



BUREAU FOR INTERNATIONAL
LANGUAGE CO-ORDINATION

STANAG 6001 Level 4 Tutorial

Testing the Receptive Skills



BILC Working Group on Level 4 Proficiency

April 2017



BUREAU FOR INTERNATIONAL LANGUAGE COORDINATION





Authors: Jana Vasilj-Begovic, Martha Herzog
Contributors: Gerard Seinhorst, Mary Jo Di Biase and Kåre Kildevang

BILC Working Group on Level 4 Proficiency Member Nations:

Bulgaria	Netherlands
Canada	Sweden
Denmark	USA
Germany	NATO/SHAPE

First published	April 2015
Edition 1.2	April 2017



CONTENTS

1.	INTRODUCTION	1
2.	LEVEL 4 PROFICIENCY AND ITS INTERPRETATION	2
3.	PRINCIPLES OF TEXT RATING	9
4.	SAMPLE TEXTS	11
5.	TEST DEVELOPMENT PROCEDURES	18
6.	SAMPLE ITEM DEVELOPMENT PROCEDURE I	19
7.	SAMPLE ITEM DEVELOPMENT PROCEDURE II	24
	List of Works Consulted	33
	Annexes	35
	ANNEX A – Sample texts at Levels 1 - 4	35
	ANNEX B – Samples of typical Level 4 prose	47
	ANNEX C – Terminology relating to the analysis of Level 4 reading texts	53
	Appendices	59
	APPENDIX 1 – Prototype Level 4 Reading Test Specifications	59



[page intentionally left blank]



1. INTRODUCTION

This tutorial consists of two parts. The first part provides guidelines for developing the STANAG 6001 Level 4 reading test prototype IAW the Test Specifications included in Annex 1. The second part, to be added subsequently, will provide similar guidelines for developing a listening test prototype.

The Working Group (WG) members and other prospective Level 4 test developers are advised to study this material carefully. Assimilating the content of this tutorial will be critical for the testers' ability to accurately rate and select texts for the prototype test development.

Since general proficiency tests need to be developed for personnel who may have gained their competence in the target language under a variety of conditions and learning environments, this test cannot be linked to any particular course of instruction, set of learning materials or specific job-related tasks.

In this Tutorial, we will use reading texts found in a variety of print and on-line publications. The Tutorial focuses on Level 4 and directly applies language from the STANAG 6001 Level 4 descriptors to the sample texts selected for this training material.

The Level-4 descriptor for the reading skill is quoted at the beginning, along with statements about the functional trisection and Ray Clifford's overview of text levels. In addition, the theoretical underpinnings of text rating drawn from James Child, as well as comments by Pardee Lowe are included in the discussion on the characteristics of Level 4 texts. In order to facilitate understanding of Level 4 text characteristics, this level will be compared and contrasted with Level 3, both from the reader's perspective and that of the text difficulty features.

The Tutorial is divided into Section I – *Self-Study Module* and Section II – *Workshop Activities*. Workshop activities will be conducted in an interactive training session.

In **Section I**, in addition to using a Level 3 and a Level 4 text for comparative purposes, two additional authentic English language texts taken from US and British sources will be used to illustrate text rating and item development. We will analyze discourse in these texts illustrating text features that are characteristic of Level 4. These two texts will be followed by questions and sample responses illustrating the test model outlined in the Test Specifications.

In **Section II**, the WG members will be asked to analyze features of one or two texts designated for item development practice by filling out the Discourse Analysis Chart, developing questions and including a sample response for each. Once moderated, revised and finalized, these texts will be used for piloting purposes. After that, following the same procedures, the prototype reading test will be developed IAW the Test Specifications.



2. LEVEL 4 PROFICIENCY AND ITS INTERPRETATION

STANAG 6001 (Ed. 5v2) describes Level 4 reading proficiency as follows:

LEVEL 4 - EXPERT
<p>Demonstrates strong competence in reading all styles and forms of the written language used for professional purposes, including texts from unfamiliar general and professional-specialist areas. Contexts include newspapers, magazines, and professional literature written for the well-educated reader and may contain topics from such areas as economics, culture, science, and technology, as well as from the reader’s own field. Can readily follow unpredictable turns of thought on any subject matter addressed to the general reader. Shows both global and detailed understanding of texts including highly abstract concepts. Can understand almost all cultural references and can relate a specific text to other written materials within the culture. Demonstrates a firm grasp of stylistic nuances, irony, and humour. Reading speed is similar to that of a native reader. Can read reasonably legible handwriting without difficulty.</p>

Table 1: STANAG 6001 Level 4 Reading descriptor.

In the presentation below, the reading descriptor has been broken down into content, tasks and accuracy statements.

READING LEVEL 4	
CONTENT	<p>All styles and forms of writing used for professional purposes, including texts from unfamiliar general and professional-specialist areas. Newspapers, magazines, and professional literature written for well-educated native readers Highly abstract concepts. Reasonably legible handwriting.</p>
TASKS	<p>Follow unpredictable turns of thought on any subject matter addressed to the general reader. Show both global and detailed understanding of texts. Understand almost all cultural references.</p>
ACCURACY	<p>Can relate a specific text to other written materials in the culture. Demonstrates a firm grasp of stylistic nuances, irony, and humor. Shows both global and detailed understanding of texts. Understands almost all cultural references. Reading speed is similar to that of a native speaker.</p>



Table 2: Trisection – Reading comprehension



According to Clifford (see Table 3 below), the reader at Level 4 understands meaning “beyond the lines”. In other words, the reader is able to make evaluative judgements or express opinion about a text, evaluate the significance of the author’s message, credibility, intent, and purpose, extrapolate beyond the text and place it in a socio-cultural and historical context. Evaluative comprehension involves making a judgement about the text genre and mode of discourse, rhetorical organization of the text, and the writer’s use of jargon, figures of speech and allusions. Evaluative comprehension skills entail taking into account unstated assumptions in order to understand, evaluate, accept or reject the writer’s arguments. While understanding “between the lines” is a required skill for both STANAG 6001 Levels 3 and 4, understanding “beyond the lines” is a skill that distinguishes Level 4 from lower proficiency levels. It is also important to mention that attaining higher levels of language proficiency implies not only possessing outstanding language skills, but also higher order thinking skills such as deductive and inductive reasoning, analysing and synthesizing.

Level 4 readers understand highly sophisticated written target language on unfamiliar general, abstract or professional-specialist topics. They generally understand specialized language on general abstract and complex topics outside of their area of expertise. These readers can discern relationships among sophisticated written materials in the context of broad experience and can follow unpredictable turns of thought in editorial, conjectural, and literary materials in any subject matter intended for the general reader. They also recognize, understand and almost always correctly interpret cultural allusions, nuance and emotional overtones and attitudes (disenchantment, satire, humour, etc.) as well as less common figures of speech, such as malapropisms and spoonerisms.



OVERVIEW OF TEXT CHARACTERISTICS BY LEVEL

(At each level, these elements must be aligned)

Level	Author Purpose	Typical Text Type	Reader Purpose
1	Orient by communicating main ideas.	Simple short, loosely and predictably organized texts, containing simple grammar and vocabulary. Sentences may be re-sequenced without changing the meaning of the text. Text organization is loose without much cohesion, but follows societal norms. <i>Examples:</i> notices, announcements, advertisements, short notes and emails, weather reports, simple and brief instructions, descriptions, etc.	Orient oneself by identifying the main topics, ideas, or facts.
2	Instruct or inform by communicating organized factual information.	Factual texts about situations and events containing compound and complex sentences organized in paragraphs that may be re-ordered without changing the meaning of the text. Information may be densely packed and these texts establish chronological relationships. Vocabulary tends to be topic specific. The identity of the author is not important. <i>Examples:</i> detailed narrations, descriptions, news reports, routine business letters, simple technical material, such as user instructions for equipment assembly, etc.	Learn by locating and understanding not only the main ideas, but also the supporting details such as temporal and causative relationships.
3	Evaluate situations, concepts, and conflicting ideas; present and support arguments and/or hypotheses with both factual and abstract reasoning, using language that is often accompanied by the appropriate use of wit, sarcasm, irony, or emotionally laden lexical choices.	A multiple-paragraph block of discourse on a variety of unfamiliar or abstract subjects. The author uses facts to build an opinion, express a point of view or provide an analysis of an issue. Texts contain concrete as well as abstract lexicon and complex grammatical structures. The author is present by asserting his/her views of presenting (comparing/ contrasting) the views of others. Texts may contain cultural references or common cultural values, hypothesis, and emotional overtones. <i>Examples:</i> editorials, opinion pieces found in articles discussing societal issues, official correspondence (reports, policies, directives) and other professional writing, etc.	Learn by relating ideas and conceptual arguments. Comprehend the text’s literal and figurative meaning by reading both “the lines” and “between the lines” to recognize the author’s tone and infer the author’s intent.
4	Project lines of thought beyond the expected, connect previously unrelated ideas or concepts, and present complex ideas with nuanced precision and virtuosity – often with the goal of propelling the reader into the author’s world of thought.	Extended discourse that is tailored for the message being sent and for the intended audience. Texts are highly individualized, abstract, innovative and culturally dense, and may contain abstract metaphors and symbolism, and unpredictable shifts in thought, tone and register. To achieve the desired tone and precision of thought, the author will often demonstrate the skillful use of low-frequency, subtle and nuanced vocabulary, cultural and historical concepts, and an understanding of the audience. Substitution of lexical choice might distort the author’s intent. <i>Examples:</i> articles from the fields of military philosophy/ethics, literary criticism, treaties, policies, professional literature, politics, economics, etc.	Read “beyond the lines”, to understand the author’s sociolinguistic and cultural references, follow innovative turns of thought, and interpret the text in view of its wider cultural, societal, and political setting.



Table 3: Overview of text characteristics (adapted from Clifford, 2013)

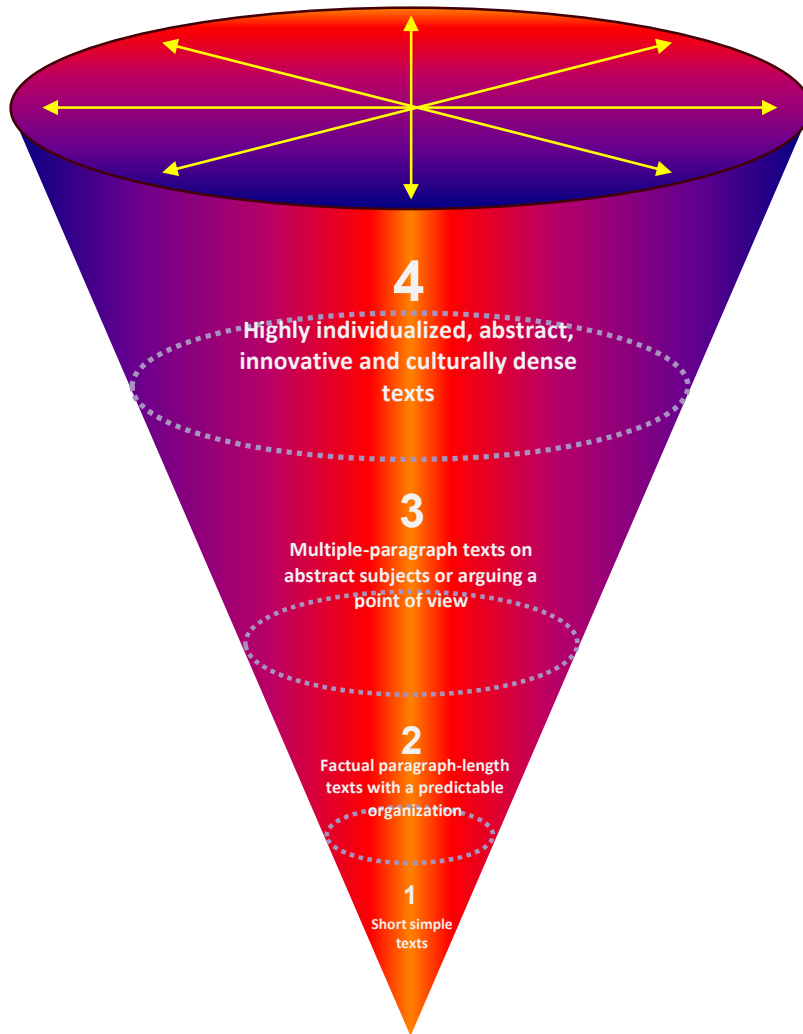


Fig. 1: Text types in relation to the expanding scale

It is also important for test developers to be able to distinguish between Level 3 and Level 4 texts. Even though both levels deal with ideas and abstraction, there are distinctly different degrees of complexity, dependence on nuance, and types of propositions relating to the shared knowledge of writer and reader. Also, the syntactical richness used by the writer and the pragmatic inference required of the reader differ distinctly by level. Tables 4 and 5 below show the differences between Level 3 and Level 4 texts from the perspectives of reader purpose and text characteristics.



What the Reader Can Do

LEVEL 3	LEVEL 4
Understands formal and informal language, for most everyday social and work-related situations.	Understands highly sophisticated language appropriate for almost all topics and social as well as professional situations.
Understands language relating to abstract concepts and hypotheses.	Shows a firm grasp of various levels of style and understands subtle nuances and shades of meaning.
Can understand literal and figurative meaning, and implied meaning (“between the lines”)	Understands meaning, “beyond the lines”. Follows innovative turns of thought.
Understands the author’s tone and intent, including humour and irony, but may not fully understand some allusions, as well as implications of many nuances and idioms.	Understands the author’s socio-linguistic and cultural references. Understands idiomatic expressions, figures of speech and colloquialisms.
Learns by relating ideas and conceptual arguments.	Interprets the text in view of its wider cultural, societal, and political setting.

Table 4: Significant differences between STANAG 6001 Level 3 and Level 4

LEVEL 3 – EVALUATIVE MODE	LEVEL 4 – PROJECTIVE MODE
Texts often contain facts used to provide analysis, commentary, personal views, apologies, justification, advocating a course of action, arguing a side of issue, for example editorials in newspapers for educated readers are often written at this level.	Texts are abstract and culturally dense; syntax used with virtuosity; rhetorical devices apparent.
Analysis and evaluation of issues take place against a backdrop of shared information.	Novel/non-linear approach to treating topics/issues or developing arguments.
Abstract concept and linguistic formulations.	Individualized or culture-specific forms of discourse, abstract metaphors, and symbolism.
Hypothesis, tone and inference.	Unpredictable turns of thought.
Shared knowledge and target culture between author and reader.	Author’s approach to shared knowledge is unique.
Tone (persuasive, ironic etc.); emotional overtones.	Distinctive tone evident.
Shaping (thoughtful, selective and purposeful use of language that often becomes the signature of the particular author) is evident.	Highly nuanced language, the substitution of which may change author’s intent.

Table 5. Text Types and Characteristics



3. PRINCIPLES OF TEXT RATING

Ray Clifford's overview of text characteristics by level is built upon the work done by James Child over the course of several decades of service in the US Government. Child supplemented such scales as the Interagency Language Roundtable and the STANAG 6001 by defining the characteristics of texts at four major proficiency levels – Base Levels 1, 2, 3, and 4. His ideas have been circulated throughout the US Government's language learning and testing community and have also influenced academic discussion of text levels. His various articles on text typology have been published in several journals and books. Readers are encouraged to consult "Language Proficiency Levels and the Typology of Texts," in *Defining and Developing Proficiency: Guidelines, Implementation, and Concepts*, which was edited by Heidi Byrnes and Michael Canale in 1987 and "Language Skill Levels, Textual Modes, and the Rating Process," published in *Foreign Language Annals* in 1998.

Child's concept defines four text modes, ranging from Base Level 1 to Base Level 4. These are the *orientation mode*, which communicates a central idea, related to the immediate context; the *instructional mode*, which communicates factual information; the *evaluative mode*, which demonstrates a clear point of view by the author, concerning either analytical or affective content; and the *projective mode*, which presents original and innovative turns of thought by the author. In general, as text modes become more complex, they also become longer. Orientation mode texts are constructed at sentence level; they are rarely more than 1-3 lines in length. Instructional mode texts are paragraph length; they are characterized by one or more fully developed paragraphs. Evaluative mode texts may be essay length, consisting of a series of cohesive, well-constructed, carefully connected paragraphs. Because projective mode texts are the subject of this Tutorial, much of our discussion will be devoted to their characteristics – that is, to a discourse analysis intended to help test developers (in this case, the WG members) recognize and rate these texts accurately.

Projective mode texts demonstrate extended discourse; they are usually essay length or longer. They represent an author's unique point of view, and the method of argumentation may be complex and innovative. Personal interests or academic and work-related requirements will dictate whether any given text is selected to be a regular, occasional or rare part of one's reading experience.

According to Pardee Lowe (June 22, 1998), Level 4 texts are "abstract & culturally dense, often have embedding syntax used with virtuosity," while Edwards (1996) states that the author "may take a novel or creative approach to a problem." She continues to say that "in these texts, the reader is likely to encounter highly individualized or culture-specific forms of discourse, abstract metaphors, and symbolism. The author assumes a great deal of reader input and leaves historical, cultural or other references and assumptions unexplained."



Level 4 texts demonstrate a considerable range of subtlety and nuance of expression. To present a novel, highly individualistic opinion or line of argumentation, the author uses tone that may be sardonic, tongue-in-cheek, playful, solemn, scathing, etc. The level of abstraction in these texts can be very high and readers familiar with the text rating premises can see clear distinctions between the layers and level of abstraction in these texts compared to those found in Level 3 texts. As a general rule, the higher the level of the text or discourse, the harder it is to convey the author's intent with synonyms.

According to Child, “*Belles-lettristic* creativity, too, is a hallmark of the Projective mode, especially at the higher end of the spectrum, which can be seen as the natural realm of artistic creativity.” For full reference, see the documents at Appendix 2.

Please note that the purpose of the texts used in this Tutorial is to contrast the characteristics between a Level 3 and Level 4 text, and in particular, to highlight the features of Level 4 texts, and not to show necessarily the text genres that would be used on a Level 4 reading test intended for military personnel. Even though general proficiency will be tested, the relevance of selected texts to the test-taking population will be given consideration.



4. SAMPLE TEXTS

Each of these two texts is an analysis of an aspect of Clausewitz's book, "On War". Both texts have ideational, that is abstract content; however, the level of abstraction differs. Texts of this length would need to be edited for purposes of testing. Here they are used for the purposes of illustrating the differences between Level 3 and Level 4 texts.

A comparative discourse analysis follows the texts justifying the rating and contrasting the two texts.

TEXT 1

Clausewitz and the Blue Flower of Romanticism: Understanding On War

by Bill Bentley

Introduction

War amongst the people, hybrid war, unrestricted warfare, counter-insurgency (COIN), stabilization operations, culture wars, new wars versus old wars, terror wars, these are examples of the conceptualization of war and conflict that have emerged in the wake of the end of the Cold War. Currently overshadowed by the contemporary security environment, there remains the historically most prevalent form of conflict; that of high intensity state-on-state war. The work of one man first published 181 years ago can be very helpful in understanding both the nature and character of all these types war and conflict in the 21st Century, regardless of how they are described and explained, however imperfectly, in today's relevant literature, military colleges, and other policy making fora. In this regard, a clear understanding of the relationship between the thought of Carl von Clausewitz and German Romanticism greatly enhances our understanding of how this Prussian military thinker constructed his unique theory of war, and what he meant by it.

As to his actual influence extending throughout the 20th Century and into the 21st Century, there are numerous testimonials to the impact and value of General Carl von Clausewitz's life-long attempt to develop a theory of war and conflict that aimed, not at prescription, but at a deep understanding of the phenomenon itself. Four will suffice here to make the point. T.E. Lawrence (of Arabia) was a great admirer of On War, finding it logical and fascinating. More recently, H. T. Hayden, editor of *Warfighting: Maneuvre Warfare in the US Marine Corps*, stipulates that Clausewitz's On War is the definitive treatment of the nature and theory of war. According to W.B. Gallie, a British philosopher and Second World War infantry major, the years 1818-1832 saw the writing and eventual publication of the first, and



to date, the only book (*On War*) of outstanding intellectual eminence on the subject of war. Finally, one of the leading strategic theorists still writing today has concluded that "...for as long as humankind engages in warfare, Clausewitz must rule.

Discussion

The subject of these accolades, Carl von Clausewitz, was born in Burg, Prussia, in 1780, and joined the Prussian army at the age of twelve. From 1792 until 1815, he fought against the French revolutionary and Napoleonic forces at the battles of Valmy, Jena, Borodino, Leipzig, and Waterloo. In 1818, he was appointed Director of the Prussian War Academy in Berlin, and for the next 12 years, devoted himself to writing about war, and especially, to the construction of his theory of war expounded in his masterpiece, *On War*. The unfinished manuscript of *On War* was published by his wife, Marie, in 1832.

On War was translated into French in 1849 and English for the first time in 1874. Notwithstanding its growing availability, the book was little read for 40 years after it first appeared in 1832. This interest was sparked by virtue of the fact that the military architect of these startling successes, General Helmut von Moltke, frequently cited von Clausewitz's *On War* as one of the major intellectual and theoretical influences that shaped his strategic thinking and the associated campaigns.

Nonetheless, *On War* remained notoriously difficult to interpret and to fully understand. It was often misquoted and frequently referred to without having been read carefully, if at all. Except for a few dedicated, serious scholars of *On War*, this remains largely the case today.

A large part of the difficulty in this regard has been the failure to place Clausewitz's thought firmly in the relevant intellectual context. It is necessary to read *On War*, as John Lynn has recently argued, not as a work that expresses eternal truth about war, but within the intellectual context that generated it. It is a Romantic work, and like the broader intellectual and cultural movement of Romanticism, it cannot be understood apart from the intellectual paradigm that it challenged—the dry rationalism of the Enlightenment. There is no doubt that the Enlightenment was itself a very complex movement that should not be reduced to a few platitudes. Taken as a whole, the Enlightenment (circa 1687-1789) sought to establish the intellectual foundation for a political system without divine sanction, a religion without mystery, a morality without dogma. Such was the edifice the Enlightenment thinkers believed man now had to erect.

Science would have to become something more than an intellectual pastime; it would have to develop into a power capable of harnessing the forces of nature to the service of mankind. Science was the key to happiness. With respect to the material world, once it was in his power, man could order it for his own benefits and for the happiness of future generations. Such are the "notes" by which the 18th century is readily



identifiable. But for the Romantics, from Coleridge, Chateaubriand, Herder, and Heine, through Schelling and Hegel, a vision of the Enlightenment emerged that was as cold, timeless, monotonous, and as calculating as the bourgeois who had supposedly embraced it. In their view, the Enlightenment enshrined mechanism as the model to explain all phenomena of matter, life and mind.

(Excerpt taken from the Canadian Military Journal)

TEXT 2

TIP-TOE THROUGH THE TRINITY

THE STRANGE PERSISTENCE OF TRINITARIAN WARFARE

Christopher Bassford

INTRODUCTION

I am a historian. Or, rather, I was formally trained as an historian. Today, as a "Professor of Strategy," I'm not sure I can still characterize myself that way. But my approach to teaching strategy is essentially an historical approach. I routinely start out a new seminar group with the question, "So: Why do we study history, anyway?" Invariably, some earnest young colonel will volunteer that hoary old line from George Santayana, "Those who cannot remember the past are condemned to repeat it." I will then fix what I hope is a withering eye on this student and say something to the effect of, "Well, Bubba/Bubbette, I've got some bad news for you: Those of us who do remember the past are also condemned to repeat it. We simply have the added pleasure of knowing we've been down this damned rathole before."

Unfortunately, even this minor pleasure does not appear to be widely shared. It has been barely one generation since the American defeat in Indochina. Nonetheless, in a dazzling display of historical forgetfulness worthy of the brain-damaged female protagonist of the movie "Fifty First Dates," our national security community appears to be stunned to discover that warfare can be waged by groups that do not look at all like the cast of the 1960s TV show, Hogan's Heroes. And many of the worst offenders are... military historians. Prompted by what evidently appears to many writers to be the utterly new kind of warfare waged by organizations like, say, Al Qaeda, they spin out bold new buzzwords designed, shaman-style, to capture the spirit of this earthshakingly new innovation by giving it a name. Some popular examples are "non-state war," "Fourth- (or Fifth-) Generation War," and the stunningly uncreative "the New Warfare." Most misleading of all (to the few who are equipped to assign any meaning whatsoever to the phrase) is "non-trinitarian war," a term coined by



Israeli military historian Martin van Creveld to encapsulate a new, allegedly “non-Clausewitzian” approach to theorizing about war.

In reality, Clausewitz’s Trinitarian concept bears little resemblance to the concept Creveld claims to be refuting. The purposes of this paper are to examine the meaning and significance of this trinity and to explore its continuing relevance to contemporary political/military problems. A warning to “hard-headed realists,” “operators,” and “practical men”: Clausewitz was a practical soldier and he intended his work to serve as a very practical approach to real-world complexities – *without avoiding the complexity*. If you are one of those people who are repelled by the allegedly “hair-splitting” character of Clausewitzian theory, I advise you to stop reading now. I would also suggest that you quit – today – any profession connected to the higher levels of politics, public policy, or war.

Universalizing Clausewitz

As a final prefatory comment, I should note a certain bias, of which I am well aware, in my own thinking. This is a bias towards universalizing Clausewitz, who, after all, sought with *Vom Kriege* to formulate a valid general theory of war. Thus I think it would be a “good” thing if the entire war-studies community could use the essentials of Clausewitzian theory as the common basis for comparative military-political studies across all human societies and history. A common understanding of the Trinity would do much to advance that project. While I am under no illusions that the achievement of such a common understanding is imminent (which would make me certifiably insane), that goal is implicit in the very existence of “military theory” as a coherent field of inquiry. Our current utter confusion will continue to prevail until we find some common theoretical structure within which to conduct our debates. I do not have space to explore that notion here to any great degree, but it certainly influences my choices in translation and in defining terms like policy and politics: We want definitions that are not confined to Prussia in the era of the French Revolution, the Westphalian-model state system, or Western civilization. As, I think, did Clausewitz. Because so much of the debate over Clausewitz tends to reflect an academic “instinct for the capillaries,” I offer this bias as a convenient target for anyone seeking the jugular vein of my analysis.

The Problem of the Trinity

Conceptual trinitaries are inherently problematic, but especially in the context of contemporary American politics and policy. It is difficult enough to convey the meaning and implications of any single idea. But the world in which we actually operate is, despite what partisans and fundamentalists of various stripes like to pretend, never the simple unfolding of any one concept or force. Unfortunately, expanding a discussion even to *two* coexisting or contending ideas requires a philosophical leap into the obscure realm of “dialectical thinking” – especially if one’s purpose is more sophisticated than seeking merely to eliminate one of the two options or to split the difference. Attempting to adjust to the complex situation routinely



created by the ongoing collision of two real-world facts thus makes one, in modern parlance, a “hopeless flip-flopper.” Juggling the meaning and implications of *three* interacting realities makes even splitting the difference obnoxiously difficult and is utterly beyond the pale of acceptable political analysis. As famed political strategist James Carville notes, “If you say three things, you say nothing.” This contemporary reality makes sensible thought impossible, so we will ignore it in this paper.

Many of the difficulties specifically with Clausewitz’s trinity seem at first to turn on issues of translation from the German into English. We will have to look at some of these issues. Ultimately, however, the problem has little to do with the German or English languages or cultures, per se. Rather, it derives from the different ways in which various individuals, disciplines, and subcultures understand the universe (or, as many phrase it, “How things really work”) – issues raised most ably by Alan Beyerchen’s analysis of Clausewitz’s worldview in terms of nonlinear mathematics and Complexity science. For reasons of space I will have to leave most of the staggering implications of Clausewitz’s choice of nonlinear scientific imagery to Alan. The two issues are not unrelated, because both turn on the interactivity of interdependent variables. But we can discuss, e.g., the word-choice issues regarding policy, politics, and Politik in rather traditional terms without invoking any new cosmic paradigms.

The overarching problem with attempting any short discussion of the Trinity in isolation is that, however much various writers may try to treat the Trinity as a discrete theoretical “nugget” – indeed, as an afterthought, a conception that allegedly popped into Clausewitz’s mind in the last phases of his unfinished writing process and was never effectively incorporated into the existing body of his theory – in fact, the Trinity is the central concept in *On War*. I don’t mean “central” in the sense that, say, Jon Sumida applied in his Oxford conference paper to Clausewitz’s concept of the inherent superiority of the defensive form of war. That is, I do not argue that the Trinity is Clausewitz’s “most important” concept, that the desire to convey it was his primary motivation in writing, or that all of his other insights flowed from this one. Rather, I mean simply that the Trinity is the concept that ties all of Clausewitz’s many ideas together and binds them into a meaningful whole. This remains true whether Clausewitz conceived his theoretical universe with this construct in mind, or instead discovered only at the end of his efforts that the seemingly divergent roads he had been traveling all led, inexorably, to this particular intersection.

An intersection is of little significance, however, without reference to the roads that run through it. Thus it is difficult to confine a discussion of the Trinity within tidy boundaries: Any comprehensive examination must lead to every major issue in *On War*.

In any case, the role of the Trinity within the narrow confines of Book One, Chapter One of *On War*, which reflects Clausewitz’s most mature thinking, is crucial. That chapter must be read in terms of Clausewitz’s dialectical examination of the nature of war. That discussion is very carefully structured but (purposefully, I suspect) largely



unmarked by clear dialectical road markers labeling thesis, antithesis, and synthesis, or even by sections clearly devoted to one stage of the dialectic or another. The Trinity itself represents the synthesis of this dialectical process. In this chapter, at least, it is no afterthought, clearly being foreshadowed throughout the discussion.

All the circumstances on which [war] rests, and which determine its leading features, viz., our own power, the enemy's power, allies on both sides, the characteristics of the people and the governments respectively, etc., as enumerated in Book I, Chapter 1 – are they not of a political nature, and are they not so intimately connected with the whole political intercourse that it is impossible to separate them from it? But this view is doubly indispensable if we reflect that real war is no such consistent effort tending to the last extreme, as it should be according to abstract theory, but a half-hearted thing, a contradiction in itself; that, as such, it cannot follow its own laws, but must be looked upon as part of another whole – and this whole is "Politik".



TEXT 1	TEXT 2
<p>Topic: Military Philosophy. <i>Clausewitz’s Theory of War and its Relationship to Romanticism</i></p> <p>Type of text: analysis and opinion (essay)</p> <p>Mode: <i>Evaluative</i>. The author uses both facts and abstract ideas to present his argument. The author presents most background information (e.g., states clearly the characteristics of the period of Enlightenment).</p> <p>Using abstract vocabulary and linguistic formulations (“divine sanction”, “morality without dogma”, etc.) builds his argument that to understand Clausewitz’s theory one must place it in its historical intellectual context.</p> <p>Tone: persuasive (achieved through careful and well-reasoned presentation of fact and opinion)</p>	<p>Topic: Military Philosophy. <i>Clausewitz’s Theory of War</i>, and in particular <i>his Concept of Trinity</i></p> <p>Type of text: analysis and opinion (essay)</p> <p>Mode: <i>Projective</i>. Innovative approach to the subject. The author does not present a lot of background information (makes reference to Hegel’s dialectics without defining it)</p> <p>Complex ideas (such as hypothesis in the last paragraph) presented with virtuosity. Nuanced and figurative language used when discussing a highly abstract idea of ‘conceptual trinities’ (“Juggling the meaning and implications of three interacting realities makes even splitting the difference obnoxiously difficult and is utterly beyond the pale of acceptable political analysis”).</p> <p>Tone: pervasive and individuated throughout the text; at times humorous (“withering eye”, “bubba/bubette”). Formal/informal mix; including references to pop culture mixed with serious ideas (“in a dazzling display of historical forgetfulness worthy of the brain-damaged female protagonist of the movie”, “Fifty First Dates”) ironic (“invoking cosmic paradigms”).</p>
<p>RATING: Level 3</p>	<p>RATING: Level 4</p>

Table 6: Comparison between Text 1 and Text 2 Characteristics



5. TEST DEVELOPMENT PROCEDURES

The steps listed below will be followed in the development of test items:

(1) *Select a text*

Compare it to the descriptor, trisection, and Clifford's overview and assign an *initial* rating to it.

(2) *Determine if editing is needed*

While we want the texts to be as authentic as possible, some selections may be too long for this purpose. If a text can be shortened and still maintain a good representation of the original content, as well as its Level 4 features, and retain coherence, there can be judicious editing to fit the text length given in the test specifications.

(3) *Analyze the text*

Using the Text Analysis Chart, analyze characteristics of each text. This step will facilitate more accurate and reliable text rating, as well as help test developers identify/map out authors' ideas and lines of argumentation, which in turn will facilitate the process of item writing.

(4) *Rate the text*

Based on the text analysis and comparison of the text characteristics with Clifford's Overview of Text Types, assign the final rating to the text.

(5) *Review appropriate authentic tasks*

Tasks can be extracted from the C/T/A statements and the Test Specifications. Even though the testing situation is artificial by its nature, this Level 4 test format reflects real world language use. Friends and colleagues do discuss their thoughts about an author's ideas and attitudes.

(6) *Develop tasks for the test items* that reflect the descriptor and real-world tasks

IAW the Test Specifications (see Appendix 1), this is an interactive test. An interactive (sometimes also called a *dynamic*) test is one in which the examinee produces spoken or written responses to a prompt.

(7) *Write, review, and revise test items*



6. SAMPLE ITEM DEVELOPMENT PROCEDURE I

1. *Select a text*

TEXT 1

An excerpt from "A Howard Nemerov Reader" – a book on literary criticism.

Poetry and Meaning

By Howard Nemerov

There is a sense, utterly true but not very helpful, in which everything is always ending and always beginning. The fabric of the generations simply is woven that way, seamlessly, and only the work of the historical intellect divides it up. Imagine someone living through the fall of the Roman Empire in a provincial town, in Marseille say, or London; he would live his life day by day, as we all do, and never know that he had lived through the fall of the Roman Empire. He would notice, perhaps, certain signs of neglect; the garrison might go slovenly and unshaven, the roads might not be so well kept up, proclamations would be fewer than they used to be... and when people began to notice the absence of something called The Roman Empire they nostalgically replaced it with a Holy Roman Empire and pretended it was the same thing, sort of. So it may be with my subject. I will present the evidence as best I can.

For one thing, the posture of the literary mind seems these days to be dry, angry, smart, jeering, cynical; as though once people had discovered the sneaky joys of irreverence they were quite unable to stop. This is one typical process of Shakespeare's tragedies, where the intelligent and crafty young destroy the stupid old and, with them, the sacred something that these complacent dodos by some accident had in their charge, and the intelligent and crafty young, at last, as Ulysses says, eat up themselves.

This symptom in itself is perhaps not much. Literary quarrels have usually been acrimonious, indeed are less personally spiteful now than in the Age of Pope. The world has always been as full of people plugging their friends as of people unplugging their enemies. Yet the public discussion, the criticism that attends on poetry, has appeared to me as coming close to the point at which a smart shallowness and verbal facility will jettison meaning altogether; the same thing has been happening in poetry itself. I shall not now give examples, but I ask you to consider whether it is not as I have said. Not only the terms of abuse, but more importantly the



terms of praise, appear in a language whose vagueness of sense is closely related to the extravagance of its claims.

This kind of shrillness may be the sign of considerable unacknowledged anguish of spirit. As though everyone felt some big thing was breaking up, and made bigger and louder noises to pretend that all is as it was. For it ought now to be possible to turn and look back over the modern period, as it foolishly goes on being called, and see how some one thing – I should date it perhaps from the middle of the last century, from Baudelaire and Swinburne, say – was gathering momentum in a direction and was assembling armies of adherents, but that not so long ago this momentum, giant as it was, divided itself among the members of the armies, diminished, and may now be flickering out in brief contingencies.

2. *Determine if editing is needed*

The text does not need to be shortened. The word count of this text is 516 words; therefore, it falls within the range of text length prescribed by the Test Specifications.

The text was written about 1980; hence, the phrase “last century” in the last paragraph might be changed to “19th century” for clarity.

3. *Analyze the text*

TEXT ANALYSIS CHART	
<i>Text Purpose</i>	To project beyond the lines.
<i>Ideas in support of purpose</i>	The author expresses ideas about the current state of literary criticism and “projects” his line of argumentation to the wider social situation which is undergoing change.
<i>Other ideas present but not required in support of purpose</i>	The author makes an analogy between the current situation and one in the past.
<i>Distinguishing features</i>	Highly abstract, projective mode, numerous literary references. There may be little shared knowledge between the author and the reader.
<i>Opinions expressed</i>	Yes. The author believes that poetry itself as well as its criticism has become shallow, losing its sense and meaning.
<i>Devices used to express opinions</i>	Abstract vocabulary, complex sentences, literary references.



<i>Presence of analysis</i>	Yes. The author analyses the reasons behind the trends of vapid and shallow literary discussions, and positive and negative criticisms that tend to obliterate meaning.
<i>Author's tone</i>	Critical, formal, somewhat derisive.
<i>Vocabulary used to contribute to argument or tone</i>	Modals “shall, should and ought”; “attends on” (low frequency formal use); “foolishly goes on to be called”; “pretended it was the same thing”; the use of adjectives describing the current literary mind, such as “dry, angry, jarring, jeering, cynical”; wordplay, such as “plugging their friends” (promoting the work of their friends) and “unplugging their enemies” (cutting off their resources), etc.
<i>Abstract ideas or abstract linguistic formulations</i>	“Sneaky joys of irreverence, extravagance of claims, vagueness of sense, unacknowledged anguish of spirit, smart shallowness, verbal facility, flickering in brief contingencies”.
<i>Presence of hypothesis (speculative mode)</i>	A fully formulated hypothesis is not present; however, the author speculates on the cause of the phenomenon he has observed and criticized.
<i>Metaphors and other figures of speech</i>	“Fabric of the generations, posture of the literary mind, armies of adherents”.
<i>Literary, cultural, historical, religious and other allusions</i>	Many, including “Complacent dodos” and “Age of Pope” .
<i>Colloquialisms, slang, street talk, words invented by writer</i>	Sort of, plugging, unplugging.
<i>Technical jargon (military legalese, scientific etc.)</i>	N/A
<i>Archaic words or expressions</i>	N/A
<i>Other types of vocabulary (puns, spoonerisms etc.)</i>	N/A

Table 7: Text Analysis Chart



4. *Rate the text*

Based on the text analysis, we would rate the text as Level 4. It fits Clifford's "Projective Mode" in that it "*projects* lines of thought beyond the expected; connects previously unrelated ideas or concepts, or presents complex ideas with nuanced precision and virtuosity".

The language of this text is abstract, and includes many abstract linguistic formulations as well as literary references. Nuanced lexical choices paint a vivid picture of noise or clamouring made by critics and poets as a possible reflection of spiritual anguish resulting from the sense that some familiar structures are disintegrating.

5. *Review appropriate authentic tasks*

The Level 4 reader, IAW the STANAG descriptor, is able to follow unpredictable turns of thought. An example of that would be the second paragraph, which connects to the first one unexpectedly, as it introduces the author's focus of discussion. Indeed, the reader may even feel that the second paragraph does not follow the first one, either thematically or organically, but s/he can readily adjust to this shift.

To understand this text, the reader needs to be able to evaluate the author's analysis, opinions and argumentation and link them to the wider social, historical and cultural context of the slice of time the author describes.

6. *Develop authentic tasks for the test items* that reflect the descriptor

Some of the reading tasks at Level 4 are:

1. Follows unpredictable turns of thought on any subject matter addressed to the general reader
2. Shows both global and detailed understanding of texts
3. Understands almost all cultural references
4. Understands highly abstract vocabulary
5. Recognizes tone

7. *Write, review, and revise test items*

The following questions would be considered as the first draft test items:

Questions:

1. What symptoms reflecting the spirit of the time has the author observed and criticized? How are these symptoms manifested?
2. What analogy can be drawn between the author's comment on the significant historical permutations referred to in paragraphs 1 and 4?
3. According to the author, what are the ramifications of the "sneaky joys of irreverence"?



4. In the last paragraph, the author speculates about the underlying causes of the trend he criticizes. To what influencing factors does the author ascribe “shrillness”?

These questions may be modified during the item review sessions.

Sample responses (based on the draft questions)

The responses below represent some possible answers to the questions. They also illustrate different levels of complexity and abstraction that a test-taker may use in his/her formulation of the response.

1. “The author has noticed a change in the nature of what he calls the “literary mind”. These changes are reflected in a display of negative attitudes toward literary works, especially poetry. Even though literary feuds have always existed, contemporary poetry critics are using such vacuous language that their criticism of poetry obliterates its very meaning.”
2. “In the author’s view, dramatic and significant historical changes may happen without being noticed by the people living through them. It is only in retrospect that these changes stand clearly delineated against the historical backdrop.”
3. “Once adopted, irreverent attitudes may be too seductive and enjoyable to change. These attitudes may lead to the destruction of the old, but in the process they also tend to destroy everything that was good and sacred, and that ultimately leads to self-destruction.”
4. “The author believes change is in the air, a change that people are not even aware of; however this change still has an effect on the spirit of the time. It also influences the acerbic, angry and irreverent tone of literary critics. All this noise and clamouring reflect a sense of something changing, vanishing and disintegrating; however, there remains a pretense that stability is still present.”

NOTE: the sample responses provided above are more representative of the formal written than the typical spoken style. Even though the spoken responses may be less formal in their choice of vocabulary or syntax, they would still be expected to be precise enough to convey the ideas expressed in the text.



7. SAMPLE ITEM DEVELOPMENT PROCEDURE II

Below is another approach to text rating, analysis and item development.

TEXT 2

From William T. Vollman: *Rising Up and Rising Down: Some Thoughts on Violence, Freedom and Urgent Means*.

COPS LOOKING FOR MOLEST CASE SUSPECT: *Penang – Police are looking for a factory bus driver who allegedly fondled the breasts of a 23-year-old woman in Air Itam here recently... It is learnt that the suspect also kissed the woman.*

Malaysia being a Muslim country, what the bus driver did was very likely a terrible assault upon the woman's pride and sacred secretness. She must have been veiled, in which case his kissing her would have been as much an act of exposure, of humiliation, as his snatching her breasts into his hands. Call it a rape. When I asked my Thai companion, D., how she would have responded had he done this to her, she said, "Me? Maybe I laugh. Don't like so much. Maybe little bit angry, but I try to talk to him. I think he need some girlfriend. Maybe I say to him, you very silly boy."

The conclusion I come to (one abhorrent to local law, but agreeable to the contradictions of international customs) is that a major defining ethical constituent of violence is the unique relationship between each victim and the perpetrator at a given time. That is why even the sternly consistent Martin Luther warns the sixteenth-century German princes that they must be guided by their own minds and consciences in any given case of legal judgment – and why it is permissible, for instance, for somebody to kill one person in wartime but not another, and neither one in peacetime. And if the motive and the context are so crucial, then we must ask whether one can with equal justification kill out of hatred, out of fear, out of rational self-defense or out of mercy?

Arendt, musing upon the "pale" and "ghostlike" figure of the accused Eichmann in his glass booth, insisted: "If he suffers, he must suffer for what he has done, not for what he has caused others to suffer." As a general rule, this principle is demonstrably untrue. D., a live-and-let-live Buddhist from a sexually more easy country than Malaysia (the rural markets usually have on the walls colored advertising posters with photos of naked women) would not have suffered what the twenty-three-year-old Muslim woman presumably did. I know this because I know D. This is not to say that she would have suffered nothing; for in public life Thais are sufficiently modest



that the public kissing of European lovers inspires them with disgust. Nor would that brute of a bus driver be any more justified in fondling her than he would have in the case of a Muslim girl. But D. never wore a veil. She would have been kissed without being uncovered. In short, she would have been humiliated, not raped. She would have shrugged it off and tried not to be shamed; the Muslim girl might well have been hurt to the core. And had, let's say, a man in Mexico City done this to D., and not a bus driver whom she could not get away from but a seatmate whose escalating flirting she had ignored; had the bus also, let's say, been filled with drunken soccer players who were groping giggling cheerleaders at least some of whom desired an orgy, I would, again, not excuse the brute, but I would be slightly less angry, thinking of him (as I do my friend from the rough town) as someone who didn't know where to stop as opposed to someone who coolly initiated something. And I would think even worse of the Malaysian bus driver had he forced his attentions on the woman if no other passengers were on the bus than if the bus were crowded; for if the bus were empty the woman would feel more alone and helpless, the act hence more the vicious one of intimidation, domination and humiliation. If the same action can cause significantly different degrees of injury to different victims, then the deed itself cannot be adequately described without context.

Arendt gets around this by substituting the social fabric as a whole for the personal vagaries of any victim – a Kantian strategy which “may be rendered by saying that the undeserved evil which anyone commits on another is to be regarded as perpetrated on himself...This is the right of retaliation (*jus talonis*); and...is the only principle which in regulating a public court, as distinguished from mere private judgement, can definitely assign both the quality and the quantity of a just penalty.”

Thus Eichmann's genocide was in Nuremberg parlance a crime against humanity because it attacked human diversity, without which the whole concept of humanity becomes reduced to ethnicity or nationality. This argument is valid in Eichmann's case, and would remain so if his expertise had killed only one person instead of millions; who, except the suicidally inclined (whose case will be taken up later in this book) would be at variance over the ultimate negativity of death? Had the Malaysian molester done his deed not with a kiss but with a dagger, then D. and the woman on the bus would have been more equally harmed. If not, then the merely normative approach fails to hold.

Interestingly enough, the importance of context from the other point of view – the aggressor's – was recognized up to a point even at Nuremberg. Just as the military



court which tried and sentenced the assassins of President Lincoln endured (and probably instigated) experts' haggles on whether they were mentally or morally insane – in the end, they decided that the law didn't care – so too the Nuremberg tribunal ground through the motions of debate as to whether Rudolf Hess was sane enough to stand trial. In Churchill's memoirs he's implied to be mad; and the prosecutor Telford Taylor portrays a defendant rarely able to concentrate, listen or remember. The remarkable point is that the issue was raised at all. In the end, expediency, justice, or perhaps vindictiveness won out, and Hess was tried. A decade and a half later, so was Eichmann, who displayed a different sort of madness. I wish that Kant had been there, for vis-a-vis the "I just followed orders" defense the philosopher expresses agreement: "The good or bad consequences arising from the performance of an obligated action – as also to the consequences arising from failing to perform a meritorious action – cannot be imputed to the agent (modus imputationis poneus)". The implication is that the social medium in which one swims (or, as Kant probably would have preferred to put it, the institutional uniform in which one clothes oneself) automatically justifies the actions which it condones and commands. By conforming and obeying, the Eichmanns are exculpated. I would have loved to see the look on the chief prosecutor's face. Well, if I'd been there when Eichmann was speaking, I guess I would have seen it; for this was precisely the argument which that monster used. Whether one agrees with his line or not (and I don't), it surely makes a difference to our moral or metaphysical understanding of his crimes ("I committed mass murder") – as opposed to our juridical comprehension ("I upheld my obligation to authority", or perhaps the very different "I violated the international laws against war crimes") – whether Eichmann donned the livery of the state in 1939 or simply flew the colors of a non-representative cabal whose 'agent' he was. In one case (the most likely one), the regime made him what he was. In the other, he would have done it regardless, like the opportunistic rapist in the elevator. Either way, let's hang him, but if the first cause is the dominant one, then we've learned that there's something very useful we can do with our lives: study Nazism in detail, in order to discover how to prevent it from coming to life again. If the second case gets privileged, it's more utilitarian to study the various Eichmanns.

Before analyzing the text's features, let us briefly review the so-called "triangulation process".

The graphic below illustrates the three aspects of the text rating process. In other words, when deciding on a text rating, test developers will review the STANAG 6001 Level 4 descriptor, consult Dr. Clifford's Overview of Text Types (author purpose and look at the text from its in-language perspective (text characteristics)).

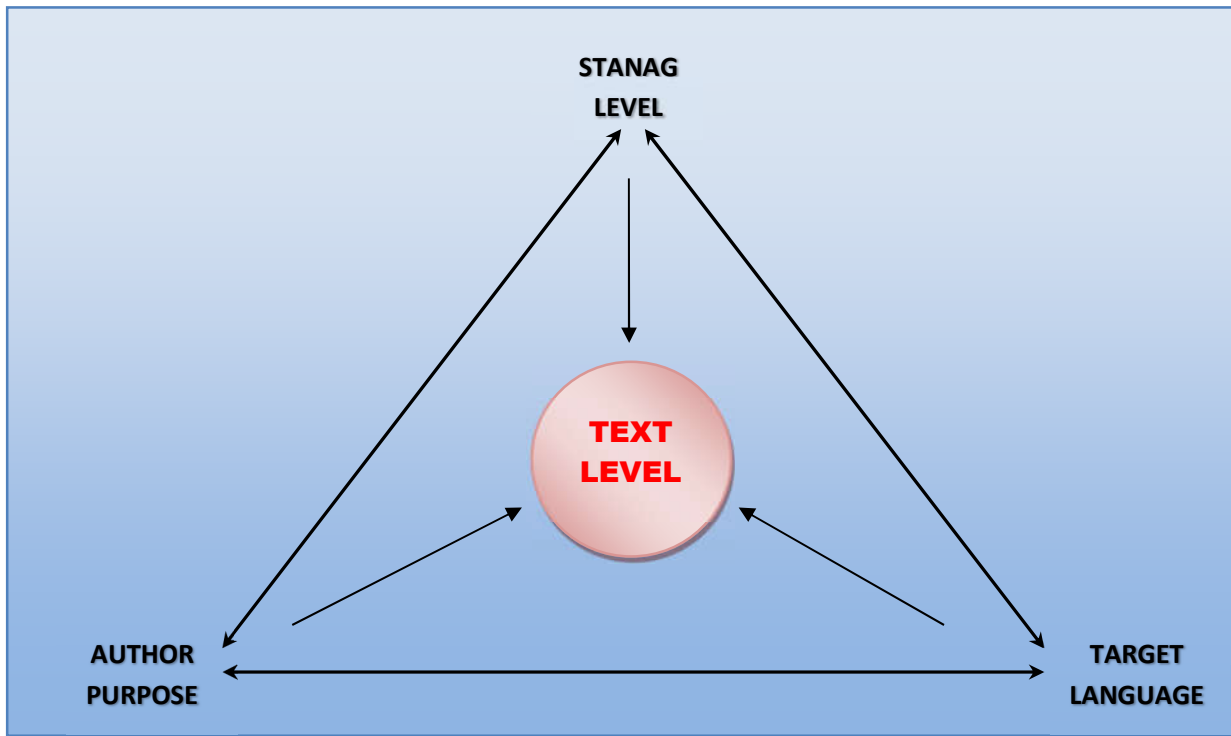


Fig. 2: The triangulation process of text rating

The Three Sides of the Triangle

1. **STANAG Level 4:** As mentioned earlier, IAW STANAG 6001, the Level 4 reader understands all styles and forms of writing; follows unpredictable turns of thought on any subject matter addressed to the general reader; understands almost all cultural references and relates a specific text to other written materials in the culture. The reader purpose at this level (see Dr. Clifford's Table of Reading Item Alignment) is to read "beyond the lines", understand the author's sociolinguistic and cultural references, follow innovative turns of thought, and interpret the text in view of its wider cultural, societal, and political setting.



The reader at this level fits Dr. Clifford's definition of a proficient reader:

Proficient reading: The active, automatic, far-transfer process of using one's internalized language and culture expectancy system to efficiently comprehend an authentic text for the purpose for which it was written. (*Dr. Clifford*)

According to Clifford, the reader at this level establishes the author's probable communicative intent by positing in real time during the reading process a series of possible interpretations of a given text and by concurrently evaluating those interpretations against his/her internalized language, culture, and real-world expectancy systems.

2. **Author Purpose:** Project lines of thought beyond the expected, connect previously unrelated ideas or concepts, and present complex ideas with nuanced precision and virtuosity with the goal of propelling the reader into the author's world of thought. (*Dr. Clifford*)

Vollman's intent and line of argumentation reflect Clifford's definition of author purpose at Level 4. The text analysis that follows will illustrate this further.

3. **Target language:** Before a final rating is awarded to a text, discourse is analysed in terms of text features, such as tone, use of abstract and nuanced vocabulary, cultural and historical concepts, innovative turns of thought etc.

NOTE: Modes of discourse and language patterns vary by language and depend on cultural norms and expectations for a particular language. For example, discourse organization in Arabic takes a circular form, while in Japanese the main idea is stated at the end of the text. English writing style tends to be more linear and terse compared to French or Spanish. Chinese writers tend to provide many examples without explicitly stating the main points of an argument. (The topic of contrastive, or intercultural, rhetoric has been discussed extensively by Robert Kaplan and, more recently, Ulla Connor.)

Text Analysis and Possible Items

The questions and answers below offer guidance in the analysis of Level 4 features of Text 2:

- Q. *What are one's first thoughts about designating a Level 4 rating to William T. Vollman's Rising Up and Rising Down: Some Thoughts on Violence, Freedom and Urgent Means.?"*
- A. Our hypothetical NATO examinee or any other general reader would probably notice that the text includes "unpredictable turns of thought" because of the author's treatment of crime and his views that the gravity of a crime may be viewed as culture or context dependent, or even dependent upon the relationship between the victim and the perpetrator.



That reader would probably also notice that his/her reading processes change when encountering a text of this difficulty level. The reader may experience a certain amount of tension trying to grasp the multi-layered meaning fully; the reader will invariably feel the text requires attentive/focused and careful reading, and that reading operations, such as skimming and scanning are of little use here.

Q. *What features make it a Level 4 text?*

A. The author begins the article by describing a concrete event and then almost startles the reader by developing an unexpected argument about Eichmann and the Nuremberg process. Through this sudden shift to a much wider picture, the author links the event of a sexual attack on a girl to a wider societal and historical framework, forcing the reader to follow the references and make his/her own linkages. The language itself is abstract and contains complex syntactical and linguistic formulations, as well as cultural, historical, philosophical and legal references, including Latin expressions. The author's tone is individualistic and insistent as he ruminates on the nature of crime within different contexts.

Q. *What kind of test questions might be viable for this text?*

A. We might think first of the kind of discussion that could possibly arise naturally between two acquaintances who had read this essay. What would they talk about or ask each other?

One might ask the other: Do you agree with this conclusion? Do you find this article controversial? Is the author really viewing the gravity of crimes as relative to each specific situation? Or, is this, in fact, a broader societal view?

Q. *How should we go about developing test questions?*

A. Follow these steps:

1. Questions should be developed in a team.
2. Team members should discuss text features and ensure they are in agreement about the level.
3. Initially, team members should begin item development individually.
4. Each team member should review all of the items/questions developed and suggest revisions or additional questions.
5. One or more team members provide should provide tentative responses to determine if there are some hidden pitfalls. For example, would more external cultural knowledge be needed than seems reasonable for Level 4? Or, does the text itself include a phrase or sentence that answers the question verbatim?
6. Team members should use their tentative responses and objections, along with other types of input, to refine the final set of questions.



7. Both the questions and the revisions should be set aside for a few days before making a final decision. One person may have to make that final decision, but all team members should be consulted. Reasons for any objections should be kept on record for possible review at later stages of the project.

Q. *What are some sample questions that could appear on an actual test if this text were selected?*

A. The following are some possible questions/items and responses. Please note that in order to help the reader focus his/her attention, it is indicated in brackets whether the question refers to the whole text or a paragraph.

NOTE: as mentioned before, responses delivered orally may be less formal and more reflective of spoken language style.

Question 1 (*whole text*): While reflecting on the event in question, the author reaches a conclusion on the unique relationship between the crime perpetrator and his victim, and the role of motive and context. According to this author's reasoning, how can identical crimes be judged differently depending on the perpetrator, the victim, and the circumstances?

Sample Response (*whole text*): the uniqueness of each relationship between an offender and his victim stems from the potential reaction of the victim to the crime being perpetrated. In one case, a devout Muslim woman might consider the act of fondling as grave a violation against her body and purity as that of a rape, while a woman from a different cultural background might be mildly offended and might good-naturedly admonish her attacker.

The author also brings in examples that evaluate the gravity of a crime in relation to the motive behind it, as well as the context or circumstances under which it is committed. Thus, what is acceptable in the times of war is not acceptable in the time of peace. In addition, committing a murder out of fear will not carry the same weight as that committed for another reason.

Question 2 (*Para 3*): How does the quote, "If he suffers, he must suffer for what he has done, not for what he has caused others to suffer", relate to the author's overall perception of the severity of crimes?

Sample Response (*Para 3*): The author refutes the veracity of the quote and mentions context as a critical factor in judging a deed. His line of argumentation is again linked to his statement that there is a unique relationship between the perpetrator and the victim. In one case, the amount of suffering from the same misdeed committed in one context is



negligible, while in another case the consequences would be very serious (the Thai versus the Muslim girl). In his view, it is the context that determines the gravity of a deed and not the deed itself. For that reason, a crime is relatively serious.

Question 4 (*Paras 8 and 9*): What do these paragraphs conclude on Eichmann's moral guilt and society's juridical interpretation of his crimes? And how may future generations utilize Eichmann's trial?

Sample response (*Paras 8 and 9*): the author states that Eichmann had used his obligation to the state and the uniform he wore during his execution of crimes as a justification. The audience listening to his testimony would have formed different impressions of his morality depending on how he presented his case. So, for example, if had admitted committing a mass murder versus having stated that he was merely upholding authority it would have shaped a different impression and reaction on part of the jury and other trial attendees. The former statement would have provoked more of a moral reaction than the latter, which would have been viewed from more of a judicial point of view. One may be swayed toward thinking that he just followed orders, or one may believe that he would have committed these crimes anyway.

If Eichmann had just blindly followed orders, one should study Nazism in order to prevent it from ever rearing its ugly head again. In the event that Eichmann acted from his own volition and character, one should attempt to learn more about that type of personality to avoid a recurrence. In either case, the author is satisfied that in Eichmann's case justice was served.

Question 5 (*whole text*): This author organizes and develops his arguments in an unusual manner. How might a reader with strong convictions and views on the issues counter these arguments?

Sample Response (*whole text*): Readers who have less of a relativist view of crimes might not welcome Vollman's giving so much emphasis to the role of context in determining the seriousness of a crime. They might argue that he provides too much legitimacy to cultural distinctions that underlie that portion of the argument concerning molestation. They might retort that a Thai woman, while living in a society that permits advertising posters with nudes in rural markets, nevertheless deserves as much sexual privacy and respect as any veiled woman in a strict Muslim country. The same would apply to a western woman in an even more open society. Vollman's insistence on relating the seriousness of the action to context would not fit into the value system of many readers who insist that all women should be treated with dignity. Some readers might find it inappropriate to wedge into a discussion of Eichmann's crime against humanity a case of sexual groping probably



feeling that each topic deserves a separate argument, to be treated with proportionate significance in terms of human values. In addition, s/he might well be offended by the attention given to examining the suffering of the perpetrator or the reasons behind the actions of an Eichmann or a Hess.



LIST OF WORKS CONSULTED

- Alschuler, C. and Moussa, N. (2002). Testing at ILR Levels 4, 4+ and 5: A Tester Training Model. Text of a Presentation to the Interagency Language Roundtable Committee, Foreign Service Institute, Washington, D.C. March 15, 2002.
- 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, Ill.: National Textbook Company, 97-106.
- Child, J.R. (1998). Language skills levels, textual modes, and the rating process. *Foreign Language Annals*, 31, 381-391.
- Child, J.R, Clifford, R., and Lowe, P. Jr. (1993). Proficiency and performance in language testing. *Applied Language Learning* 4 (1&2), 19-54.
- Clifford, R. and Cox, T.L. (2013). Empirical Validation of Reading Proficiency Guidelines. *Foreign Language Annals*, 46 (1), 45-61.



[page intentionally left blank]



BUREAU FOR INTERNATIONAL
LANGUAGE CO-ORDINATION

Annex A

Sample texts at Levels 1 - 4



[page intentionally left blank]



SAMPLE TEXTS AT LEVELS 1-4

This appendix contains sample texts at each of the base levels from Level 1 through Level 4. These texts will serve as the basis of discussion of the proficiency scale and testing.

Sample texts

LEVEL 1

Aircraft Carrier USS Hornet Museum

This World War II aircraft carrier is a national historical landmark with historic planes on deck. The museum is open from 10 a.m. until 5 p.m. daily. Pier 3, Alameda Point. (510) 521-8448. www.hornet.com

LEVEL 2

Plane Crashes; Pilot Survives

A small plane spraying insecticide on crops near Donaldston crashed Thursday. The pilot survived with only minor injuries, and no one else was on board.

The plane went down in a lettuce field about 80 yards east of Highway 90, north of Donaldston, at about 2:15 p.m., according to Police Chief Harold Quinn.

Damaged power lines in the area may have been a factor, Quinn said. The Federal Aviation Administration will conduct an investigation.

The pilot, whose name was not provided by police, was taken to City Hospital in Donaldston where he was treated and released.



LEVEL 3

This is an editorial from 2006.

A Real-World Army

Military reality finally broke through the Bush administration's ideological wall last week, with President Bush publicly acknowledging the need to increase the size of the overstretched Army and Marine Corps.

Larger ground forces are an absolute necessity for the sort of battles America is likely to fight during the coming decades: extended clashes with ground-based insurgents rather than high-tech shootouts with rival superpowers. The president's belated recognition is welcome, though it comes only after significant damage has been done to the Army's morale, recruitment standards and fighting readiness. Given the time required to recruit and train the additional troops, the proposed increase will not make much difference in Iraq's current battles. But over time it will help make America more secure and better prepared to meet future crises.

The need for more troops has for some time been obvious to Americans. They have heard from neighbors or from news reports of tours of duty involuntarily extended, second and even third deployments to Iraq, lowered recruiting standards and members of the National Guard and Reserves vowing to get out. That is the inevitable consequence of trying to squeeze out an additional 160,000 soldiers for Iraq and Afghanistan year after year without significantly increasing overall ground forces.

But it took the departure of Donald Rumsfeld – the author of the failed Iraq policy and the doctrine of going to war with less than the Army we needed – for Mr. Bush finally to accept this reality.

There is no permanent right number for the size of American ground forces. The current size – just over 500,000 for the active duty Army and 180,000 for the Marine Corps – is based on military assessments at the end of the cold war. As the world changes, those assessments must be constantly reviewed. When the 21st century began, Pentagon planners expected that American forces could essentially coast unchallenged for a few decades, relying on superior air and sea power, while preparing for possible future military competition with an increasingly powerful China. That meant investing in the Air Force and Navy, not the Army and Marines.

Then 9/11 changed everything, except the Pentagon mind-set. During the Rumsfeld years, reality was subordinated to a dogma of "transformation," which declared that with a little more technology, the Army could do a lot more fighting with fewer soldiers than its senior generals believed necessary.

Every year since 2001 has brought increased demands on America's slimmed-down and dollar-starved ground forces, while billions continued to flow into sustaining the oversized and underused Air Force and Navy, and modernizing their state-of-the-art



equipment. As a result, the overall Pentagon budget is larger than it needs to be, while the part going to overtaxed ground forces is too small.

Increasing those ground forces will cost roughly \$1.5 billion a year for every 10,000 troops added, as well as tens of billions in one-time recruitment and equipment expenses. But American can afford it and it can be done without any significant increase in the annual military budget.

For example, the estimated \$15 billion a year (plus start-up costs) needed to add 100,000 more ground troops could easily be found by slashing military pork and spending on unneeded stealth fighters, stealth destroyers and attack submarines, and by trimming the active duty Air Force and Navy to better reflect current battlefield requirements.

Over time, bigger ground forces will mean more sustainable troop rotations, fewer overseas deployments of the National Guard and better battle field ratios of American to enemy fighters. That is the least America owes to the men and women who risk their lives to keep us all more secure.

LEVEL 4

This text was written during the cold war era.

It is said that Woodrow Wilson was shocked to learn, in 1915, that the Army War College was studying plans for wars against other countries. The story, sometimes cited as an example of Mr. Wilson's naivete, seems to be the somewhat distorted version of a real episode. But it is suggestive of a historically significant reality. When a military planner selects another country as the leading hypothetical opponent of his own country - the opponent against whom military preparations and operations are theoretically to be directed - the discipline of his profession obliges him to endow that opponent with extreme hostility and the most formidable of capabilities. In tens of thousands of documents, this image of the opponent is re-created, and depicted in all its implacable formidability, until it becomes hopelessly identified with the real country in question. In this way, the planner's hypothesis becomes, imperceptibly, the politician's and the journalist's reality. Even when there is some degree of substance behind the hypothesis, what emerges is invariably an overdrawn and distorted image.

And so it has been in the case of the Soviet Union, with the result that what began as a limited political conflict of interests and aspirations has evolved into a perceived total military hostility; and what was in actuality a Soviet armed-forces establishment with many imperfections and many limitations on its capabilities has come gradually to be perceived as an overpowering paragon of military efficiency, standing at the



beck and call of a political regime consumed with no other purpose than to do us maximum harm. This sort of distortion has magnified inordinately in the public eye, the dimensions of what was initially a serious political problem, and has created, and fed, the impression that the problem is one not to be solved otherwise than by some sort of a military showdown.

Apply the trisection to the sample texts

LEVEL 1	
CONTENT	The text consists of very simple connected written material. This is an announcement with a brief description of a museum. The announcement gives the times that the museum is open and its location along with a phone number and website.
TASKS	A Level 1 reader would be expected to understand the basic meaning of a Level 1 text. The intended reader of this text would be able to find specific information such as opening and closing times, location, and purpose of the museum.
ACCURACY	The Level 1 reader would understand the basic message and would often be able to guess the meaning of an unfamiliar word, such as “landmark,” from context. However, a Level 1 reader may not understand every Level 1 text that s/he encounters.

LEVEL 2	
CONTENT	The text is concrete and factual, and it narrates past events. It is an authentic news item about a frequently occurring action; that is, an accident. Such events are usually described in a predictable way, and this predictability aids the Level 2 reader in understanding the text.
TASKS	The Level 2 reader can locate and understand the main ideas and details. The intended reader would be able to answer factual questions about the text, such as the location of the plane crash, a possible cause, and any injuries.
ACCURACY	The Level 2 reader would be able to read this text with understanding, although s/he may be slow in comprehending some texts of this type that s/he encounters.



LEVEL 3	
CONTENT	This text is an editorial in a major publication for educated native readers. While the subject matter was familiar to native readers in the 2006 timeframe, the editorial supports a specific point of view through discussion of abstract concepts. The basis for the author’s argument involves competing ideas about the US military’s future. Relatively concrete information is used to support evaluative language (such as “sustaining the oversized and underused Air Force and Navy, and modernizing their state-of-the-art equipment”).
TASKS	The Level 3 reader should be able to learn through reading. The native reader would be able to gain information about the results of the previous and new policies. Even the non-native reader who might be unfamiliar with the term “military pork” could probably determine its meaning from the author’s examples. The intended reader would be able to follow the argumentation and supported opinion, to relate ideas, and to understand implicit information (such as the indication that US security was reduced by budgetary decisions that made it difficult to retain and recruit soldiers).
ACCURACY	The Level 3 reader should be able to read this text with almost complete comprehension and follow the line of argument made by the author.

LEVEL 4	
CONTENT	These are the introductory paragraphs in a text found in a periodical for well-educated readers. The author juxtaposes an anecdote from early in the twentieth century with his/her own original conclusions about the exaggeration of the Soviet threat to the west during the mid-twentieth century. According to the author, military planners designed and developed this outsize image of the Soviet Union and convinced politicians, journalists, and, implicitly, the public to accept it. Interestingly, s/he labels the planners’ work a “hypothesis” without revealing that his own explanation is itself a hypothesis. This framework makes the text somewhat more abstract than it might be if the ideas were presented differently.
TASKS	The intended reader will need to follow unpredictable turns of thought and understand highly abstract concepts. The Level 4 reader will need to show both global and detailed understanding of the concepts presented by the author. S/he would benefit from having had some exposure to the views of Woodrow Wilson and the period in which he was US President. In addition, s/he would need to know enough about Cold War rhetoric to apply the author’s argument to some of the language prevalent then.
ACCURACY	The Level 4 reader should show strong competence in following the presentation of an original set of ideas.

Determine some ways to develop test questions for the text at each level

First, we might consider why each text might interest the intended reader and, possibly, how that reader might discuss or refer to it with another person.

Then, we can try to use that information to develop test questions in alignment with the level of the text.



LEVEL 1

This text could interest a person who was scanning newspaper listings of local events to plan an outing with family or friends. S/he might focus on such points as the purpose of the museum, the hours it is open, and its location. The reader might convey these points to others as they decided whether to go.

Typical questions in a constructed-response test would be:

- *When can you visit the museum?*
- *What can you see at this museum?*

These Level 1 questions are in alignment with the Level 1 text.

LEVEL 2

This text would probably attract the attention of a reader who was interested in news about the Donaldston area, as well as readers specifically interested in general aviation or plane crashes. This reader might want to pass the information along to others with similar concerns.

Typical questions in a constructed-response test would be factual:

- *What was reported from the Donaldston area?*
- *How many people were involved?*
- *Has the cause been determined?*
- *What was the exact location?*
- *When did this happen?*

LEVEL 3

This text would attract the interest of readers concerned about military preparedness in the US and could spark debate between those who accept its conclusions and those who disagree.

Appropriate questions in a constructed-response test would involve a linguistically abstract presentation of the issues, requirements to compare and contrast approaches, or demands for synthesis. These might include:

- *What contrast is made in this editorial between recently announced changes in US military philosophy and that of previous years?*
- *What is the position of the author?*
- *How is the author's point of view revealed?*
- *What opinion does the author ascribe to the American people?*

LEVEL 4



Bearing in mind that the text in this appendix represents only the introductory paragraphs of a longer text, we can surmise that both the full and the partial text would be appreciated by readers interested in US military policy, as well as students of propaganda. It could attract people who share the author's point of view but also those who disagree with his/her conclusions and intend to dispute them. Such readers would probably seek out other people who have read the text and elicit their responses.

Appropriate questions in a constructed-response test would require processing highly abstract concepts and comprehension of cultural references. These might include:

- *What is the function of the author's reference to Woodrow Wilson in this text?*
- *What impression does the author create of military planners in this text?*
- *Which word or words best convey the author's attitude toward the Cold War planners?*
- *What are the competing images of the Soviet military within this text?*



[page intentionally left blank]



BUREAU FOR INTERNATIONAL
LANGUAGE CO-ORDINATION

Annex B

Samples of typical Level 4 prose



[page intentionally left blank]



SAMPLES OF TYPICAL LEVEL 4 PROSE

Below are a few samples illustrating typical Level 4 prose.

Text 1

American political legitimacy has to be reestablished each time the presidency changes hands. It isn't simply because presidents write, produce and perform plays about the meaning of national life, it being hard to start over with a new script when it feels as if one has been stranded in the middle of the old one. For all of the sophistication of a Constitution which is at once a machine and a nearly Cubist work of modern art, America remains in some ways a politically primitive country: a country less than fully convinced that, as a national polity, it even exists, or, if it does, that it should. America remains a young country: itself a strange land, with a before and an after. The country was invented. The country may or may not have existed as a society before it existed as a polity, but the fact of its invention overwhelms all that is social and, by so doing, deprives the political of the social bases of its legitimacy. Other societies can pull their social history, which in the case of, say, Japan, China, Russia, France, Italy, Germany, Ethiopia, Mexico or Great Britain, seems to extend back into the mists of time, into the unrecorded and the unknowable, into its political history, and the two are not completely separable. The Constitution does not countenance the social, or even the American specific. It makes certain fundamental assumptions about the basic motives that govern all human beings and then attempts to design a censorious but self-renewing set of rules and procedures by which government might protect citizens from one another and citizens might be protected from government. The arrangement is neither ordered nor blessed by God; it presents itself as the product of reasoned suspicion and perhaps irrational hope. There is no State, in the European sense: an apparatus of power that carries its own principle of authority. The only authority given the national government of the United States of America is the consent of the governed. At any given moment, any given American can and likely will feel that his or her consent has not been given, or that it should be withdrawn.



Text 2

Excerpt from "Politics among Nations" by Hans J. Morgenthau

When one reflects upon the development of American thinking on foreign policy, one is struck by the persistence of mistaken attitudes that have survived- under whatever guises-both intellectual argument and political experience.

Once that wonder, in true Aristotelian fashion, has been transformed into the quest for rational understanding, the quest yields a conclusion both comforting and disturbing: we are here in the presence of intellectual defects shared by all of us in different ways and degrees. Together they provide the outline. of a kind of pathology of international politics. When the human mind approaches reality with the purpose of taking action, of which the political encounter is one of the outstanding instances, it is often led astray by any of four common mental phenomena: residues of formerly adequate modes of thought and action now rendered obsolete by a new social reality; demonological interpretations of reality which substitute a fictitious reality-peopled by evil persons rather than seemingly intractable issues-for the actual one; refusal to come to terms with a threatening state of affairs by denying it through illusory verbalization; reliance upon the infinite malleability of a seemingly obstreperous reality.

Man responds to social situations with repetitive patterns. The same situation, recognized in its identity with previous situations, evokes the same response.

The mind, as it were, holds in readiness a number of patterns appropriate for different situations; it then requires only the identification of a particular case to apply to it the preformed pattern appropriate to it. Thus the human mind follows the principle of economy of effort, obviating an examination de novo of each individual situation and the pattern of thought and action appropriate to it. Yet when matters are subject to dynamic change, traditional patterns are no longer appropriate; they must be replaced by newness reflecting such change. Otherwise a gap will open between traditional patterns and new realities, and thought and action will be misguided.



Text 3

In many respects the work of the School of Bologna represents the most brilliant achievement of the intellect of medieval Europe. The medieval mind had, indeed, a certain natural affinity for the study and development of an already existing body of Law. The limitations of its knowledge of the past and of the material Universe, were not, to any appreciable extent, a bar to the mastery of a Science which concerns itself simply with the business and relations of every-day life. The Jurist received his Justinian on authority as the Theologian received the Canonical and Patristic writings, or the Philosopher his Aristotle, while he had the advantage of receiving it in the original language. It had only to be understood, to be interpreted, developed, and applied. The very tendencies which led men of immense natural powers so far astray in the spheres of Theology, of Philosophy, and still more of Natural Science, gave a right direction to the interpretation of authoritatively prescribed codes of law. An almost superstitious reverence for the *littera scripta*; a disposition to push a principle to its extreme logical consequences, and an equally strong disposition to harmonize it at all costs with a seemingly contradictory principle; a passion for classification, for definition and minute distinction, a genius for subtlety--these, when associated with good sense and ordinary knowledge of affairs, are at least some of the characteristics of a great legal intellect. Moreover, the exercises which were of such doubtful utility in other branches of knowledge formed an excellent course of legal education. The practice of incessant disputation produced a dexterity in devising or meeting arguments and a readiness in applying acquired knowledge, of comparatively little value to the student of History of Physical Science, but indispensable to the Advocate and even to the Judge. While it fostered an indifference to the truth of things fatal to progress in Theology or Philosophy, it gave the pleader the indispensable faculty of supporting a bad case with good, and a good case with the best possible arguments.

In estimating the place of the Civil Law in the history of medieval culture, we must carefully distinguish between its cultivation as a science and its pursuit as a profession. During the most brilliant period of its cultivation as a science its Professors were almost all congregated in Bologna itself. That period embraces the century and a half after its revival by Irnerius. It was in the hands of the 'Glossators'--of Irnerius, of the famous 'Four Doctors'..., of Rogerius, Placentinus, Azo, and Hugolinus - that the most real progress was made. The works of these men are, perhaps, the only productions of medieval learning to which the modern Professor of any science whatever may turn, not merely for the sake of their historical interest, not merely in the hope of finding ideas of a suggestive value, but with some possibility of finding a solution of the doubts, difficulties, and problems which still beset the modern student.

**Text 4**

Used by James Child in his training in the 1980s. (Good example of an extended metaphor.)

In this season of apostasy, the high priests of economics face a bitter challenge. As soothsayers they have lamentably failed, showing themselves quite as fallible as the Greek oracles or the readers of sheep's entrails. As practitioners, their cures have often proved worse than the disease.

During the long period of sustained innocence that lasted through the '50s and '60s, we basked in the happy conviction that economics was a developed science. The revered prophet, John Maynard Keynes, had taught that we could, by careful demand management, render business cycles as flat as a desert highway, so that, by keeping the faith, we might look forward to a long future of steady economic growth. Economists were, they assured us, on the verge of mastering man's economic ills just as the medical doctors were mastering disease. We assumed, in other words, that economics had reached a level of advancement roughly equivalent to the sulfanilamide age in medicine.

No wonder we were shocked when we found that the state of economic science more nearly paralleled Dr. Harvey's discovery of the circulation of blood. To be sure, that disturbing insight has not stopped the production of new economists; doctoral candidates are still filling thousands of pages with opaque formulas and devising elaborate charts that strikingly resemble a phrenologist's schematic drawing of the human cranium. But honest practitioners no longer try to hide their dubiety; many are quite frankly resorting to leeches and poultices.

At the moment, the doctors are prescribing protracted bleeding in the hope of curing the American economy of the high fever of inflation, but, as is traditional with doctors, they stridently disagree. In some quarters there is a great vogue for Dr. Milton Friedman, who is busy prescribing not only for Ronald Reagan but also for Margaret Thatcher. I am grateful to Friedman, for it was he who in the fall of 1974 reassured me that I need not worry about high oil prices, since inflation would render them nugatory while OPEC would promptly fall apart. Today, he is as single-minded as a naprapath. Naprapathy holds that human disease can be cured by adjusting the ligaments; Dr. Friedman would cure the economic dislocations by regulating the money supply.

New healers are consistently emerging, such as the young Dr. Laffer, who asserts that we could painlessly cure our ills by cutting taxes - a remedy enthusiastically echoed by any number of lay practitioners. But some, like the venerable Dr. Hayek, sturdily adhere to the true faith of Adam Smith, refusing, like the legendary medieval priest,



to change his mumpsimus for the new fangled sumpsimus. Finally, a zealous core of true believers still fall back on the time-proven quackery of Dr. Karl Marx.

Disagreement persists even as to the most fundamental issues; thus economic model builders vie with their more classical confreres just as the ancient School of Cnidos, which concentrated on symptoms from the patient, was disputed by the School of Hippocrates, which studied causes and held that disease was the improper balancing of blood, phlegm, yellow bile and black bile.

To an ancient like myself who was in Washington early in the New Deal, all this resembles the orgy of theorizing and experimentation practiced on the American economy by President Roosevelt's changing stable of PhDs. No one knew what nostrums, if any, would work, but that did not discourage us from trying. Today, the American public is playing the same guinea pig role as in the '30s.

Though I have no special brand of snake oil to offer, I have faith that our economy is sufficiently robust to survive. Yet, I cannot forget the infant Louis XV who, after contracting small pox, was saved from death only because his nurse hid him from the ministrations of the doctors whose vigorous attentions killed both his brother and his father. For Americans there is no place to hide, nor, unlike medical doctors, can economists even be sued for malpractice.

No wonder that the patient citizen is worried, baffled and frustrated by the different advice and indifferent results of economists who are equally worried, baffled and frustrated. We are still decades away from economic antibiotics.

Humility is, however, the first requisite to understanding, and it is comforting to discover economists who are at least privately admitting that they no longer have full confidence in their own omniscience.

It is a good sign and the beginning of wisdom. They would do well to recall the famous comment of the distinguished medical doctor, the senior Oliver Wendell Holmes: "If the whole *material medica*, as now used, could be sunk to the bottom of the sea, it would be all the better for mankind—and all the worse for the fishes." Unfortunately, Holmes's prescription could never be applied to our economic text books: it would violate the Clean Water Act.



Text 5

The hermeneutic motion, the act of elicitation and appropriative transfer of meaning, is fourfold. There is initiative trust, an investment of belief, underwritten by previous experience but epistemologically exposed and psychologically hazardous, in the meaningfulness, in the “seriousness” of the facing, or strictly speaking, adverse text. We venture a leap; we grant *ab initio* that there is “something there” to be understood, the transfer will not be void. All understanding, and the demonstrative statement of understanding which is translation, starts with an act of trust. This confiding will, ordinarily, be instantaneous and unexamined, but it has a complex base. It is an operative convention which derives from a sequence of phenomenological assumptions about the coherence of the world, about the presence of meaning in very different, perhaps formally antithetical semantic systems, about the validity of analogy and parallel. The radical generosity of the translator (“I grant beforehand that there must be something there”), his trust in the “other,” as yet untried, unmapped alterity of statement, concentrates to a philosophically dramatic degree the human bias towards seeing the world as symbolic, as constituted of relations in which “this” can stand for “that,” and must in fact be able to do so if there are to be meanings and structures.



BUREAU FOR INTERNATIONAL
LANGUAGE CO-ORDINATION

Annex C

Terminology relating to the analysis of Level 4 reading texts



[page intentionally left blank]



Abstract topics: (*as opposed to concrete topics*); topics dealing with theoretical or ideational aspects of issues (for example: global implications of free-trade agreement).

Argumentation: process of providing supporting evidence when discussing an issue or expressing an opinion, such as the acts of affirming, denying, agreeing with, opposing or justifying one's stance or views.

Comprehension between the lines (inferential comprehension): the ability to comprehend information that is not explicitly expressed by the spoken or written text but which can be derived or assumed on the basis of the text.

Comprehension beyond the lines (evaluative comprehension): the ability to make judgements or express opinion about a text, evaluate the significance of the author's message, credibility, intent, and purpose, extrapolate beyond the text and place it in a socio-cultural and historical context.

Concrete topics: (*as opposed to abstract topics*); referring to physical, tangible, material objects; topics that can be discussed using factual and descriptive language (job, house, family, hobbies, etc.).

Description: linguistic task focusing on physical characteristics of an object.

Elaboration: detailed, extensive and thorough development or explanation of a thought or idea.

Explicit information: (*as opposed to implicit or implied*); information that directly states or expresses all that is meant and intended.

Extended discourse: a stretch of meaningful and organized language providing a comprehensive or detailed treatment of a topic.

Facility: ease or dexterity by which language is used.

Factual content: information based on facts and not on opinions, theories or other abstract postulates.

Familiar topic: (*as opposed to unfamiliar topic*); experiences or subject matter encountered by the individual before or well known to him/her (one's job, family, etc.).

Function: communicative task or competency showing what the individual can do using a language (for example: give instructions/directions, request service, etc.).



High-frequency vocabulary and structural patterns: (*as opposed to low-frequency vocabulary and structural patterns*); vocabulary, expressions or grammatical structures that occur often in language use; common, everyday vocabulary.

Hypothesize: communicative act by which one proposes or speculates about a potential outcome, implications, or consequences of an issue being treated from an abstract perspective.

Idiosyncratic language use: language use that is highly personal and peculiar to an individual.

Implicit information: (*as opposed to explicit*); information which is not expressly stated, but implied or hinted at indirectly.

Inference: the act of arriving at conclusions or making deductions from given facts.

Innovative turns of thought: presentation of ideas, opinions and concepts in a way that is new, original and, therefore, often surprising or unanticipated.

Low-frequency vocabulary and structural patterns: (*as opposed to high-frequency vocabulary and structural patterns*); vocabulary and structures not occurring often or with a high degree of predictability in language use.

Malapropism: the, usually deliberate, use of an incorrect word in place of a word with a similar sound, resulting in a nonsensical, often humorous utterance (for example: “The flood damage was so bad they had to evaporate (evacuate) the city.”)

Narration: oral or written account of events (storytelling) organized in a manner that is often chronological and logical. This will usually show a clear development of between events.

Non-routine topics: (*as opposed to routine topics*); non-predictable, unfamiliar topics that may transcend the realm of one’s own experience.

Paragraph: a group of chronologically, thematically and logically related sentences, thoughts or ideas.

Pragmatic inference: a conclusion, deduction, or opinion that is formed based on real-world knowledge or experience.

Projective mode: a writer’s stance intended to convey a highly individual approach to a message.

Propositional inference: a conclusion, deduction, or opinion that is formed based on a statement to be proved, explained, or discussed.



Register: form or level of language used according to the social context or specific content matter, for example, colloquial, informal, highly formal, professional.

Rhetorical devices: language used in a certain way to persuade or to evoke an emotion on the part of the reader or audience.

Shaping: the thoughtful, selective and purposeful use of language that often becomes the signature of the particular author.

Spoonerism: an error in speech or deliberate play on words in which corresponding consonants, vowels, or morphemes are switched between two words in a phrase (for example: ‘a blushing crow’ instead of ‘a crushing blow’).

Stylistic devices: linguistic features (vocabulary, structures) employed in different language use contexts to indicate specific manner (style) of expression (formal, official, technical, literary, legal, informal, intimate, etc.).

Stylistic level: manner of oral or written expression indicating various degrees of formality or informality.

Subject matter: topics treated orally or in writing.

Supported opinion: viewpoint well explained, justified and substantiated with reasoned arguments.

Tailoring: adaptation or adjustment of language (in terms of style, register, etc.) according to context and audience.

Transition: a word or group of words that connects ideas and shows relationships in sentences and in paragraphs (for example: ‘eventually’, ‘nonetheless’, ‘consequently’, etc.).

Unfamiliar topic: (*as opposed to familiar topic*); topic that that may transcend the realm of one’s own experience.



[page intentionally left blank]



**BUREAU FOR INTERNATIONAL
LANGUAGE CO-ORDINATION**

Appendix 1

Prototype Level 4 Reading Test

Test Specifications

(draft)



BUREAU FOR INTERNATIONAL LANGUAGE COORDINATION



[page intentionally left blank]



1. BACKGROUND

The Working Group (WG) on Level 4 Proficiency was formed at the 2010 BILC Conference in Istanbul. Its goal was to develop a conceptual framework of Level 4 language proficiency, amplify the STANAG 6001 Level 4 descriptors in all four skills, and develop a prototype Level 4 test as a model, which nations could use as the basis of their national Level 4 test development.

The English language proficiency requirement of a number of posts at NATO HQs is set at Level 4. One of the goals of a Language Need Analysis (LNA), conducted in 2015 was to establish whether the Standardized Language Profiles (SLPs) of posts designated as Level 4 were, indeed, Level 4. The assumption was that most of these posts do not require incumbents with Level 4 proficiency; however, it was conceivable that there were a few positions that warranted such a high SLP in some skills. Within the LNA framework, a proficiency test developed in accordance with the test specifications, as set out in this document, was used, as a pilot test, to assess whether the incumbents selected for, or already occupying these posts had the required proficiency.

For practical reasons, the WG decided to begin this development with a test of the reading skill. A test of reading proficiency seems to be somewhat easier to develop and operationalize than a test of the more complex listening skill. In addition, Level 4 written texts can be found more readily than Level 4 audio texts.

Experience and research have shown that many foreign language learners master the reading skill faster than the other three skills, so that finding an appropriate trialling test population would not be a significant challenge. A number of military institutions in Europe deliver officer training content in English, in addition to separate English language lessons. For that reason, officers are exposed regularly to higher level texts in English and attain high reading proficiency levels. Also, with the current prevalence of on-line and print articles written in English, it is assumed that out of personal interest military officers, just like the general population, are exposed to a multitude of English reading texts, even though the content may not always be of Level 4 difficulty.

The knowledge gained from the development and trialling of this test will be applied to the development of tests in the other three skills.

2. GENERAL TEST OVERVIEW

Purpose of the Test

This test is designed as a prototype that will be used to illustrate how Level 4 language proficiency might be measured IAW NATO STANAG 6001, Ed. 5v2. Because language testing within the NATO context is the responsibility of individual nations, those nations that need to test at this level, and feel under-resourced for the development of their own tests, may follow this testing model in order to meet the requirements of practicality, reliability and validity. It is expected that after it has been trialled and validated, the format and testing protocol of this test may be transferrable to languages other than English.

3. TEST CONSTRUCT

The construct, or theoretical model of the reading ability at Level 4, is derived from the NATO STANAG 6001, Ed. 5v2 descriptor of the reading skill:

Level 4 – Expert

Demonstrates strong competence in reading all styles and forms of the written language used for professional purposes, including texts from unfamiliar general and professional-specialist areas. Contexts include newspapers, magazines, and professional literature written for the well-educated reader and may contain topics from such areas as economics, culture, science, and technology, as well as from the reader's own field. Can readily follow unpredictable turns of thought on any subject matter addressed to the general reader. Shows both global and detailed understanding of texts including highly abstract concepts. Can understand almost all cultural references and can relate a specific text to other written materials within the culture. Demonstrates a firm grasp of stylistic nuances, irony, and humour. Reading speed is similar to that of a native reader. Can read reasonably legible handwriting without difficulty.



As per the amplification of the descriptor done by the WG (*NATO STANAG 6001 Level 4 Language Proficiency – A Conceptual Model and Implications for Testing*, April, 2013), the reader at this level can readily understand all forms of written language, including lengthy, highly abstract, structurally or linguistically complex texts dealing with unfamiliar general and professional topics. In these texts, the reader appreciates subtle distinctions of style and demonstrates strong competence in identifying finer points of detail, including attitudes, author tone, implied opinions, and subtle references to cultural phenomena, such as religion and politics.

A Level 4 reader can read highly individualized literary and non-literary writings that show the author's virtuosity with language, often by mixing registers (e.g., formal and informal), evincing tone (e.g., humour, irony, sarcasm), and generally challenging the reader to follow innovative turns of thought. The author assumes the reader shares target-language culture at a high level. In order to be capable of fully constructing meaning from such texts, a Level 4 reader possesses a high level of linguistic and extra-linguistic knowledge, as well as the ability to read and think critically, and employ general intellectual reasoning strategies. A Level 4 reader understands subtle variations and complex syntax typical of creative language, including the use of metaphors. Language control may not be complete when rarely used structures and/or idiomatic uses of grammar are involved.

At this level of proficiency, target culture is internalized to such an extent that the reader can read "beyond the lines", i.e. understand the full ramifications of texts as they are situated in the wider cultural, political, or social environment. A Level 4 reader can fully utilize these external references to better understand the wider socio-cultural or political ramifications of the author's assertions. A reader at this level is able to make evaluative judgements about the author's assertions, and may recognize logical fallacies, hidden bias and assumptions, as well as the author's ulterior motives. S/he can follow unpredictable turns of thought readily, for example, in editorial, conjectural, and literary texts in any subject matter area directed to the general reader. Nevertheless, some limitations in comprehension will occur if texts deal with issues and ideas that are narrowly bound to the socio-linguistic or regional context.

4. TEST TARGET POPULATION

Military and civilian defence personnel selected for NATO HQ positions requiring a STANAG 6001 Level 4 for reading.



5. CHARACTERISTICS OF TEST POPULATION

Generally, examinees at this level have formal education that exposed them to texts of Level 4 difficulty (see *Overview of Text Characteristics*, Table 3 of the Tutorial) and have demonstrated higher order cognitive skills, such as analysis, synthesis, extrapolation, etc. They will have already demonstrated in classroom/work settings that at times they can deal with complex authentic texts written by native writers for native English readers who share the wider cultural and societal context to which the ideas expressed in these texts belong. These examinees possess a strong ability to learn about new ideas through reading in a foreign language.

6. TRIAL TEST POPULATION

The prototype test would be piloted in two phases:

Phase 1: Among native and non-native speakers of English who work within different MODs as English language teachers/testers. These individuals are known to the WG members as high level readers/speakers.

Phase 2: Among military personnel who have been posted to or have returned from NATO HQ postings requiring a Level 4 for reading in English, or who have been identified by their teachers as suitable candidates for the trialling.

All examinees will be given a code number. Information, such as their country of origin, educational background etc. will also be captured.

7. TESTERS/RATERS

Testers/raters selected to administer the prototype reading test to the trial populations, or to act as independent reviewers, will receive training. The duration/location and method of training will be determined later. All testers will be highly qualified testers or teachers, who can be trained as testers. They may or may not be members of the Working Group.



8. TEXT TYPES

Articles and essays intended to persuade, analyze critically, or develop a unique approach to a complex subject will be selected. In accordance with Clifford's *Text Type Characteristics* for the projective mode (see Table 5 of the Tutorial), the author's presence, tone, and point of view will be clear to higher level readers. For trialling purposes, texts between 1200-1300 words in length will be used. The final text length will be determined after the first trialling has been conducted. Appropriate text editing will be extremely important. The texts will normally include the full title as used in the original publications, although there could be an occasional exception.

9. TOPICAL CONTENT

Content areas for reading items are targeted to the general reader and may include:

- military and security issues
- military ethics/philosophy of war/military diplomacy
- economic and political matters
- scientific and technical topics
- cultural and social issues

Texts with military topics will not be job-specific, but intended for the general well-educated reader.

10. TEXT SOURCES

Texts will be taken from online or print publications known for publishing high level texts in the "Projective Mode" (see Clifford's *Overview of Text Characteristics*, Table 3 of the Tutorial). Such texts may be found in specialized journals, for example, the *Canadian Military Journal* and the *Journal of Political Philosophy*, or in higher level general interest publications, such as *The New Yorker*.

All submitted texts will undergo the text rating procedure by all WG members. Specific reading tasks and accuracy requirements will be based on the STANAG 6001 standard.

Texts selected will be comprehensible to Level 4 readers regardless of where they acquired or regularly use the English language.



11. TEST FORMAT

- The prototype test will be a single-level test comprised of two (2) authentic reading texts. Two equivalent versions of this test may be developed.
- The test will be delivered through two modalities, that is, reading comprehension will be evaluated through the speaking or writing modality.
- Each text will be followed by six (6) questions.
- The total number of questions/items per version will be twelve (12).
- In addition to the 12 core questions, the testers administering the test via the speaking modality may also ask follow-up questions. The purpose of the follow-up questions is to:
 - a) elicit more evidence of comprehension in case the examinees' responses to the core questions are not precise, complete or extensive enough;
 - b) redirect discussion to the task in case the examinees deviate from it in their response or do not address it fully or adequately.
- Depending on how content-rich a text is or how comprehensive the examinees' responses are, the number of follow-up questions may vary from one candidate to the next.
- The questions may be formulated to resemble the core questions or may be stated as 'encouragers', such as "could you be more specific?", "please tell us more", etc.
- All passages and questions will be given in English.

12. TEST TASKS AND ACCURACY EXPECTATIONS

The test tasks will be derived from the STANAG 6001 Level 4 descriptor and adapted to the testing situation. The tasks will be extracted from the construct statement and may include, but will not be limited to:

- follow unpredictable turns of thought on any subject matter addressed to the general reader;
- show both global and detailed understanding of texts including highly abstract concepts;
- extrapolate;
- understand cultural references and allusions;
- understand tone, such as irony, and humour;
- understand linguistic and stylistic devices used by the author for persuasive purposes.

Some of the sub-skills tested will be:

- understand the author's choice of words in relation to the tone and persuasive stance;
- link ideas from various parts of the text;
- interpret the text in its wider cultural and societal context.



It is not expected that even at this high level of proficiency, accuracy will be complete. Accuracy will be predicted using a modified Angoff validation procedure, in which a panel of expert judges estimates the expected percentages of examinees that will meet the minimum accuracy requirements at the threshold target level.

13. DELIVERY

In its trialling stage, this test will be administered face-to-face (speaking modality) where possible or through a videoconferencing system/Skype by two trained testers/raters. When administered via the written modality, test protocols laid out in the instructions for the test administrators and test-takers will be followed.

14. TEST ADMINISTRATION

- Examinees taking the test *via the speaking modality* will be given 5 minutes to read the instructions. The instructions will state that they will have 30 minutes to read each text (an average native reader reads 200 wpm). They will see the questions along with the text. After 30 minutes, they will respond to the questions orally before spending another 30 minutes reading text 2 and responding to the questions. In other words, they will work with one text and set of questions at a time. At their discretion, the testers may ask follow-up questions in order to seek clarification and engage examinees in a discussion about the texts.
- Examinees will be allowed to underline or mark the texts, take brief (bullet style) notes on a separate note pad capturing the main points of the texts, and/or formulate in point form their intended response to the 2 core questions. The final reading and text processing time including the note taking will be determined after the trialling. Examinees must not read responses verbatim or read out full paragraphs, but engage interactively with the testers. Test instructions will emphasize this aspect of the testing format.
- Examinees will be allowed to consult the text while answering the questions.
- Total test administration time is anticipated to be about 75 minutes excluding the instructions. Final suggested text length and time allotment will be determined after the first trialling session.
- Testers' questions and examinees' responses will be recorded.
- Lexical aids will not be allowed as examinees at this level are expected to have near-native reading proficiency. In case there is a word they do not understand and cannot guess from context, examinees may address the issue openly with the testers without being penalized by a lowered score, provided they demonstrate good overall understanding of the text by addressing the given task successfully.
- Examinees taking the test *via the writing modality* will have 150 minutes to complete it without using any lexical aids.



NOTE: if this testing model is adopted, testing organizations will need to decide which of the two modalities to administer. Because the number of potential Level 4 examinees is assumed to be low at any given point, the test administration via the speaking modality is practical enough. Considering the complexity of testing at this level, and the cognitive and linguistic requirements placed on test-takers, allowing a test modality choice would add a layer of fairness to the test and increase its face validity. The Table below captures some of the advantages and disadvantages of the two testing modalities:

Test Modality	PROS	CONS
Written response	<ul style="list-style-type: none"> • Raters can analyze output repeatedly. • Rating may appear more objective to the raters in the presence of visual evidence. • Rating written output is less time-consuming than re-listening to recordings. • Examinees may feel less stress and pressure while structuring/refining their response in writing, because they may feel more in control of the time they spend on individual responses. • Examinees may feel more confidence in showing evidence of proficient reading if they believe writing is their stronger productive skill. • Test administration only requires the presence of invigilators. 	<ul style="list-style-type: none"> • Raters may strain to read written output (illegible writing if hand-written). • Raters have no possibility of intervening with follow-up questions if examinees stray off the topic. • Raters may be more lenient/biased toward written responses as they tend to be more formal. • Examinees may focus more on form than the substance of their response that would provide evidence of reading comprehension. • Examinees who are poor spellers may feel disadvantaged if the test is administered in the pen and paper form. Can have spell check.
Spoken response	<ul style="list-style-type: none"> • Testers can redirect examinees' responses with follow-up questions in case the test-takers stray off the topic. • Some test-takers may feel more comfortable responding to questions orally. • Test-takers may find the oral test more authentic than the written test format (higher face validity). 	<ul style="list-style-type: none"> • Raters may find rating time-consuming/less practical (via recordings) and the outcomes of rating the spoken versus written production more subjective. • Testers may find the test administration impractical (90 minutes per test-taker approximately). • Some examinees may feel nervous/uncomfortable speaking.

Table 1: Advantages and disadvantages of two test modalities.



15. SCORING/RATING

The test will be administered and rated by well trained and carefully normed testers who will be native or near-native speakers of English. Candidates' oral responses will be recorded for further evaluation by at least another independent reviewer. A detailed answer key will be developed providing a variety of possible sample responses that would be considered as adequate for demonstrating comprehension of the text. To be considered successful, all responses must be supported by the information contained in the text and the information that may be plausibly extrapolated in order to demonstrate evidence of "reading beyond the lines" ability. The proficiency range of the spoken/written responses may span Levels 3 and 4; however, the quality of the spoken or written responses will not be rated. The responses will be rated as Successful/Partial and Unsuccessful. The number of Successful responses will have to be greater than those of the Partial or Unsuccessful ones for the performance to be considered a pass.

16. VALIDATION

The initial item validation work and setting of cut scores will be accomplished using the modified Angoff rating procedure and a combination of holistic and analytic rating procedures (Retrodictive Modelling Approach). The performance of the test items and the cut scores will be verified through collection of data in subsequent test administrations.

Questionnaires and interviews with the examinees from the trial test populations will be used in order to collect qualitative data on the testing protocol.

Note: These test specifications may need to be adjusted based on the findings that arise during the test development process, as well as the data collected from the trialling.



[page intentionally left blank]