuals in the DLIFLC who, through contact with other stakeholder personnel throughout the DoD, have put forward a process for funding and creating the VLR DLPT with this Framework as the documented point of departure. The designers attend to (1) the process for input

---

[14] DLPT5-F. pp.5-6.

into the instrument's conceptualization, (2) parameters defining its design (for example, its definition as being part of the DLPT5), (3) the management and guidance of its development, and (4) the assurance that the instrument will support all intended uses and interpretations. Designers also oversee all aspects of technical quality and adherence to industry standards.

Test users make up the last VLR DLPT stakeholder group. Test users encompass two distinct groups – those who take the test and those responsible for making inferences or decisions based on individual test scores. The interests of the second group are served with proper guidance from the policy makers on the appropriate application of the VLR DLPT within the framework of test use policies.

Those who take the test are stakeholders for a variety of obvious reasons, not the least of which is that the results of the VLR DLPT potentially affects the career course and pay of individuals in the DoD, and in the case of a score lower than expected, there is an unfavorable effect on career for a significant period of time, since (like the DLPT5) the test cannot be retaken for six months, unless there has been a significant qualifying learning event.

2.3 TEST DESIGN

The VLR DLPT is intended to serve military personnel and civilian employees of the DoD not currently served by the DLPT5, and is an additional component of the DoD support for "language-enabled" personnel. The DLIFLC has determined the design of the VLR DLPT based on its data collection surveys and the policies and decision-making procedures already in place for the DLPT5.

Design decisions:
1. VLR DLPT will test ILR levels 0+, 1, 1+;
2. Test length will be 90 minutes or less per skill (Reading, Listening);
3. The VLR DLPT will be delivered on the existing DLPT web platform.

With respect to test length, it bears mentioning that generally speaking, longer tests are more reliable. When the stakes attached to a test score are high, most test designers prefer a longer test. There is always a tradeoff between longer, more accurate tests, and shorter, more practical ones. The DLIFLC will explore the impact of test length on feasibility of administration and reliability.[15]

Point 3 has implications for the test format: The test will be text- and audio-based. No use can be made at this time of formats involving video or heavy use of graphics. While the platform currently has these limitations, it provides a stable and proven environment in which to deliver the VLR DLPT. Nevertheless, stakeholders wish to acknowledge that the existing DLPT web platform does not allow for graphics and animation.[16] This is relevant to the discussion of what can be tested because at low levels, the ILR for reading makes reference to graphical context.

Finding time and space to administer is also an issue influencing the decision of what platform to use.[17]

---

[15] Email communication from DLIFLC to EduMetrica. March 19, 2010.
[16] Email communication from DLIFLC to EduMetrica. March 19, 2010.
[17] Email communication from DLIFLC to EduMetrica. March 19, 2010.

Finally, as a DLPT, the VLR DLPT will follow a format where orientation statements, question stems, and questions are all in English. The relevant assumption is that a requirement of the examinee population is proficiency in English at the 8th grade level. As stakeholders have notes, this assumption is appropriate because this level of English is a requirement for enlistment.[18]

2.4 TEST USES

The purpose of this section is to mention *potential* uses of the VLR DLPT, as discussed by VLR DLPT stakeholders. These potential uses do indeed suggest policy, but the purpose in mentioning them in this document is not to suggest policy. Rather, it is to outline and reason through the requirements for supporting such uses and to discuss possible consequences of such uses.

Several potential uses for the VLR DLPT have been outlined in the CONOPS-VLR document of May 2009. Two ~~of these~~ uses frequently mentioned in source documents and in discussions between the contractor and government are:

1. For Foreign Language Proficiency Bonus (FLPB);
2. To establish eligibility for job or mission assignment.

It is worth noting that in the 2007 survey of command language program managers on the need for low-level testing, only two of the 16 respondents stated that eligibility for FLPB was relevant at level 1 or 1+. No respondent stated that performance at level 0+ granted or should grant eligibility for bonus pay.[19] Because it has implications for measurement accuracy requirements and incorporating the relative cost of false positive and false negative decisions in setting the 0+ cut score, it is important to determine whether and for whom there will be FLPB for level 0+. This decision has implications for the design of the operational test.

With respect to job or mission assignment, five of those surveyed stated that at 0+ there could be eligibility for certain job or mission assignments. Seven respondents said the same of levels 1 and 1+.[20] From a test validation perspective, a test is appropriate for placement if it measures knowledge, skills, or abilities that are considered necessary for performing well in the new position. But the often-mentioned test use principle of not relying too heavily on one instrument or criterion for making important selection decisions bears repeating here. The principle is especially relevant when "success" in the setting to which a candidate is to be selected is much broader than what available instruments measure.

In addition to being an extension of the current DLPT5 to lower levels, the VLR DLPT thus addresses a new set of needs and users, opening the door to a refinement of the test construct. However, the range of uses and stakes associated with decisions based on VLR DLPT test scores are about the same as those for the DLPT5. The implication for design here is that in several aspects – nature and design of the items, relation to the ILR, and mode of administration – the VLR DLPT will be similar to the DLPT5. But with

---

[18] Email communication from DLIFLC to EduMetrica. March 19, 2010.
[19] CONOPS-VLR 1.1. p.5.
[20] CONOPS-VLR 1.1. p.5.

respect to the range of language situations and (obviously) the focal levels, the VLR DLPT will address a different target audience and level of proficiency than does the DLPT5.

2.5 INTENDED CONSEQUENCES OF TEST USE

Selection and placement testing programs are instituted in order to achieve specific policy aims. In the case of programs such as the DLPT5, those aims include rewarding individuals who demonstrate skills that are of primary importance to the sponsoring organization, and staffing operations with qualified personnel. The broad intended consequence of test use is a more effective and efficient organization.

For the VLR DLPT in particular, the intended consequences are based on the premise that test takers who are truly proficient in one of the receptive skills for a particular language at level 0+, 1, or 1+ or above, are correctly identified into 0+, 1, or 1+ or above, with a level of classification accuracy roughly equivalent to the DLPT5.

Beyond this, one consequence of test use is that open positions requiring a specific level of low-range language proficiency will be filled, given sufficient qualified test takers. Another potential consequence is that there will be a greater number of individuals obtaining FLPB, especially in high-demand languages, and a greater number of promotions, given sufficient open positions requiring proficiency levels 0+, 1, or 1+ or above. It is reasonable also to expect that, for positions where language qualifications tests were not given because they were not available, the performance of staff in those positions will improve as qualified personnel move into or out of them.

Because of the high stakes associated with this test, and its relative dependence on circumscribed vocabulary, text types, and grammatical structures, the VLR DLPT will benefit from a frequent refresh rate.

Research studies can be conducted, once the test has been administered "live," to help answer questions such as, "Are we actually measuring what we intended to measure?" and "Does the test bring about the intended effects in the system overall?"

3.0 RELATION OF VLR DLPT TO DLPT5 AND ILR LEVELS 0+, 1, AND 1+

The VLR DLPT is based on ILR levels 0+, 1 and 1+, and on the DLPT5 framework in general. This decision has both benefits and implications for test design and test development. These benefits and implications are detailed below.

3.1 THE VLR DLPT AND THE DLPT5 SYSTEM

Because the VLR is contained within the DLPT5 system, to some extent the test content parameters are pre-defined. The content of reading and listening passages is determined in reference to the FLOs, Child's text typology, and text typologies relevant to level 0+, which Clifford calls "formulaic" and Lowe "enumerative."[21] Thus, this VLR DLPT framework assumes the theoretical and strategic elements of the DLPT5 system already addressed in the larger DLPT5 framework.

In brief, the VLR DLPT will address the formulaic-enumerative, orientation and instructive text modes. A decision was made to make the delivery platform for the VLR DLPT the same as the DLPT5 delivery platform, which is both proven, and readily available, and is considered by the DLIFLC to meet the essential delivery requirements for the VLR DLPT. Items will follow DLPT format, i.e., multiple choice, with four options, one of which is the key. The tasks to which candidates will respond will have the following features:

1. The task is passage-dependent. Comprehension of the content is a prerequisite for answering the question correctly. A near-exception is level 0+, where identification of isolated words or phrases (and knowledge of their meanings in English) can be sufficient to answer a question correctly.
2. In the case of listening, the totality of the tasks cannot be answered if the candidate has not established a certain degree of automaticity in listening competency.
3. The task will require the candidate to provide some evidence of comprehension.
4. Passages and tasks will be based on the ILR levels. (Passages or tasks that cannot be classified into an ILR level are not part of the test domain.)

3.2 APPLICATION OF THE ILR IN VLR DLPT ITEM DEVELOPMENT: GENERAL PRINCIPLES

The ILR levels are applied in the VLR DLPT item development process principally by attending to three aspects: the FLOs, the text modes, as defined by Child *et al*.[22], and by a determination of "memorized proficiency" as addressed by the ILR. These essential elements will guide test developers at all times, and serve as a basis for ensuring the content-related validity of the resulting tests.

---

[21] VLR-PO.

[22] See Child, J.R. (1987) Language proficiency levels and the typology of texts. In H. Byrnes and M. Canale (eds.), *Defining and developing proficiency: Guidelines, implementations and concepts*. Lincolnwood, IL: National Textbook Co.; Child, J.R. (1998). Language skill levels, textual modes, and the rating process. *Foreign Language Annals*, 31(3), 381-391; Child, J.R. (1999) Analysis of texts and critique of judgment. Retrieved on January 25, 2010 from http://dspace.wrlc.org/bitstream/1961/3455/9/gurt_1999_08.pdf

3.2.1 FINAL LEARNING OBJECTIVES

As with the DLPT5, the FLOs, outline the topic areas from which content for the VLR DLPT is derived. As with the DLPT5, in the VLR DLPT there is no strict guideline for the range of coverage across the FLOs, and sub-categories under each of the FLOs vary from level to level. The FLOs for the VLR DLPT are: Military and security issues; Economic and political matters; Science and technology; Cultural and social issues; and Geography: physical, political, and economic.

In the listening test for VLR DLPT, the FLOs contribute to the test's situational authenticity, defined as the degree to which the test requires processing of samples of realistic, spoken language, automatically, and in real time. The FLOs contribute to the test's interactional authenticity in that they ensure that it assesses knowledge, skills, and abilities critical to target language-use situations.

3.2.2 TEXT TYPOLOGY MODES

Child's text typology is a scale for classifying the purpose of written and spoken texts. It describes texts as ranging in complexity from being simple and orientational in nature to sophisticated and projective. All four of Child's text modes are represented in the DLPT5.

With the introduction of 0+ in the VLR DLPT, however, another mode referenced by Clifford and Lowe, called the "formulaic" or "enumerative" mode[23], enters the test framework. Also, because the range of proficiency assessed by VLR DLPT is by definition more limited and focused than that assessed by the DLPT5, only the first two of Child's modes are relevant to the test. Thus, passages on the VLR DLPT can be classified in one three modes: "Formulaic-Enumerative", "Orientational", and "Instructive". These text modes align roughly to the descriptors of the ILR levels for 0+, 1 and 1+, and provide additional guidelines to test developers with respect to the types of passages selected for inclusion on the test.

3.2.3 VLR DLPT, "MEMORIZED PROFICIENCY", AND RELATION TO *HEADSTART*

Because ILR level 0+ refers to memorized proficiency, a determination must be made as to what "memorized proficiency" implies for the VLR DLPT. To some extent this is defined in the ILR levels and one can assume that "common words and phrases" are those typically learned in beginning language courses. However, introductory language courses can vary, depending on the needs of the target audience. The "common words and phrases" learned in an introductory course in French for business professionals will differ from words and phrases learned in a course in French for travelers. More to the point, assessments of language mastery at the end of either of these courses will be markedly different.

---

[23] VLR-PO.

The VLR DLPT inherits the domain of language proficiency defined by the DLPT5 as reflected in the FLOs, particularly the Military and Security category. In addition, ~~it has been determined that~~ the "essential content domain" for the VLR DLPT is to be *Headstart*,[24] a web-based language-learning program developed by the DLIFLC, especially with respect to the "memorized proficiency" of level 0+, and the understanding of connected prose at Levels 1 and 1+. The following outlines the position taken by this framework with respect to the relationship between the VLR DLPT and *Headstart*.

1. Although not strictly defined, there is a domain of common words and phrases particular to the VLR DLPT.
2. *Headstart*, a language-learning program geared to address the learning needs of DoD individuals, draws much of its content from this domain.
3. Assessment of that domain is salient for the VLR DLPT because assessment of knowledge and understanding of isolated words of phrases is a focus at the lower levels of the ILR, particularly 0+.
4. The VLR DLPT, however, is NOT a direct assessment of knowledge of the words and phrases present in any *Headstart* program. Instead, the VLR DLPT and *Headstart* all attempt to reference a common domain of language situations, acceptable texts, and a general conception of "common words and phrases".
5. In developing VLR DLPT items, lists of *Headstart* words and phrases in English are consulted to understand the nature of the language domain to which both DLPT5 and *Headstart* refer. (Bilingual lists have not been made available as of this writing; therefore there is no explicit, exact correspondence provided in the VLR passages.)

Thus, although *Headstart* is a learning program geared toward DoD-relevant language proficiency, the VLR DLPT is not designed to measure *Headstart* performance only.

The above positions imply the following:

1. All else being equal, a group of candidates going through a *Headstart* program are predicted to do better on the VLR DLPT than a group going through a hypothetical equally rigorous learning program the language domain of which, as reflected in its conceptualization of common words and phrases, is different from that of the DLPT5 and *Headstart*.
2. No candidate going through a *Headstart* learning program can be guaranteed any particular score, but groups taking *Headstart* should generally score favorably on sections of the VLR DLPT whose tasks correlate most closely with *Headstart*, namely the Military and Security FLO, and the test of listening skills.

---

[24] <mark>W9124N-09-C-0072. C.3.1.3.1.: "DLIFLC will provide DLIFLC-produced *Headstart* materials that will be the essential content domain of the tests." [Note from contractor to government: A stakeholder wishes to know the originator of this decision, its rationale, and whether stakeholders were involved. The contractor has insufficient background information to answer this, and will defer to government for guidance on this matter.]</mark>

3.3 APPLICATION OF THE ILR IN VLR DLPT: LEVEL AND SKILL SPECIFICS

A typical testing system is based on a flow of events that begins with setting goals, designing an instrument to measure attainment of those goals, capturing performance on the instrument, reporting and interpreting the results, and making decisions. For the VLR DLPT, proper application of the ILR underlies this process. Adherence to the ILR ensures alignment between the test content, test taker performance, interpretations of test scores, and decisions based on test scores.

This adherence to alignment permits us to make several claims about the VLR DLPT that support test validation, addressed more fully in Section 4. The claims include:

1. The test and its passages and items are aligned to the ILR.
2. The performance of the test taker on a given task can be referenced to the ILR.
3. The resulting score is based on performance on tasks defined by the ILR.
4. Resulting scores reference ILR levels in a defensible manner.
5. Decisions taken as a result of VLR DLPT scores are based on inferences of performance described in the ILR levels, together with an argument that those inferences justify the decisions taken.

In order to make a decision based on a given test instrument, the test needs to distinguish reliably among levels of proficiency. The use of the ILR throughout the test development process allows for the ability to distinguish among four levels of language proficiency – below 0+, 0+, 1, and 1+ or above – defined by the three cut scores – one at 0+, another at 1, and another at 1+.

At level 0+, for example, we can associate the ILR descriptors with the test taker performance on multiple choice tasks because those tasks have been developed in alignment with text type descriptors, content area descriptors, and clearly defined eligible tasks appropriate to that level. Thus we can assign a description for test takers who achieve a rating of 0+, in terms of what they can or cannot do with regard to language comprehension in the target language. This same reasoning can be applied to each of the other two levels covered by the VLR DLPT. Those descriptors are already defined by the ILR.

The unique ILR descriptors for reading and listening at levels 0+, 1, and 1+ are described below, in terms of their contribution as descriptors, and their interaction with the FLOs and Child's text typology. One can look at this interaction as a process of refinement: the text typology defines the type of passage and its characteristics that are appropriate to the level in question; the FLOs then determine the language setting in which the texts will occur.

3.3.1 READING COMPREHENSION LEVEL 0+

ILR Descriptor. R-0+: Reading 0+ (Memorized Proficiency). Can recognize all the letters in the printed version of an alphabetic system and high-frequency elements of a syllabary or a character system. Able to read some or all of the following: numbers, isolated words and phrases, personal and place names,

street signs, office and shop designations. The above often interpreted inaccurately. Unable to read connected prose.[25]

Acceptable Text Modes. Formulaic-Enumerative and Orientational. Texts that have level 1 characteristics can be appropriate for level 0+ because some such texts can include elements important to level 0+ tasks. These elements include common words and phrases, often repeated, and relevant to the main purpose of the text. Also, short level 1 texts may tend toward level 0+. Finally, there are a limited number of sufficiently distinct purely level 0+ texts that can be included in any one test. The general principle of a potential half-level mismatch between text and task is that it is acceptable for a slightly less complex task to be asked of a slightly more complex text, and a slightly more complex task can be associated with a slightly less complex text. Without adoption of this principle, texts and tasks must be strictly related in such a way that places too great a restriction on the test domain. Also, without this principle it would be very difficult to associate more than one question with a passage, rendering the assessment less authentic.

At level 0+ for reading, VLR DLPT texts will tend to exhibit the following properties:

- Linguistic Characteristics
  - Texts should consist of basic survival vocabulary, focused on areas most likely to be needed by military personnel.
  - Texts should deal with concrete topics and should be context-independent. Texts should not require cultural knowledge.
  - The organization of texts should be highly predictable with minimal discourse markers; syntax is usually lists or the simplest sentence types.
- Text types: Formulaic, phatic, and transactional.
  - Examples; Signs, memos, forms, greeting cards, postcards, advertisements.

Acceptable FLO Sub-categories. Examples of appropriate texts aligned to FLOs for VLR DLPT reading level 0+ are indicated in Table 1.

---

[25] Interagency Language Roundtable Language Skill Level Description: Reading. Cited in DLPT5-F. p.91.

TABLE 1. EXAMPLES OF POSSIBLE TOPICS FOR FLOS AT LEVEL 0+

FLO CONTENT AREA	EXAMPLE SUB-CATEGORIES
Military and Security	Police station signs. Traffic signs
Economic and Political	Names of rulers. Announcements of national holidays. Price tags
Scientific and Technological	Hospital signs. Pharmacy signs. Ads for computers, TVs, etc.
Cultural and Social	Family relationships. Religious holidays
Geography: Physical, Political, Economic	Basic terms: ocean, city, mountain, etc.

Note: Table adapted from VLR-GTS, p.10.

3.3.2 READING COMPREHENSION LEVEL 1

ILR Descriptor. R-1: Reading 1 (Elementary Proficiency). Sufficient comprehension to read very simple connected written material in a form equivalent to usual printing or typescript. Can read either representations of familiar formulaic verbal exchanges or simple language containing only the highest frequency structural patterns and vocabulary, including shared international vocabulary items and cognates (when appropriate). Able to read and understand known language elements that have been recombined in new ways to achieve different meanings at a similar level of simplicity. Texts may include descriptions of persons, places or things: and explanations of geography and government such as those simplified for tourists. Some misunderstandings possible on simple texts. Can get some main ideas and locate prominent items of professional significance in more complex texts. Can identify general subject matter in some authentic texts.[26]

Acceptable Text Modes. Orientational and Instructive. (Some instructive is allowed, but will be quite rare for level 1.) At level 1 for reading, VLR DLPT orientational texts will tend to exhibit the following properties:

- Linguistic Characteristics
    o Texts should have a linguistically non-complex syntactic structure (i.e., simple sentences with little embedding).
    o Texts should consist of basic survival vocabulary.
    o Texts should deal with concrete topics and should be context-independent.
    o Texts should not require culture-specific knowledge.
- The organization of the texts should be either:
    o Very loose so that sentences can be rearranged without a loss of the original meaning.
    o Highly predictable and sequential with minimal discourse markers.
- Text types: Formulaic, phatic, and transactional.

---

[26] Interagency Language Roundtable Language Skill Level Description: Reading. Cited in DLPT5-F. p.91.

- o Examples: Signs, forms, personal requests, announcements of events, greeting cards and congratulatory messages, simple forms of personal correspondence, postcards, invitations, emails, flyers, brochures, pamphlets, simple business correspondence.

According to the ILR for this level, the language learner is able to read texts that may include "descriptions of persons, places or things: and explanations of geography and government such as those simplified for tourists," as well as representations of verbal exchanges. This description of the learner's skill permits the limited introduction of the Instructive mode at this level, as well as aligning fully to the orientational mode, which is characterized by connected text, loosely ordered information, highly predictable order of information, and simple sentences, that may occasionally be compound.

Acceptable FLO Sub-categories. Examples of appropriate texts for VLR DLPT reading level 1 are shown in Table 2.

TABLE 2. EXAMPLES OF POSSIBLE TOPICS FOR FLOS AT LEVEL 1

FLO CONTENT AREA	EXAMPLE SUB-CATEGORIES
Military and Security	Military ranks. Basic police subjects (arrests, etc.). Customs officials. Traffic regulations
Economic and Political	Government ministries. National events. Hiring and promotion. Marketplace activities. Basic bank transactions. Basic travel and tourism
Scientific and Technological	Health services (appointments, prescriptions, hospitals etc.). Technological devices of daily life. Simple texts on scientific discoveries / research
Cultural and Social	Family event announcements. Religious holidays. School events. Cultural fairs. Sports. Entertainment. Obituaries
Geography: Physical, Political, Economic	Landmarks and spatial orientation. Weather and climate. Basic geographic relations

Note: Table adapted from VLR-GTS, p.13.

3.3.3 READING COMPREHENSION LEVEL 1+

ILR Descriptor. R-1+: Reading 1+ (Elementary Proficiency, Plus). Sufficient comprehension to understand simple discourse in printed form for informative social purposes. Can read material such as announcements of public events, simple prose containing biographical information or narration of events, and straightforward newspaper headlines. Can guess at unfamiliar vocabulary if highly contextualized, but with difficulty in unfamiliar contexts. Can get some main ideas and locate routine information of professional significance in more complex texts. Can follow essential points of written discussion at an elementary level on topics in his/her special professional field. In commonly taught languages, the individual may not control the structure well. For example, basic grammatical relations are often misinterpreted, and temporal reference may rely primarily on lexical items as time indicators. Has some difficulty with the cohesive factors in discourse, such as matching pronouns with referents. May have to read materials several times for understanding.[27]

Acceptable Text Modes. Orientational and Instructive. (Some instructive is allowed, but will be somewhat uncommon for level 1+.) At level 1+ for reading, the most salient VLR DLPT text characteristics for the orientational and instructive modes are as described for level 1.

Acceptable FLO Sub-categories. Examples of appropriate texts aligned to FLOs for VLR DLPT reading level 1+ are the same as those indicated for level 1 in Table 2. Even though reading 1+ text and

---

[27] Interagency Language Roundtable Language Skill Level Description: Reading. Cited in DLPT5-F. p.91.

item characteristics are similar to those for level 1, texts can be slightly longer and require the processing of a greater amount of text in order to answer questions correctly.[28]

3.3.4 LISTENING COMPREHENSION LEVEL 0+

ILR Descriptor. L-0+: Listening 0+ (Memorized Proficiency). Sufficient comprehension to understand a number of memorized utterances in areas of immediate needs. Slight increase in utterance length understood but requires frequent long pauses between understood phrases and repeated requests on the listener's part for repetition. Understands with reasonable accuracy only when this involves short memorized utterances or formulae. Utterances understood are relatively short in length. Misunderstandings arise due to ignoring or inaccurately hearing sounds or word endings (both inflectional and non-inflectional), distorting the original meaning. Can understand only with difficulty even such people as teachers who are used to speaking with non-native speakers. Can understand best those statements where context strongly supports the utterance's meaning. Gets some main ideas.[29]

Acceptable Text [30] Modes. Formulaic-Enumerative and Orientational. Texts that have level 1 characteristics can be appropriate for level 0+ because some such texts can include elements important to level 0+ tasks. (Please see discussion under 3.3.1.)

At level 0+ for listening, VLR DLPT texts will tend to exhibit the following properties[31]:

- Linguistic characteristics:
  - o Texts should consist of basic survival vocabulary, focused on areas most likely to be needed by military personnel. They are either dialogues or monologues delivered at a normal speech rate or slower.
  - o Texts should deal with concrete topics and should be context-independent. Structure should be at the simplest level. Texts should not require culture-specific knowledge. Sources: Authentic audio media or in-house recordings of exchanges of greetings or very simple conversations focusing on one or two words or phrases. May include non-native speakers.
- Structure: Very simple structures, mostly simple sentences or fragments. Focus is on one or two words or phrases. Exchanges marked by repetition. No embedding, cohesive devices, subjunctive or subordinate clauses. Should not require culture-specific knowledge.

---

[28] VLR-GTS. p.14.
[29] Interagency Language Roundtable Language Skill Level Description: Reading. Cited in DLPT5-F. p.87.
[30] This and subsequent references to "text" with respect to listening refer to the scriptable content of utterances.
[31] Adapted from VLR-GTS. p.15.

Acceptable FLO Sub-categories. Examples of appropriate texts for VLR DLPT listening level 0+ are shown in Table 3.

TABLE 3. EXAMPLES OF POSSIBLE TOPICS FOR FLOS AT LEVEL 0+

FLO CONTENT AREA	EXAMPLE SUB-CATEGORIES
Military and Security	Simple police instructions Basic military commands
Economic and Political	Names of rulers Announcements of national holidays Merchants informing customers of prices
Scientific and Technological	Basic medical instructions Ads for computers, TVs, etc.
Cultural and Social	Family relationships Religious holidays
Geography: Physical, Political, Economic	Basic terms: ocean, city, mountain, etc.

Note: Table adapted from VLR-GTS, p.13.

3.3.5 LISTENING COMPREHENSION LEVEL 1

ILR Descriptor. L-1: Listening 1 (Elementary Proficiency). Sufficient comprehension to understand utterances about basic survival needs and minimum courtesy and travel requirements. In areas of immediate need or on very familiar topics, can understand simple questions and answers, simple statements and very simple face-to-face conversations in a standard dialect. These must often be delivered more clearly than normal at a rate slower than normal with frequent repetitions or paraphrase (that is, by a native used to dealing with foreigners). Once learned, these sentences can be varied for similar level vocabulary and grammar and still be understood. In the majority of utterances, misunderstandings arise due to overlooked or misunderstood syntax and other grammatical clues. Comprehension vocabulary inadequate to understand anything but the most elementary needs. Strong interference from the candidate's native language occurs. Little precision in the information understood owing to the tentative state of passive grammar and lack of vocabulary. Comprehension areas include basic needs such as: meals, lodging, transportation, time and simple directions (including both route instructions and orders from customs officials, policemen, etc.). Understands main ideas.[32]

Acceptable Text Modes. Orientational and Instructive, although texts in the instructive mode[33] texts will be quite rare at level 1. At listening level 1, VLR DLPT texts will tend to exhibit the following properties[34]:

---

[32] Interagency Language Roundtable Language Skill Level Description: Reading. Cited in DLPT5-F. p.88.
[33] *Texts in the instructive mode* provide facts and details about people, places and events. News articles are prime examples. This is not to be confused with *texts that give instruction(s)*.
[34] Adapted from VLR-GTS. p.16.

- Linguistic characteristics:
    - Texts are either dialogues or monologues typically delivered at a rate slightly slower than normal. They may have non-intrusive background noise.
    - Texts have linguistically non-complex structures (i.e., simple sentences with basic word order) and basic survival vocabulary.
    - Some texts will have some 'oral' characteristics, with the linguistic characteristics of spoken language.
- Sources
    - Authentic audio media sources (e.g., TV/radio programs, movies, public announcements, the Internet, language intended for the non-native traveler)
    - Recorded in-house: Everyday social exchanges (greeting, apologizing, congratulating, etc.) and basic daily transactional activities and service encounters.

Acceptable FLO Sub-categories. Examples of appropriate texts for VLR DLPT listening level 1 are shown in Table 4.

TABLE 4. EXAMPLES OF POSSIBLE TOPICS FOR FLOS AT LEVEL 1

FLO CONTENT AREA	EXAMPLE SUB-CATEGORIES
Military and Security	Military ranks, Basic police subjects, customs officials, traffic problems
Economic and Political	Government ministries. National events. Hiring and promotion. Marketplace activities. Basic bank transactions. Basic travel and tourism
Scientific and Technological	Health services (appointments, prescriptions, hospitals, etc.) Technological devices of daily life Simple discussions or announcements of scientific discoveries/research
Cultural and Social	Family event announcements Religious holidays. School events Cultural fairs. Sports. Entertainment
Geography: Physical, Political, Economic	Landmarks and spatial orientation Weather and climate Basic geographic relations

Note: Table adapted from VLR-GTS, p.13.

3.3.6 LISTENING COMPREHENSION LEVEL 1+

ILR Descriptor. L-1: Listening 1+ (Elementary Proficiency, Plus). Sufficient comprehension to understand short conversations about all survival needs and limited social demands. Developing flexibility evident in understanding a range of circumstances beyond immediate survival needs. Shows spontaneity in understanding by speed, although consistency of understanding is uneven. Limited vocabulary range necessitates repetition for understanding. Understands more common time forms and most question forms, some word order patterns, but miscommunication still occurs with more complex patterns. Cannot sustain understanding of coherent structures in longer utterances or in unfamiliar situations.

Understanding of descriptions and the giving of precise information is limited. Aware of basic cohesive features (e.g., pronouns, verb inflections), but many are unreliably understood, especially if less immediate in reference. Understanding is largely limited to a series of short, discrete utterances. Still has to ask for utterances to be repeated. Some ability to understand facts.[35]

Acceptable Text Modes. Orientational and Instructive. (Some instructive <mark>mode</mark> is allowed, but will be somewhat uncommon for level 1+.) At level 1+ for listening, the most salient VLR DLPT text characteristics for the orientational and instructive modes are as described for level 1.

Acceptable FLO Sub-categories. Examples of appropriate texts aligned to FLOs for VLR DLPT listening level 1+ are the same as those indicated for level 1 in Table 2. Even though listening level 1+ test and item characteristics are similar to those for level 1, texts can be slightly longer and require the processing of a greater amount of text in order to answer questions correctly.[36]

3.4 LANGUAGE-SPECIFIC ISSUES FOR VLR DLPT

The ILR, text typologies, and FLOs are largely language-independent schemes. They are frameworks that reflect attempts to summarize properties of texts and levels of skill in understanding and using language, regardless of the language. The diversity of human languages, however, cannot be fully accounted for by such schemes, and the practical implication of this for the development of the VLR DLPT is that designers and developers must attend to language-specific issues.

To a great extent, language-specific issues are more properly elements of item writing guidelines than a test framework. Thus, only the major, construct- or policy-relevant language-specific issues will be addressed here. Special considerations and constraints are outlined here. All decisions, unless otherwise specified, are based on communications with the leadership of, and language-specific managers and directors in, the DLIFLC.

3.4.1 A GENERAL NOTE ON COGNATES AND BORROWED WORDS

Technically, "cognate" refers to two or more words that share the same etymological root. Many times, written cognates look and/or sound similar and mean the same or nearly the same thing. Romance languages and English share many cognates, and an English reader can often guess the meaning of a word or phrase in French, Spanish, Portuguese, or Italian, merely by its spelling and/or sound.

When "cognate" is used in this framework, it refers to words in a foreign language that an English reader or listener can reasonably guess the meaning of (and quickly locate its English counterpart). Examples abound: "policía", "taxi", "autobús", "apartamento", "computadora", etc.

---

[35] Interagency Language Roundtable Language Skill Level Description: Reading. Cited in DLPT5-F. p.88.
[36] VLR-GTS. p.18.

A decision was made to be alert to the presence of cognates in passages, especially for French and Spanish, and to avoid trivializing a question by limiting it to the identification of a salient cognate in the passage. But ~~developers have~~ the DLIFLC has adopted no hard-and-fast rule regarding cognates, as we are aware of two important positions with respect to this issue.

1. If a language contains many cognates, then knowledge of the language necessarily implies knowledge of several cognates, and knowledge of those cognates aids in learning and understanding the language, and
2. To properly assess knowledge of a foreign language, even one with many cognates, special attention must be given to assessing the non-cognate elements in the language, or risk trivializing the assessment.

Thus, the test development contractor has ~~developers have~~ adopted a policy of generally avoiding the most obvious cognates as focal words or phrases for passages and answer keys (that is, those answer choices which designate the correct, credited responses to the test questions). Some may appear, but they will be rare. Cognates will often be present in passages, and may at times offer English readers and listeners clues, but this is unavoidable, and not a major issue unless it becomes the focus of the assessment.

Similarly, many languages make use of words borrowed directly from English. The policy with respect to borrowed words is essentially the same as the policy for cognates.

Other specific issues per language:

Arabic (Iraqi) (spoken only)[37]
- Dialects. Spoken Iraqi consists of more than one dialect. The VLR DLPT test will be limited to the dialect common to Baghdad and its regions.
- Relation to Modern Standard Arabic (MSA). Iraqi is a mixture of formal language (spoken MSA) and informal dialect (Arabic Iraqi). Because the Iraqi VLR DLPT is intended as a test of the latter, efforts will be made to develop passages and items where the latter elements are more salient, and the object of the assessment. However, test takers who know MSA will clearly be at an advantage over those who do not.

Chinese (Mandarin)
- Written systems. In China, readers will generally see the simplified characters for Mandarin. ~~Chinese speakers from other countries may be accustomed to the traditional form.~~ A decision has been made that ~~15% of the~~ all written passages will be in the simplified ~~traditional~~ form at all levels. ~~level 1+. For levels 0+ and 1, only simplified characters will be used.~~

---

[37] There is no reading VLR test for MSA or any dialect of Iraqi at this point.

Dari

- Context and social conventions. Achieving authenticity in Dari passages requires attention to a number of social conventions common to Dari speakers. These conventions are marked by strong gender roles in professions and rules about who can speak to whom, when, and about what.

Korean

- Dialects. A decision was made to use Seoul as the spoken standard. Developers will not include southern regional dialects in listening passages.
- Materials. Contexts will be limited to South Korea.
- Vocabulary. Sino-Korean vocabulary will be avoided. If present in a passage, that vocabulary will not be focus of the question.
- Registers. Informal / spoken forms and neologisms are to be avoided in written passages.

Pashto

- Dialects. In consultation with the leadership of the DLFLC, a decision was made to use the Central/Eastern and/or the Kandahari dialect.
- Context and social conventions. As with Dari, developers attend to rules pertaining to the gender match of speakers in a dialogue, appropriate and inappropriate topics, and women's roles and professions.
- Tenses. The past tense is very challenging in Pashto. Whereas in many languages, mastery of level is independent of tense, this is not the case for Pashto. Pashto developers will make efforts to avoid passages in which understanding the tense of an action is crucial to comprehension and correctly answering test questions.

Persian-Farsi

- Accents. As this is what is commonly heard on the radio and on television, the Tehran accent will be used in listening passages.

Urdu

- Regional variants. A decision was made to use Urdu from Pakistan.

4.0 A VLR DLPT VALIDATION FRAMEWORK

4.1 A DEFINITION OF TEST VALIDATION

Test validation is the process of validating the interpretation and use of a test; validation begins by presenting a clear statement of proposed test-based interpretations and uses and a critical evaluation of these interpretations and uses.[38] For new testing programs, validation has a legitimate confirmationist bias[39], because the focus is primarily on creating the test and an interpretive argument, which "specifies the proposed interpretations and uses of test results by laying out the network of inferences and assumptions leading from the observed performances to the conclusions and decisions based on the performances."[40] For more established programs, a more critical approach is appropriate, in which test-based claims can be challenged explicitly using evidence supporting alternative interpretations of test scores and uses other than or contrary to those of test sponsors.

4.2 INTERPRETIVE ARGUMENTS FOR THE VLR DLPT

The interpretive arguments for hiring decisions, awarding FLPB, and job or mission assignments, are distinct. In general, each of these arguments will contain four broad classes of inferences: 1) Scoring, 2) generalization, 3) extrapolation, and 4) decision.[41] The test instrument development process focuses heavily on ensuring that the first two classes of inferences are well supported. The scoring inference is from an observed performance to an observed score, and it rests on assumptions of the appropriateness of the scoring rule and the consistent application of this rule.[42] In the case of VLR DLPT, these assumptions can be supported through the adoption of item response theory (IRT), because the latter has a strong theoretical foundation supporting the assignment of numerical scale scores to different profiles of test performance. The DLPT5 uses IRT.

The generalization inference is of special relevance in test design. In order to generalize from an observed score to a so-called "universe score", or score over all possible items in the domain, it is essential that the set of tasks to which an examinee is exposed be (1) representative of the item domain and (2) large enough to control any error in "sampling" the domain.[43] The generalization inference for DLPT5s, including the VLR DLPT, is supported by specifications and training materials.[44] These help

---

[38] Kane, Michael T. (2006). "Validation", p.17. In Brennan, Robert. L. (Ed.) *Educational Measurement, 4th Edition*. American Council on Education. pp.17-64.

[39] Kane, Michael T. (2006). p.25.

[40] Kane, Michael T. (2006). p.23.

[41] Kane, Michael T. (2006). p.24.

[42] Kane, Michael T. (2006). p.24.

[43] Kane, Michael T. (2006). p.24. In most test development settings, "sampling" the item domain in any statistical manner is impractical or impossible; instead, careful guidance on task development within the boundaries of test blueprint helps support the generalization inference.

[44] Such as, for example, the DLPT5 System Specifications and Training Documents for New DLPT5 Test Writers of Very Low Range Tests. DLIFLC-TD. 10 October 2009.

operationalize the construct definition of DLPT5 and also specify the domain of DLPT5 items by making it easier to judge, of any particular task (item), whether or not it belongs to the item domain.

Therefore, given that the test development process itself supports the first two arguments, a validation program for the VLR DLPT will focus on the extrapolation and decision inferences. The first of these is the inference from a "universe score" to a level of proficiency. In practical terms, for example, what will require justification is the assumption that the Urdu reading skills of an examinee who receives a score of 1+ in Urdu reading is indeed accurately described by the ILR level 1+ reading performance level description. The decision inferences depend on the test use.

4.3 VLR DLPT SCORES AND SCORE-BASED INFERENCES

VLR DLPT scores will be reported as level designations – the possible scores are "less than 0+", "0+", "1", and "at least 1+". (By definition, the VLR DLPT cannot distinguish between examinees at level 1+ and 2 or above.) The test scale will be set through item response theory (IRT), as is required of all multiple-choice DLPT5s[45] and, unlike the current DLPT5s[46], cut scores will be established through a judgmental[47] standard setting process, by any of several standard methodologies yet to be determined.[48] [Question from contractor to government: A stakeholder expressed the view that the standard setting process should be identified in the framework document. Shall the process for establishing cut scores be decided at this time and identified in this document?]

An examinee designated at level *x* on the VLR DLPT is characterized as indicated in the performance level descriptions of Section 3. This is the test-score-based inference. For any given (valid) use of the test, the decision to be taken on the basis of test scores should be supported by that test-score-based inference.

4.4 VALIDATING DECISIONS BASED ON THE VLR DLPT

For FLPB award, the inference from a level of skill, as described in the performance level descriptors, to a specific monetary award will require justification and support. This inference will be based on assumptions about the relative value of personnel with particular levels of language proficiency, for certain languages. Those assumptions will likely require support/validation from DoD resource managers.

Similarly, if the test use is assignment on a mission, the inference is again from the performance level descriptors, but here it is to the *foreign-language-relevant requirements* of successful performance on the mission. The emphasis is added to exclude other factors of successful performance on the mission which may be unrelated to foreign language proficiency. This highlights why it is considered good

---

[45] CONOPS-VLR 1.1. p.5.

[46] Cut scores for the DLPT5s are determined through mastery level definitions embedded in ILR interpretations (examinees at level *x* have probability 0.7 of correctly responding to randomly-selected level *x* items) and the assumptions about the representativeness of level *x* items on the test. See DLPT5-F pp.103, 105-106.

[47] The term "judgmental" is used here to indicate that standard setting is to be conducted with panelists, who will bring their expert judgment to bear on the standard setting task.

[48] W9124N-09-C-0072. C.3.4.4.7.1 – C.3.4.4.7.3.

practice not to rely solely on one instrument for making important placement decisions. The VLR DLPT will only address the foreign-language-relevant requirements of the mission or job, not its other requirements. Other instruments, measures, or criteria can provide information about those other requirements.

4.5 RESEARCH SUPPORTING TEST USE

The design of the VLR DLPT and validation research should support the assumption that an examinee designated at level *x* is accurately characterized as such by the performance level description of *x* in the ILR. Moreover, the array of specific decision inferences discussed in this section will require initial and continuous support. Such support can be addressed through a validation agenda.

A validation agenda is a research agenda for a test premised on the need for continuously supporting the uses made on the basis of test scores. The agenda begins well before test development, in the design phases, and extends through operational deployment. A validation agenda for VLR DLPT can be organized around two main areas – one focused broadly on the question "Does this test measure what it purports to measure?" and the other addressing the question "Is this use of the test justified?"

4.5.1 VALIDATION RESEARCH AREA A: SUPPORTING GENERALIZATION AND EXTRAPOLATION INFERENCES

The first area of a validation research agenda for the VLR DLPT continually asks and attempts to answer the question "Does the VLR DLPT measure what it purports to measure?" Test specifics aside, this is considered the fundamental question in test validation. As noted earlier, the establishment of clear test development guidelines and documentation that they have been followed provides a good basis for supporting the claim that the VLR DLPT indeed assesses what it purports to assess. But external sources of evidence are also required, especially as the program goes "live," to support this claim. A proposed agenda for this research area would include research into the extent to which:

1. Scores on alternative forms on the VLR DLPT are equivalent[49];
2. The cut scores established through the standard setting process adequately distinguish between individuals who are at the different levels covered by the VLR DLPT, as supported by external evidence;
3. Scores on the VLR DLPT in a specific language skill are appropriately related to other well-established measures of proficiency for that language and skill;
4. Scores on the VLR DLPT are related in expected ways to candidate performance on *Headstart* (See Section 3.2.3.);
5. The pattern of item responses for individuals in a given score level implies a characterization of these individuals' language proficiency which reasonably corresponds to the ILR descriptor for that level.

---

[49] This statement reflects the generally accepted view that if different test forms (specific collections of items) exist for the same test, then any randomly selected form should yield the same hypothetical true score for any given examinee. In other words, the score one obtains on one form should be the same one obtains any other form, except for random measurement error. This "same score" need not be the same *raw* score, as test scores from different forms are often adjusted statistically for differences in difficulty to ensure equivalence of scores. Also, a "different" form is a form that differs from a reference form by at least one item. In practice, pairs of forms normally have many fewer items in common. In high-stakes testing programs where tests can be retaken multiple times, it is recommended that the number of items in common be very few, even zero if possible.

4.5.2 VALIDATION RESEARCH AREA B: SUPPORTING DECISION INFERENCES

The second area of a validation research agenda for the VLR DLPT continually asks and attempts to answer the question "Is this use of the test justified?" The use, of course, can vary, but the focus will likely be making a decision about job or mission assignment, based on the assumption that the particular VLR DLPT given establishes eligibility for the job or mission assignment in question. A secondary focus for decision inference validation support is likely to be the awarding of FLPB as a result of test scores. A proposed agenda for this research area would include research on the extent to which:

1. The language proficiency of those who have been hired as a result of VLR DLPT scores is significantly better than those who have been denied a position as basis of VLR DLPT scores.
2. The language proficiency of those who have received FLPB for a given language, skill, and ILR level, is significantly better than those who are both intending to apply for FLPB and are in an introductory language program;
3. The estimated rate of false positives and false negatives is acceptable for FLPB award and job and mission assignment decisions[50];
4. The introduction of the VLR DLPT has brought about intended consequences without also bringing about unintended consequences.

---

[50] There is a tradeoff, for example, between reducing the number of truly unqualified people ultimately selected for a position, and reducing the number truly qualified people rejected for a position.