



**BUREAU FOR INTERNATIONAL
LANGUAGE CO-ORDINATION**

**NATO STANAG 6001
LEVEL 4 LANGUAGE PROFICIENCY**

– A Conceptual Model and Implications for Testing –



BILC Working Group on Level 4 Proficiency

April 2013

[page intentionally left blank]

EXECUTIVE SUMMARY

1. INTRODUCTION

The purpose of the Framework document is to provide a comprehensive interpretation and a conceptual model of language competence associated with Level 4 IAW the NATO STANAG 6001, Ed. 4 proficiency scale. In addition, the paper discusses requirements for testing language competence at Level 4. This includes the testing construct, suitable techniques, methods and approaches for each language skill area (listening, speaking, reading and writing) and the type of testing organization and quality control system required to support Level 4 testing. The paper is intended to enhance the common understanding of the Level 4 descriptors of STANAG 6001. In addition, it can be used as a reference to assist language testers in developing valid, reliable, and practical language tests at Level 4.

2. BACKGROUND

The paper was developed by a BILC working group formed to respond to a concern expressed by some nations that occasionally were expected to conduct Level 4 language tests. These nations often felt under-resourced and ill-equipped to meet this requirement. Even though the paper is primarily intended for testing specialists and test developers, it is also expected to serve as a tool for analyzing job positions in NATO that require high level foreign language skills and for attaching appropriate Standardized Language Profile (SLP) designations to them.

3. GENERAL DESCRIPTION OF LEVEL 4 PROFICIENCY

Level 4 language is linguistically complex.* The STANAG 6001, Ed. 4 label for Level 4 is “Expert”, indicating that this level of language proficiency is typically, although not necessarily and not exclusively, achieved by individuals who use the foreign language extensively on a daily basis as part of their profession or specialization, usually in an environment where the target language is the primary means of communication.

Level 4 proficiency is usually more characteristic of individual ability than of job requirements. A Level 3 speaker may be able to represent an official position or participate in the negotiation of an agreement. However, a Level 4 speaker will bring to that process a greater understanding and more effective linguistic expression of cultural norms and expectations. The Level 4 user will feel sufficiently at ease when using the language to find socio-culturally appropriate ways to tailor/adjust language to the

* It should be noted that higher level language is also required to perform Level 3 tasks. Level 3 language users need to understand and use abstract linguistic concepts and formulations; their discourse needs to include hypothesis, analysis, and arguments supporting or opposing issues and positions.

situation and the person to whom one is speaking (for example, to persuade them to reconsider their official point of view). He/she can employ those skills in international professional situations related to his/her specialization as well as in social settings and more informal encounters. Language users at this level are also comfortable with the entire range of styles and registers in the target language, from very formal settings (such as lectures and speeches) to very informal (such as discussions about complex or sensitive topics with close friends). A similar range of abilities is observable in Level 4 users when performing in the other skill areas – listening, reading, and writing.

4. IMPLICATIONS FOR TESTING

The paper also reviews the implications of testing at Level 4 across the four skill areas. It addresses some of the complexities associated with test development at this level. These include selection of appropriate texts for testing the receptive skills; techniques for testing at higher levels; and the selection, training, and quality control of appropriate personnel for conducting and rating tests of the productive skills. In addition, the paper emphasizes the importance of establishing a closely-monitored and standardized testing system, which consistently produces ratings that can withstand external challenges and legal scrutiny.

5. CONCLUSION

In conclusion, it is important to point out that the BILC Working Group believes that Level 3 is the highest level of language needed for most requirements of most jobs. Even the language needed for a very high level job task, such as negotiating a treaty, may combine discourse from Level 2 (concrete, factual language) and Level 3 (abstract linguistic formulations) with Level 4 language (nuanced and very precise language, drawn from the individual's area of specialization). One implication of this situation is that a person commanding all of the Level 4 job tasks may not necessarily pass a Level 4 test, because general proficiency is not always congruent with job performance.

CONTENTS

1. Introduction	1
2. Background	1
3. The Conceptual Model	2
3.1 A Framework for Level 4 Proficiency	2
3.2 General Description of a Level 4 Language User	4
3.3 Significant Differences between STANAG 6001 Level 3 and Level 4	6
3.4 Listening at Level 4	7
3.5 Speaking at Level 4	9
3.6 Reading at Level 4	11
3.7 Writing at Level 4	13
4. Implications for Testing at Level 4	15
4.1 Testing Listening at Level 4	15
4.2 Testing Speaking at Level 4	24
4.3 Testing Reading at Level 4	27
4.4 Testing Writing at Level 4	29
4.5 Pre-testing and Test Administration	32
References	35
Working Group on Level 4 Proficiency – <i>members and contributors</i>	36
Appendices	37
Appendix A : Trisections	37
Appendix B : Sample of Level 4 reading text	39

[page intentionally left blank]

1. INTRODUCTION

This paper deals with the issue of how to interpret the STANAG 6001 descriptors of language proficiency at Level 4. Drawing on recent research on higher level proficiency attainment in foreign language learning and on the current interpretation and understanding of the descriptors for STANAG 6001 Levels 1, 2, and 3, it presents a conceptual framework of Level 4 language proficiency. In addition, implications for testing at Level 4 in terms of construct, suitable techniques, methods and approaches for each skill area are considered. The aim of the framework is to enhance the common understanding of the ramifications of the Level 4 STANAG 6001 proficiency descriptors. In particular, it aims to assist language testers in developing valid, reliable, and practical language tests at Level 4 to elicit ratable samples of proficiency.

2. BACKGROUND

At the BILC Conference in Istanbul, 2010, the *Study Group on Level 4 Testing* recommended that a language needs analysis be conducted in order to ascertain how many Peacetime Establishment (PE) positions within NATO Headquarters designated as requiring the incumbents to possess Level 4 proficiency do, indeed, require that level in order to carry out their job tasks. The reason behind this was the fact that some member nations had reported that they were required, albeit infrequently, to test at Level 4 and that they did not feel resourced and prepared enough to meet that requirement. Consequently, at the above conference, a decision was taken by the BILC Steering Committee to form a working group that would research Level 4 proficiency IAW NATO STANAG 6001 and, in case the language needs analysis would corroborate a veritable need for Level 4 testing, propose an action plan on how to assist nations to develop and administer tests at this level. The *Working Group on Level 4 Proficiency* was established in August 2010, and met several times in the course of 2010-2012.

In the initial stage of its complex tasking, the Working Group aims to present a theoretical model of Level 4 proficiency in the form of an amplification of the Level 4 descriptors as set forth in NATO STANAG 6001, Edition 4, followed by an overview of the implications for testing at Level 4 in the four skill areas. For illustration purposes, sample Level 4 texts and test items are included.

3. THE CONCEPTUAL MODEL

The starting point of any assessment of Level 4 language proficiency is the interpretation and thorough understanding of the level itself, i.e. the criterion against which proficiency is measured. To that effect, the STANAG 6001 Level 4 descriptors are analysed and the statements contained therein expanded upon. It is noteworthy to mention that the STANAG 6001 framework is conceptual, and not (English) language specific.

3.1 A Framework for Level 4 Proficiency

- *Individualized Language Use*

Level 4 language is linguistically complex, with low redundancy and multiple meanings embedded in the socio-cultural context. At this level, the spoken and written texts are highly individualized and make the greatest demands on the listener and reader. Linguistic production at Level 4 involves an innovative approach to whatever topic the speaker or writer addresses (Child, 1987).

Level 4 texts are not often encountered in the colloquial language of exchange and are also infrequent even in the written language for professional purposes. By definition these texts are challenging either because they reflect unfamiliar cultural concepts or highly sophisticated language behaviour or a combination of the two. Comprehending the full ramifications of language at this level entails interpretation of the layers of explicit and implicit meaning of the spoken or written text.

- *Meaning and Comprehension*

The use of productive and receptive language skills implies the construction of meaning and the act of comprehension. Language use is a meaning making process (Holtgraves & Kashima, 2007, p. 1). Successful extraction of meaning from a spoken or written text results in comprehension.

Meaning can be understood as that which is intended by the speaker or writer. Meaning is conveyed at all levels; however, the higher the proficiency level, the more linguistic means the speaker and the writer have at their disposal to convey it, and the more means the listener and the reader have to comprehend it. The kind of linguistic resources a speaker/writer brings to the task is related to the purpose of the text (see *Overview of Text Characteristics* further on). If the purpose is to inform, the speaker/writer uses exactly that language needed to provide information as efficiently as possible. But if the purpose is to project a personal approach to a message, then the speaker/writer feels free to use any and every linguistic resource available to him/her to convey the message and influence the listener or reader. Meaning is not only contained in the spoken or written text itself but is also constructed by the receiver (the listener or reader). It is dynamic, variable and different for the same receiver at different times and in different contexts. Meaning is personal to the extent that the similarities between the encoded and decoded meanings vary according to the receiver's purpose and motivation, attitudes, biases, and interests. It

also varies according to the receiver's own framework of world and background knowledge and experience. Meaning is social and cultural, because language and conventions work only as shared meaning from cultural codes which have been handed down within culture (Lye, 1996). Level 4 language use is often characterized by a huge gap between the relatively simple syntax and sentence level discourse and the complex semantic meaning embedded within a cultural context.

Comprehension is the act of simultaneously extracting and constructing meaning through interaction and involvement with spoken and written language (*cf.* Snow, 2002, p. 11). At higher language proficiency levels, comprehension involves not only the ability to distinguish main ideas from supporting details (Level 2), but also to summarize and paraphrase accurately, to fill in gaps and infer on the basis of contextual clues, to draw conclusions, raise questions, see the underlying causal patterns, note bias and distortion, evaluate the speaker's or writer's effectiveness, synthesize personal experience or background knowledge with the text, and understand allusion and references to the socio-cultural or historical context.

Comprehension can roughly be categorized into three interrelated main components: literal, inferential, and critical/evaluative (Clymer, 1968).¹ Each component is interrelated to and dependent upon the other components. The literal comprehension component (understanding explicitly stated information) is fundamental for lower proficiency levels and will therefore not be addressed in the present model. Full comprehension at Level 4 goes beyond literal understanding of explicitly stated information; it involves the interpretation and critical evaluation of the text and the speaker or writer, who may have a unique attitude toward the subject of their discussion. In other words, the comprehension tasks that characterize the upper proficiency levels are inferential and evaluative by nature.

Inferential comprehension or understanding meaning "between the lines" is 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. It includes making a case for one's interpretation and/or drawing a conclusion regarding a text based on clues and evidence, revising conclusions and interpretations as new clues or evidence are found, and generalizing interpretations to a larger theme or area. Complete comprehension of a Level 4 text requires the receiver also to interpret correctly the speaker's or writer's use of connotation and tone. Connotation carries the meaning of a word beyond its literal meaning and gives it an added dimension. Connotation reflects the speaker's, or writer's stance and viewpoint and determines the emotional effect of a statement. Tone refers to feelings, mood and attitude that the speaker or writer wants to convey. In concrete, factual texts the tone is neutral or objective, whereas the tone becomes subjective when the author's purpose is to persuade, shock, or arouse the receiver's emotions. Detecting the author's tone is often taxing for non-natives because the devices used to express the

¹ A creative dimension could also be added since language use can be seen as a creative act.

tone, such as metaphors, similes, humour, irony, satire and sarcasm, are usually embedded in the culture.

Evaluative comprehension or understanding meaning “beyond the lines” is 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. Evaluative comprehension involves making a judgement about the text genre and mode of discourse, rhetorical organization of the text, and the speaker’s or 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 speaker’s or 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.

- *Higher Order Thinking Skills*

Full comprehension of higher level language becomes more complex as density of information increases, signalling and redundancy decrease, ideas and opinions are more implicit and content, concepts and language are more abstract. Consequently, the present conceptual model acknowledges 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.

3.2 General Description of a Level 4 Language User

The STANAG 6001, Ed. 4 label for Level 4 is “Expert”, indicating that this level of language proficiency is typically, although not necessarily and not exclusively, achieved by individuals who use the foreign language extensively on a daily basis as part of their profession or specialization, usually in an environment where the target language is the primary means of communication. As such, these individuals are often considered authorities or experts in the foreign language by other non-natives.

Level 4 language proficiency is usually called for in very taxing academic and professional settings, and is typically only required of people who need to have an outstanding ability with the foreign language in order to deal with demanding subjects at the highest levels. However, this level of language can also be used in social settings where interlocutors are comfortable with this type of discourse. Such individuals can use language to lead in challenging and complex professional settings, and negotiate and persuade effectively in international professional environments. Examples of military-related tasks at this level include serving as the spokesperson responsible for press releases and press conferences requiring nuanced, culturally appropriate communications necessary to win support for national policies or actions; acting as an arbiter between warring factions during a sensitive peace-keeping assignment; providing official

interpretation services² at international meetings; analyzing the real communicative intent of diplomatic pronouncements.

At Level 4, the use of the foreign language becomes highly precise, nuanced and effective. One important characteristic of individuals with this level of proficiency is that they readily adapt and tailor their language to suit the purpose and situation in order to achieve intended effect. A Level 4 language user shows a firm grasp of various levels of style and register, and can understand and express subtleties and shades of meaning. Communication at this level is highly successful in a wide variety of interactions, ranging from highly formal to very casual. Level 4 language proficiency presupposes that a high degree of competence in the handling of abstract ideas and concepts has been achieved by the language user. Individuals at this level also possess the discourse skills that would be expected of an educated native speaker in the target culture, such as the ability to hypothesize and persuade.

Level 4 users can typically be expected to understand documents, correspondence and reports with multiple layers of information, and comprehend the nuances and finer points of complex texts. They can advise on or handle complex, delicate or sensitive issues without awkwardness, understand and use colloquial expressions and culturally embedded references, and deal smoothly with challenging linguistic settings. In their professional field or other areas of interest, these individuals can write for public purposes with precision and accuracy.

At this level, the individual has the linguistic competence to use the foreign language in a range of culturally appropriate ways, and with a fluency and sophistication approaching that of an educated native speaker. Users at this level are able to enrich their use of the language by expanding their vocabulary and refining their usage and command of style and register. At this level, individuals can understand any type of press or other media, and other areas of culture.

Since Level 4 language users have an effective and multi-dimensional understanding of a variety of topics, and are able to understand or express meaning “beyond the lines”, they can readily comprehend and participate in discussions relating to the full range of the target culture’s history, society, customs, traditions, current events, and national politics. In other words, these individuals have spent enough time in the target language environment that they have become acculturated to such a high degree that they might be perceived as full members of that culture.³

² However, one should keep in mind that interpretation is a separate skill requiring skill specific training.

³ It should be noted that staying for an extended period of time in a foreign culture alone is not a guarantee for attaining Level 4 proficiency. Studies (e.g., Flege and Liu, 2001) have demonstrated that other factors such as the proportion of contact with native speakers and the motivation of language users to continue improving their language skills are at least equally critical in attaining this level of proficiency.

3.3 Significant Differences between STANAG 6001 Level 3 and Level 4

FOREIGN LANGUAGE PROFICIENCY

LEVEL 3	LEVEL 4
<p>Understands/Can produce formal and informal language, for most everyday social and work-related situations.</p>	<p>Understands/Can produce highly sophisticated language appropriate for almost all topics and social as well as professional situations.</p>
<p>Understands/Can produce language relating to abstract concepts and hypotheses.</p>	<p>Readily adapts language to any situation. Shows a firm grasp of various levels of style and can understand/express subtle nuances and shades of meaning.</p>
<p>Understands/Can produce extended discourse and conveys meaning correctly and effectively. Understands arguments/Can argue for and against different opinions.</p>	<p>Uses the language with great precision for all social and professional purposes, including persuasion, negotiation, or the representation of an official point of view.</p>
<p>Can understand/express meaning 'between the lines'.</p>	<p>Can understand/express meaning 'beyond the lines'.</p>
<p>Understands/demonstrates humour and irony, but may not fully understand some allusions, as well as implications of nuance and idioms.</p>	<p>Good command of idiomatic expressions, figures of speech and colloquialisms.</p>
<p>Rarely requests repetition. Has a natural flow, without searching for words.</p>	<p>Consistently maintains a high degree of grammatical and lexical accuracy. Uses the language effortlessly; only conceptually difficult subjects can hinder a natural, smooth flow of language.</p>
<p>Language use is socio-linguistically not always entirely appropriate for the situation.</p>	<p>Language use is socio-linguistically almost always appropriate for the situation.</p>
<p>Occasional errors rarely disturb the native speaker.</p>	<p>Errors are rare.</p>

3.4 Listening at Level 4

- *Level 4 descriptor for Listening (taken from STANAG 6001, Ed. 4)*

Level 4 – Expert
<p>Understands all forms and styles of speech used for professional purposes, including language used in representation of official policies or points of view, in lectures, and in negotiations. Understands highly sophisticated language including most matters of interest to well-educated native speakers even on unfamiliar general or professional-specialist topics. Understands language specifically tailored for various types of audiences, including that intended for persuasion, representation, and counselling. Can easily adjust to shifts of subject matter and tone. Can readily follow unpredictable turns of thought in both formal and informal speech on any subject matter directed to the general listener. Understands utterances from a wide spectrum of complex language and readily recognizes nuances of meaning and stylistic levels as well as irony and humour. Demonstrates understanding of highly abstract concepts in discussions of complex topics (which may include economics, culture, science, and technology) as well as his/her professional field. Readily understands utterances made in the media and in conversations among native speakers both globally and in detail; generally comprehends regionalisms and dialects.</p>

- *Amplification of the Descriptor*

A Level 4 listener understands highly sophisticated spoken language of well-educated native speakers on unfamiliar general or professional-specialist topics. S/he generally understands specialized language outside of his/her area of expertise. This listener can follow with relative ease most lectures, briefings, discussions, debates, and presentations, including those employing a degree of colloquialism, language variety or unfamiliar terminology. S/he can follow extended speech on general abstract and complex topics beyond his/her own field; s/he can discern relationships among sophisticated listening materials in the context of broad experience. A listener at this level can follow unpredictable turns of thought readily, for example, in informal and formal speeches covering editorial, conjectural, and literary materials in any subject matter directed to the general listener.

A Level 4 listener recognizes, understands and almost always correctly interprets cultural allusions, nuance and tone in the behaviour and language of members of the target culture. This listener can appreciate the subtle differences that exist in the underlying fundamental value systems and their sub-components in a given society, and may use this ability to further his/her skills through appropriate interpretation of the speaker's manner, tone, use of rhetorical devices, modes of expression, and possibly of body language. A Level 4 listener is aware of the full range of formal and informal styles within specific language registers and comprehends most of the common regional varieties of the target language provided that the topic is not culturally bound to the region. S/he understands discrete use of idiomatic expressions, colloquialisms, humour, puns

and references to regional cultures. This individual generally comprehends less common figures of speech such as understatements, malapropisms, and spoonerisms.

Excited or highly emotional language rarely causes difficulty for a Level 4 listener, except where the pronunciation is grossly affected. S/he can record information about an incident from several points of view and discriminate between slight variations in the reported observations. With only occasional lapses of concentration s/he fully comprehends radio or television news broadcasts, plays, films and documentaries, even when a considerable degree of slang and idiomatic usage is employed. This listener can identify finer points of detail including implicit attitude, speaker viewpoints and relationships between speakers. Comprehension is limited only in parts which include oblique cultural or historical references. S/he can extract specific information from poor quality, audibly distorted public announcements, e.g., in a station or a sports stadium.

3.5 Speaking at Level 4

- *Level 4 descriptor for Speaking (taken from STANAG 6001, Ed. 4)*

Level 4 – Expert
<p>Uses the language with great precision, accuracy, and fluency for all professional purposes including the representation of an official policy or point of view. Can perform highly sophisticated language tasks, involving most matters of interest to well-educated native speakers, even in unfamiliar general or professional-specialist situations. Can readily tailor his/her use of the language to communicate effectively with all types of audiences. Demonstrates the language skills needed to counsel or persuade others. Can set the tone of both professional and non-professional verbal exchanges with a wide variety of native speakers. Can easily shift subject matter and tone and adjust to such shifts initiated by other speakers. Communicates very effectively with native speakers in situations such as conferences, negotiations, lectures, presentations, briefings, and debates on matters of disagreement. Can elaborate on abstract concepts and advocate a position at length in these circumstances. Topics may come from such areas as economics, culture, science, and technology, as well as from his/her professional field. Organizes discourse well, conveys meaning effectively, and uses stylistically appropriate discourse features. Can express nuances and make culturally appropriate references. Speaks effortlessly and smoothly, with a firm grasp of various levels of style, but would seldom be perceived as a native speaker. Nevertheless, any shortcomings, such as non-native pronunciation, do not interfere with intelligibility.</p>

- *Amplification of the Descriptor*

A Level 4 speaker produces clear, smoothly flowing, well-structured extended discourse, showing controlled use of organizational patterns, connectors and cohesive devices. At this level, the individual has the linguistic competence to speak the foreign language in a range of culturally appropriate ways, and with a fluency and sophistication approaching that of a well-educated native speaker. S/he can perform extensive, sophisticated language tasks, encompassing most matters of interest to well-educated native speakers, including tasks which do not bear directly on a professional specialty. A Level 4 speaker communicates effectively and effortlessly with various audiences on a wide range of topics. S/he can give elaborate and precise descriptions and narratives, integrating sub-themes, developing particular points and rounding off with an appropriate conclusion. S/he can handle all social and professional content areas and discuss all topics normally pertinent to professional needs. The context may vary widely, from highly formal to very casual situations.

Level 4 speakers are able to tailor language to fit an audience. They can also express nuances and subtleties, such as reticence or doubt. Level 4 speakers know how to address people in a wide variety of conflicts and at various social levels appropriately. They can advise on or discuss complex, delicate or sensitive issues without awkwardness, use appropriate culturally embedded references and deal confidently with argumentative or

unsympathetic interlocutors. They can represent and defend a point of view other than their own. They can convey a specific tone or attitude and consequently can counsel and persuade as well as negotiate. They can establish credibility and participate successfully in all aspects of public discourse, such as meetings, conferences, debates, presentations, and public interviews, including those conducted on public media and press conferences. They can discuss in detail concepts which are fundamentally different from those of the target culture and make those concepts clear and accessible to the native speaker. As these speakers have an effective and multi-dimensional understanding of a variety of topics, they can discuss in a comprehensive way the target culture's history, society, customs, traditions, current events, and national politics. They control and use the full range of formal and informal styles of expression and are able to serve as an effective mediator, advisor, or an informal interpreter.

Although this speaker would rarely be mistaken for someone born and brought up within the target culture due to his/her non-native accent or non-native linguistic slips, s/he makes effectively use of cultural references, proverbs, sayings, and quotes from literature and other culturally significant sources. S/he is able to synthesize knowledge and information from various sources and disciplines, and may be able to present them in a new and idiosyncratic way by adding his/her interpretation on these concepts. At the other end of the spectrum, this individual uses colloquialisms, slang and a wide range of idiomatic expressions required in more informal settings.

A Level 4 speaker consistently maintains a high degree of grammatical and lexical accuracy, with only occasional unpatterned errors. When errors are made the speaker is capable of correcting them. Some of the errors may be faulty intonation or pronunciation indicating that this is not a native speaker or using language, for example idiomatic expressions, metaphors and proverbs in a non-native way; however, errors of this type do not detract from communication. S/he demonstrates a good command of a broad lexical repertoire, incorporating regional variations common to the language as a whole, and allowing gaps to be readily overcome with circumlocutions. The speaker is to a certain extent also able to use specialist language from outside his/her own field. There is little obvious searching for expressions or avoidance strategies, while only a conceptually difficult subject can hinder a natural, smooth flow of language.

3.6 Reading at Level 4

- *Level 4 descriptor for Reading (taken from STANAG 6001, Ed. 4)*

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>

- *Amplification of the Descriptor*

A Level 4 reader can readily understand all forms of written language, including lengthy, highly abstract, structurally or linguistically complex texts dealing with unfamiliar general and professional topics. S/he 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 belief systems and ideology, 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, such as the use of metaphors. Confusion may arise over archaic structures and/or idiomatic uses of grammar.

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 achieve full understanding of 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.

Vocabulary extends across a broad range of topics and the reader requires use of a dictionary only for unique terminology outside his/her own field. This individual recognizes all professionally relevant vocabulary known to the educated nonprofessional native, although s/he may have some difficulty with slang. Phrases and idioms commonly quoted from regional varieties of the target language cause little difficulty. Accuracy is often nearly that of a well-educated native reader.

3.7 Writing at Level 4

- *Level 4 descriptor for Writing (taken from STANAG 6001, Ed. 4)*

Level 4 – Expert
<p>Can write the language precisely and accurately for all professional purposes including the representation of an official policy or point of view. Can prepare highly effective written communication in a variety of prose styles, even in unfamiliar general or professional-specialist areas. Demonstrates strong competence in formulating private letters, job-related texts, reports, position papers, and the final draft of a variety of other papers. Shows the ability to use the written language to persuade others and to elaborate on abstract concepts. Topics may come from such areas as economics, culture, science, and technology as well as from the writer's own professional field. Organizes extended texts well, conveys meaning effectively, and uses stylistically appropriate prose. Shows a firm grasp of various levels of style and can express nuances and shades of meaning.</p>

- *Amplification of the Descriptor*

A Level 4 writer produces clear, well-structured extended discourse, showing controlled use of organizational patterns, connectors and cohesive devices. This writer employs a variety of organizational patterns, uses a wide range of cohesive devices such as ellipsis and parallelisms, and subordinates in a variety of ways. S/he can write effectively and with confidence on complex topics, underlining the relevant salient issues, expanding and supporting points of view at some length with subsidiary points, arguments and relevant examples, and rounding off with an appropriate conclusion. S/he can handle all social and professional content areas, and all topics normally pertinent to professional needs. S/he can write letters on any subject and full notes of meetings or seminars with good expression and accuracy.

Level 4 writers are consistently able to tailor language to fit the addressees by using appropriate registers as well as appropriate styles of writing, such as narrative, descriptive, argumentative, and expository prose. They can also express nuances and subtleties, such as reticence or doubt. Level 4 writers are capable of appropriately addressing readers in a wide variety of conflicts and at various social levels. They can advise on complex, delicate or sensitive issues without awkwardness, and use appropriate culturally embedded references. They can represent and defend a point of view other than their own. They can convey a specific tone and present arguments in a logical and convincing way, and consequently can counsel, persuade and insinuate. As these writers have an effective and multi-dimensional understanding of a variety of topics, they can discuss the full range of the target culture's history, society, customs, traditions, current events, and national politics.

At this level of proficiency, the writing reflects the socio-cultural and linguistic norms of the target culture, including the appropriate use of the full range of formal and informal

styles of expression. Writings produced by such individuals may appear completely “non-accented” to the native reader. A Level 4 writer uses precise, highly abstract, nuanced and sophisticated lexicon, as well as figurative language such as metaphor and imagery. The language often includes complex grammatical and syntactical structures. It is characterized by effective transitions from idea to idea and discourse markers, highly coherent discourse, and adroit rhetorical devices. Because an individual with this level of proficiency has had a significantly long exposure to the target culture, s/he makes effectively use of cultural references, proverbs, sayings, and quotes from literature and other culturally significant documents and is able to synthesize knowledge and information from various sources and disciplines, and may be able to present them in a new and idiosyncratic way by adding his/her interpretation on these concepts. At the other end of the spectrum, this individual uses colloquialisms, slang and a wide range of idiomatic expressions required in more informal settings. In addition, language proficiency at this level is so developed that this individual can use word play, humour and irony to achieve the desired tone, effect and intent.

A Level 4 writer consistently maintains a high degree of grammatical and lexical accuracy, with only occasional unpatterned errors in low-frequency, complex structures. Vocabulary is extensive. Use of colloquialisms and literary devices such as metaphor, proverbs, allegory, may be occasionally inappropriate. Synonyms, antonyms and idioms are used with precision. The writer is able to exploit specialist language from outside his/her own field. In formal writings little or no editing by a native writer is required.

4. IMPLICATIONS FOR TESTING AT LEVEL 4

Within the BILC framework, it has been widely acknowledged that training students to Level 3, as well as testing them at this level already poses significant challenges, for even this proficiency level is not easily attainable through the medium of classroom instruction alone. Persons displaying Level 3 proficiency have usually acquired it by frequent or extended exposure to the target language country and culture.

Due to the fact that in NATO relatively few positions require foreign language proficiency at Level 4, not many institutions possess the “know-how” to develop and administer tests at this level. Moreover, a survey of the existing literature has shown that studies analysing assessment of this level of proficiency are exceedingly rare.

Nonetheless, institutions that endeavour to develop tests at Level 4 should follow the standard procedures for test development comprising a thorough needs analysis, design of test specifications, test development, trialling, analysis, administration, evaluation and monitoring. Selecting and training test developers, testers and raters play an even more important role at this level than it does at lower levels; this will be discussed at the section on pre-testing and test administration further on.

4.1 Testing Listening at Level 4

- *Selecting Appropriate Texts*

One of the challenges in defining the levels of texts and tasks for testing listening comprehension, particularly at the highest levels, stems from the “unstable” nature of the texts and the “fluid” nature of the listening construct. Defining, categorizing, and identifying the difficulty level of authentic listening texts and tasks is often complex and challenging. The gradual progression in content and structure that can be observed in most authentic reading texts – usually moving from personal, most frequent, everyday content expressed in linguistically simple language, to more societal, complex, infrequent, unpredictable content communicated in linguistically challenging structures – may or may not appear in listening texts in the same hierarchy, sequence, or expansion. In other words, a topically simple interpersonal conversation can require a very low or a very high proficiency level to decipher the meaning, depending on the surrounding conditions.

There seem to be several reasons for the challenge of defining the level of texts and tasks at the high end of the listening scale. Firstly, the listening skill domains that Level 4 listeners deal with can be encountered in both interpersonal (informal) settings as well as wider and societal (formal) settings with larger audiences. And because of the transitory nature of spoken versus written language, during a listening act one requires heavier reliance on the context in which the linguistic input is produced. Secondly, the monologue or dialogue that is heard can be a communication between just two people or groups that have already pre-established the references, referents, and content/context domains of those factors. This may result, even in cases where simple forms and content are used, in a linguistic task of a higher level, in which the listener must fill in the gaps using both socio-

cultural and socio-linguistic schemata, as well as higher-tier linguistic strategies, in addition to a very high level of accuracy in processing and reconstructing the meaning.

There are other factors that may play a significant role and elevate the level of difficulty in listening tasks. One is the possible presence of situational and/or non-linguistic factors such as background noise, signal quality, interference; the other factor is the features and properties of language that may appear in interactive listening texts, such as false starts, digressions, accent, overtalk, turn-taking issues, and other unique properties of spoken language that may place an additional burden on the listener's skill level.

In addition to the schematic load, one of the challenges Level 4 listeners generally face is the prosodic elements that are prevalent in higher-level listening texts, in which the tone, attitude (for example, hostility, subdued anger, defensiveness) and even the semantic message is conveyed simply through skilful and purposeful use of pitch, stress, and intonation. These prosodic elements require additional work and reflection by both native and non-native listeners at this level.

Another important aspect to keep in mind is the non-visual aspect of the listening process in non-participatory situations (radio shows, for example in real life, or audio input only under testing conditions). The lack of visual support, in addition to the caveats mentioned above, can directly impact the accuracy of a listener because the listener needs to engage auditory skills in the semantic reconstruction process. Extensive research is being conducted on the issue of input length and its impact on the short-term memory, and whether the length factor aids or hinders listening comprehension with regard to the memory taxation. The conditions and caveats that surround each input, as well as the impact of visual input on the listening comprehension process, need to be further investigated since much listening related input these days comes from TV or the Internet – media that deliver a mixture of both visual and auditory input, which needs to be factored in the descriptors that deal with the listening skill.

Nonetheless, there are certain types of listening texts that could be defined and categorized in terms of the “text modes” for reading passages (see section 4.3), where the listeners' purpose as well as the typological features of the texts seem to progress in the same direction and hierarchy as reading passages. These listening texts are usually found in the media where the focus of the text is directed to a wider audience. In such cases, the approach to reading typology may apply to listening as well.⁴

The Overview below (*Table 1*) has been adapted from Clifford's Overview of Text Modes (Child *et al.*, 1993). The second table shows characteristics and examples of listening text types in participatory and non-participatory situations (Level 4 only), and also captures the listener's purpose.

⁴ An example of a Level 4 listening text can be viewed at <http://www.charlierose.com/view/interview/12276> .

Table 1: Overview of listening text modes.

OVERVIEW OF TEXT MODES BY LEVEL

LISTENING

(At each level, these elements must be aligned)

Level	Speaker Purpose	Typical Text Type	Listener Purpose
1	Orient by communicating main ideas.	Simple short sentences with simple vocabulary. Text organization is somewhat loose, but follows cultural norms.	<i>Orient oneself</i> by identifying topics and main ideas.
2	Instruct by communicating factual information.	Connected factual discourse with compound and complex sentences dealing with concrete information. Discourse is paragraph-length, with sentences carefully sequenced within paragraphs. The speaker's personal viewpoint is subordinated to a factual presentation.	<i>Acquire information</i> by understanding not only the main topics and facts, but also supporting details such as temporal and causative relationships.
3	Evaluate situations, concepts, conflicting ideas; present and support arguments and/or hypotheses with both factual and abstract reasoning.	Extended discourse on a variety of unfamiliar or abstract subjects; using either formal or informal speech such as might be found in professional discussions, supported opinion, hypothesis, argumentation and elaboration. References may be made to previous statements, to common cultural values, etc. The speaker's unique personal point of view is evident.	<i>Learn</i> by relating ideas and conceptual arguments, by understanding the messages "in the lines" and "between the lines," by recognizing the speaker's tone and intent.
4	Project lines of thought beyond the expected; connect previously unrelated ideas or concepts; present complex ideas with nuanced precision and virtuosity.	Extended discourse that is tailored for the message and the intended audience. To achieve the desired tone and precision of thought, the speaker will often skillfully use low-frequency vocabulary, cultural and historical concepts, and demonstrate understanding of the audience's shared experience and values.	<i>Listen "beyond the lines,"</i> understand the speaker's sociolinguistic and cultural references, follow innovative turns of thought, and interpret the message in view of its wider cultural, societal, and political setting.

Table 2: Overview of listening text characteristics.

LEVEL 4 LISTENING TEXT CHARACTERISTICS

Typical Text Type / Genre	Speaker Purpose	Listener Purpose/Task IAW STANAG 6001
<ul style="list-style-type: none"> ○ All types of discourse on all types of subjects, presented in a manner unique to the speaker. ○ Planned varieties of such discourse reflect the description from <i>Overview of Listening Text Modes</i>. When more spontaneous, texts may contain natural language features (such as false starts, interruptions, overlapping speech, etc.). 	<p><i>Project</i> lines of thought beyond the expected; connect previously unrelated ideas or concepts, and present complex ideas with nuanced precision and virtuosity in a variety of registers. May include persuasion and tailoring of language.</p>	<p><i>Listen “beyond the lines,”</i> understand the speaker’s sociolinguistic and cultural references; follow innovative turns of thought; interpret a message in view of its wider, cultural, societal, or political setting.</p>
Non-Participatory Situations*		
<ul style="list-style-type: none"> ○ Lectures, broadcast interviews and discussions, briefings, conference presentations. <p>For example: an interview contrasting critical thinking with emotional intelligence.</p>	<p><i>Analyze</i> issues, present complex or novel ideas or arguments on professional, political, economic, cultural or societal issues (through public media, overheard conversations).</p>	<p><i>Understand</i> unpredictable shifts of subject matter and tone; understands highly sophisticated language and references embedded in a wider socio-cultural context.</p>
Participatory Situations*		
<ul style="list-style-type: none"> ○ Negotiations ○ Debates ○ Professional meetings. <p>For example: influencing others to change a policy position</p>	<p><i>Elaborate</i> on various highly abstract topics</p> <p><i>Counsel/persuade</i> or influence using nuanced language specifically tailored to the listener(s).</p>	<p><i>Understand</i> highly complex discourse and all styles of speech (highly formal and abstract or highly informal/colloquial).</p> <p><i>Respond</i> and contribute satisfactorily, although not necessarily in the projective mode.</p>

* Typical but not comprehensive examples in participatory and non-participatory situations

- *Selecting Appropriate Tasks*

The step preceding the decision on the testing methods or techniques would be to develop detailed test specifications, and list the tasks to be tested. At Level 4, the tasks include the ability to:

- Adjust to shifts of subject matter and tone
- Follow unpredictable turns of thought in both formal and informal speech on any subject matter addressed to the general listener
- Recognize nuances of meaning and stylistic levels, irony, and humour
- Understand language specifically tailored for various types of audiences, including that intended for persuasion, representation, and counselling
- Demonstrate understanding of highly abstract concepts in discussions of complex topics (which may include economics, culture, science, and technology)

Topical domains for test items at this level would be broad, encompassing subject matter requiring the use of highly abstract, nuanced and sophisticated language. Domains span an almost limitless range, including unfamiliar and professional-specialist topics. Topics themselves do not determine whether a text is a Level 4 listening text; it is the manner in which a topic is treated that will influence its level of difficulty. Level 4 texts are characterized by low-frequency lexicon and grammatical structures and, generally, contain numerous lexical, propositional and pragmatic inferences. Texts at this level may also be rich in colloquialisms, street talk, idioms, and figurative usage. During the text selection process, texts should be evaluated in terms of their level of difficulty, and in addition to consulting the STANAG 6001 Level 4 descriptor, which represents the construct for test development, and the Content/Task/Accuracy statements, texts should also be compared with the above Overview of Text Modes and Overview of Listening Text Characteristics. Proficiency tasks to be developed should reflect the real world use of language.

- *Selecting Appropriate Testing Methods*

There are many different testing techniques, such as gap-filling, matching, multiple-choice, constructed response, cloze, and ordering tasks. When selecting the most appropriate testing techniques, one needs to keep in mind that higher level linguistic skills are linked to higher level cognitive skills, which include the ability to analyse, synthesize and evaluate. For that reason, gap-filling, vocabulary matching and other forms of discrete-point testing, must be considered inappropriate as Level 4 testing techniques.

As possible testing methods or techniques the following may be considered:

- *Multiple-choice format*

The multiple-choice (MC) format is a test method in which the examinee selects a response from two or more choices. As one of the leading principles of STANAG testing is that the level of the task should reflect the level of the text, the following should be taken into account if the MC technique is used. The options must be aligned with the text in terms of lexical abstractness and linguistic complexity. However, their length should be moderate enough in order to avoid that the task itself becomes more of a reading than a listening type challenge. Given that a Level 4 text typically contains multiple layers of meaning embedded in the social-cultural context, one potential problem with the MC technique at this level is that it might be difficult to formulate four distinctly different options that are not open to subjective interpretations of the test takers. Constructing plausible and clear options that at the same time reflect the complexity of the text itself without being obscure or deceptive is extremely challenging. Piloting these items on the right target population, including the well educated native speakers of the language tested, and analysing the gathered data will be of essence.

- *Interactive test format*

An interactive test is a test in which the examinee produces spoken or written (constructed) responses to a test prompt/question. If an interactive testing technique is selected, it is critical that the scoring key be comprehensive and decisions be taken in advance as to what would constitute an acceptable spoken or written response. Decisions on whether the spoken and written response would still be acceptable even though the level of the response may be lower than Level 4 must also be taken into consideration. Because listening comprehension is tested and not the speaking or writing skill, responses below Level 4 that still demonstrate unequivocally presence of comprehension might still be acceptable. Nonetheless, one should consider such decisions very carefully, as it is also crucial for testers to be able to discriminate between Level 3 and Level 4 responses. Allowing a Level 3 written or oral response to represent proof of Level 4 listening comprehension may pose a threat to test reliability. In any event, spoken responses should be recorded so that they can be rated by a second/third examiner. Moreover, tester training will play a critical role in order to ensure a standardized and consistent application of rating criteria and scoring rubrics.

One issue associated with testing techniques is whether to have monolingual or bilingual tests, i.e. whether the questions and answers should be in the target language or in the candidates' native language. Since Level 4 proficiency implies that a person at this level is both linguistically and culturally very advanced, it might be extremely disruptive to have a bilingual test because of code switching. At this level, because thought processes also occur in the target language, monolingual tests should be considered preferable.

- *Developing Scoring Rubrics*

Detailed scoring rubrics will have to be developed if the constructed written or spoken response format is chosen, especially because this involves the testing of mixed skill areas. Detailed criteria for evaluation reflecting the accuracy requirements IAW STANAG 6001, Ed. 4 need to be included, as well as the procedures for scoring. The scoring criteria must clearly indicate what is expected of language input with regard to its organization, structural correctness, lexical precision, socio-linguistic appropriateness, etc. As in every testing situation, clear and detailed instructions have to be provided to test takers in terms of what their task is and what is expected in terms of their language production and input.

- *Sample Level 4 Listening Item*

From a broadcast interview

[Transcript]

Man: Dr. Albertson, you have expressed impatience with arguments whether the Democratic Party or the Republicans first thought of some position on moral and social issues.

Woman: I genuinely believe the public is bored with this debate. Concerned people are asking whether their representatives will ever, conceivably, get around to discussing the positions themselves. I am interested in a collection of these shared or swiped or imitated positions, the ones concerning the derailed American young. I include in their number the growing support for uniforms in the schools, for curfews for minor children, and for welfare regulations that make staying in school and living at home with parent or guardian a condition of getting a grant for unwed teenage mothers.

Man: Are these issues subject to charges of larceny between Democrats and Republicans in the presidential campaign?

Woman: These positions have more in common than that. Often, for instance, the curfew proposition will be accompanied by sensible additional measures meant to discourage adolescent crime.

Man: And, where states and localities have already tried curfews, haven't some shown encouraging results?

Woman: Umm. But it strikes me that there is something else these proposals have in common. Even allowing for their better features and the preferable versions already in effect. All rest on the assumption that what we think of as the good old days are still here or at least can be made to seem to be.

Man: Isn't that an unfair charge? Aren't many politicians and private citizens genuinely alarmed about juvenile crime?

Woman: There is, in the first place, a kind of cosmetic component to all the proposals. *They* will make things look different, from the uniformed kids to the relatively safe and quiet nighttime streets to the households where the welfare checks go. But inside the uniform or off the street corners or wherever else *they* may be, the same kids, with many of the same disorders, will still be somewhere. And if it is unfair to say that *this* represents a mere façade of political and social progress, it is not unfair, I think, to say that *it* does represent at least a measure of confusion of nostalgia with reform.

Man: Shouldn't this country be able to retrieve what was best in an earlier incarnation, just as we have discarded much that was wrong and hypocritical in the past?

Woman: Certainly. But that will mean addressing life as it is, not as it used to be or as it didn't used to be but is romantically reconstructed. Politicians have an obligation to address the real problems in real ways and let the look of things take care of *itself*. I think we should keep that distinction in mind and try to hold both parties to it.

- *Sample selected response item:*

Dr. Albertson expresses the view that:

- A** such mandatory measures as curfews and school uniforms were appropriate in the past, but not today.
 - B** regimenting young people may create apparent public order but lead to violation of civil liberties.
 - C** facile appeals to past values have obscured the political debate on programs involving alienated youth.
 - D** politicians try to disguise their own misconduct by focussing on social problems associated with young people.
- *Sample constructed response item (eliciting spoken or written response in case of interactive testing technique):*
 - What political proposals are discussed?
 - What is Dr. Albertson's objection to these proposals?

Task to be performed:

Follow unpredictable turns of thought. Democrats and Republicans have argued about which party first thought of certain political proposals. Dr. Albertson is more concerned with the content of the proposals – particularly those related to young people.

The interviewer points out that some of these ideas have worked well. However, Dr. Albertson counters that a common feature of the proposals is a desire to recapture the past. The interviewer objects to her charge that politicians and others are insincere. She replies that all the proposed measures will create a façade without solving the problems. She insists that nostalgia is being confused with reform. The interviewer

asks whether the country cannot indeed recapture some of the values of the past. Dr. Albertson agrees but states that would require a realistic examination of the problems, not a focus on appearances. She concludes that both political parties should be held accountable.

If the constructed response format is chosen one should consider, in order to avoid creating an overload on memory, allowing test takers to take notes and to listen to the passage twice.

4.2 Testing Speaking at Level 4

- *Selecting Appropriate Tasks*

In order to elicit a ratable sample of Level 4 proficiency, the selection of appropriate tasks is of crucial importance. According to the STANAG 6001 descriptor, the tasks a Level 4 speaker is able to perform include:

- Tailor speech to a specific audience
- Counsel
- Persuade
- Advocate a position at length
- Elaborate on abstract concepts
- Communicate with native speakers at conferences, negotiations, lectures, briefings, debates
- Shift and adjust to shifts of subject matter and tone

When conducting speaking tests at any level, it is important that the level of the prompt or task formulated by the tester and presented to the test taker be at the level tested, thus signalling the level of language that is expected from the test taker. There is ample evidence that prompts phrased at a level below the level being tested will usually not elicit language at the required level from the test taker. In other words, if a test taker fails to meet the minimum requirements for a given proficiency level, the reason may be the testers' inaptness of eliciting language at the required level, rather than the test takers linguistic shortcomings.

To elicit a Level 4 sample, testers need to weave into their speech unfamiliar topics requiring discussion and the use of low frequency and precise abstract vocabulary and structures. The testing technique should incorporate methods of verifying the test taker's ability to persuade, tailor his/her speech to a specific audience, and the ability to shift registers (from formal to informal and vice versa). A Level 4 speaker can usually perform all the tasks a speaker at Level 5 (equivalent to a Highly-Articulate Native Speaker) can, but displays non-native intonation, accent, use of idioms and/or cultural references.

- *Selecting an Appropriate Testing Method*

An oral proficiency interview (OPI) containing a sufficient number of Level 4 tasks and covering a variety of high level topics would serve the purpose of testing speaking at this level. Just like for any other skill, as a first step test specifications should be developed outlining in detail how this test would be operationalized.

One of the hallmarks of Level 4 proficiency is the ability to tailor language, to convince, persuade, advise and negotiate. The use of proper register (formal vs. informal), expressions, references and nuances that appear in such tasks when they are carried out by native speakers in such situations are linked strongly to the relationship between the two participants in a conversation, and are also highly embedded in culture; in other

words, one and the same situation would play out differently, dependent on the socio-linguistic and cultural context in which it takes place (this varies from language to language). A good way to elicit these tasks or functions is via role-play situations.

- *Sample role-play situation for informal tailoring:*

You are living in Washington D.C. A good friend of yours is concerned because his 17-year-old daughter/son spends hours on his/her mobile device. Your friend is taking him/her to a family reunion and wants to ensure he/she spends face to face time with family members. Your friend has asked you to try to convince his daughter/son to leave his/her phone at home and try to connect more with people face to face. The tester will play the role of the daughter/son.

- *Sample role-play situations for formal tailoring:*

You have been working in an import-export firm in New York for about a year. Your employer is considerably older than you and your relationship is friendly but rather formal. You have good reason to believe that one of your employer's favourite "trusted" employees, who has worked in the firm for twenty years, has been taking merchandise home and not returning it. Alert your employer to the fact. The tester will play the role of the employer.

The language used in the role-play situations does not necessarily have to be of Level 4 difficulty as this language does not represent a text that needs to be aligned with the level of the task. Rather, it represents instructions for the test taker on how to enact the situations. Just like the instructions preceding listening and reading tests, these ones should be written as clearly as possible.

All role-play situations should be designed to be culturally appropriate and authentic for the context of the tested language. During the verification of the test taker's ability to make shifts between levels of formality, the testers themselves must be able to make such shifts with ease and naturalness, thus indicating what is expected of the test taker. Moreover, when testers engage the test takers in a discussion on highly abstract topics, it is critical that their own level of language reflect what is expected of the test takers in terms of sophistication, precision, rhetorical speech devices, native cultural references, etc.

At Level 4, socio-cultural elements of the language tested must be part of the speech sample. If they do not occur naturally and spontaneously during the conversation, the tester can elicit them directly and ask the test taker to explain the meaning of proverbs, sayings, idioms, colloquialisms, etc.

Training testers adequately is one of the prerequisites for increasing the reliability of tests. The higher the level tested, the more skill required by the testers. Testers should be able to tailor their language naturally and effortlessly. They should also be able to change registers smoothly to keep test questions abstract and formal, or colloquial and informal, as needed to support the test objectives. Role plays require testers who can themselves play the role of youngsters, elderly people, high-ranking officials, uneducated people, etc. However, for the role play to be persuasive, the tester should not assume any role in which s/he is not comfortable.

Testers who are not native speakers of the language they test, may have to rely more on scripted questions and preludes, as they might have difficulty producing unrehearsed language of required level on the spot. Recognizing Level 4 speech and being able to distinguish it from high Level 3 speech will be of importance. Extra-linguistic knowledge, i.e. knowledge of and the ability to discuss with test takers a wide variety of abstract topics also plays a significant role in one's ability to conduct tests at Level 4. One important consideration is face validity of Level 4 speaking tests. If test takers' level of proficiency is higher than Level 3, they will most likely notice if the testers lack the required proficiency or sound less proficient than the test taker.

Training itself should be conducted by trainers who have sufficient experience with testing at this level, and whose level of proficiency is, at least close to that of a well educated native speaker of the target language.

- *Developing Rating Criteria*

The rating of any speech sample should be holistic by using the STANAG 6001, Ed. 4 level speaking descriptors as a rating criterion. To facilitate the rating process it is recommendable to develop a rating factor grid, extracted from the Level 4 speaking descriptor and capturing the factors contributing to the final rating. The rating factor grid should reflect the handling of topics and tasks, lexical and structural control, fluency, pronunciation, socio-linguistic appropriateness, discourse organization, etc. All these factors will contribute to the final rating; however, some of them may play a more important role than others. At Level 4, the ability to interact culturally and not only linguistically plays an important role and is usually assessed through the above type role-play situations. For that reason, the content and context of the role plays may vary significantly from one target language to another.

- *Level 4 Speech Sample*

Follow the link below to listen to a clip taken from a Level 4 oral proficiency interview.

[insert here link to audio file [Europe.mp3](#)]

4.3 Testing Reading at Level 4

- *Selecting Appropriate Texts*

When selecting texts for reading proficiency tests, it is useful to consult Clifford's *Overview of Text Characteristics* (Child, *et al.*, 1993). The Overview captures texts from the perspectives of the author purpose, text type and reader purpose. The table below is an adapted version, illustrating these three approaches to text classification.

Table 3: Overview of reading text modes.

OVERVIEW OF TEXT MODES BY LEVEL - READING

(At each level, these elements must be aligned)

Level	Author Purpose	Typical Text Type	Reader Purpose
1	Orient by communicating main ideas.	Simple short sentences with simple vocabulary. Sentences may be re-sequenced without changing the meaning of the text.	<i>Orient oneself</i> by identifying topics and main ideas.
2	Instruct by communicating factual information.	Connected factual discourse with compound and complex sentences dealing with factual information. Sentences are sequenced within paragraphs, but the paragraphs might be re-sequenced without changing the meaning of the text. The identity of the author is not important.	<i>Learn</i> by locating and understanding not only the main ideas, but also supporting details such as temporal and causative relationships.
3	Evaluate situations, concepts, conflicting ideas; present and support arguments and/or hypotheses with both factual and abstract reasoning.	Multi-paragraph prose on a variety of unfamiliar or abstract subjects such as might be found in editorials, official correspondence, and professional writing. References may be made to previous paragraphs, to common cultural values, etc. The "voice" of the author is evident.	<i>Learn</i> by relating ideas and conceptual arguments. Understand the text's literal and figurative meaning by reading both "the lines" and "between the lines". 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, or present complex ideas with nuanced precision and virtuosity.	Extended discourse that is tailored for the message being sent and the intended audience. To achieve the desired tone and precision of thought, the author will often demonstrate the skillful use of low-frequency vocabulary, cultural and historical concepts, and an understanding of the audience's shared experience and values.	<i>Read "beyond the lines"</i> , 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.

An example of a text that meets the above criteria in terms of author purpose, text type and reader purpose can be found in Appendix B .

- *Selecting Appropriate Tasks*

Once the level of the text itself is confirmed, the tasks tested can be extracted from the Level 4 reading descriptor. As with a listening test, test specifications would outline the types of texts and tasks that would be selected. According to the STANAG 6001 descriptor, the tasks a Level 4 reader is able to perform include:

- 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
- Understand almost all cultural references

If the article in Appendix B were to be used for testing purposes, it would need to be edited in terms of its length, as a test should contain a number of texts sampling from a variety of Level 4 topical domains. Appropriate testing techniques would be comparable to those discussed under the listening skill.

Multiple-choice items might be difficult to construct as the options might be too close to each other in potential interpretation of the text to be considered distinct and unequivocally right or wrong. Also, constructing multiple-choice options at this high level sometimes has a tendency to move away from language comprehension into problem solving and, possibly, to rely too much on one problem-solving style. Therefore, an interactive constructed response method would perhaps be a more suitable testing technique. In such a method the examinee reads the text and provides spoken answers in the form of a summary, outline, or set of responses to key questions. A major advantage of such an approach would be that it allows the tester to keep shifting the line of response until any doubts are resolved. Apparent disadvantages of such a method would be that it requires a very well-trained, experienced tester with excellent oral skills, and that responses need to be recorded to make a second independent rating possible. Alternatively, the answers might be given in writing. However, this would not solve the problem of the mixing of skills tested and rating the response. For example, a response may be rated as Level 4 in terms of the accuracy and quality of speech or writing; but even though the speech or writing itself may be at level, it still may not show that the reading comprehension is at Level 4.

- *Sample constructed response item (based on the text in Appendix B):*

- What line of argumentation does the author present for his relinquishment of morality?
Write at least 300 words.

A similar question could be used to elicit a spoken response.

4.4 Testing Writing at Level 4

- *Selecting Appropriate Tasks*

As writing is typically done both with a specific purpose in mind and for a pre-defined audience, the tasks selected for a test may loosely be related to the topical domain associated with the professional lives of the test takers, in this case, to a military environment and, more specifically, to NATO, which requires Level 4 proficiency in writing for a certain number of positions. The tasks, however, have to be broad enough that the test itself does not become a narrowly defined performance test intended for a particular military field or profession. According to the STANAG 6001 descriptor, the tasks a Level 4 writer is able to perform include:

- Counsel
- Persuade
- Advocate a position at length
- Elaborate on abstract concepts
- Express nuances and shades of meaning

- *Selecting an Appropriate Testing Method*

The ability to convince, persuade, and advise is one of the hallmarks of Level 4 writing proficiency. As such, a position paper or essay, with a well-formulated prompt as the stimulus, is the recommended format to test Level 4 writing skills. It is essential that writing prompts be formulated in such a way that they elicit the required level of language that is expected from the test taker. To elicit Level 4 written language, tasks or prompts should relate to unfamiliar, inherently complex topics requiring discussion and careful argumentation, and the use of stylistically appropriate language (register), low frequency and precise abstract vocabulary, idiomatic expressions and cultural references. Just like for any other of the skill areas, test specifications should be developed outlining in detail how the writing test would be operationalized.

- *Sample writing prompt:*

There is a tendency in the military to foster a dichotomy between training and education. Because education is considered elusive and intangible, while training can be evaluated in terms of practical outcomes, those with a predilection for the less intellectual side of the dichotomy argue for a clear military focus on training followed by applicable experience.

Others have posited that, without education, an officer cannot fully grasp the interconnected human conditions, motives, and relationships manifested as war, conflict, and peace. They argue that today's military needs warrior scholars who are not only trained to react predictably to predictable situations, but who also possess the critical and analytical skills, and the cultural background, to react adeptly and creatively to the unpredictable and the unknown. As the Canadian General Jean Victor Allard stated, "without a properly educated, effectively

trained professional officer corps, the armed forces would, in the future, be doomed at best to mediocrity, and at the worst, to disaster.”

The international organization to which you have been posted is currently planning long-range budgetary allocations for both training and education. You have been asked to write a paper arguing one side of the issue and present convincing arguments for major investments.

Members of the review board that will read and evaluate your ideas have long, detailed familiarity with this subject. You know that they do not want to read the same old ideas underlying the traditional dichotomy. Your argument should provide the decision-makers with a fresh approach that is persuasive and viable.

Length: approx. 750 words

- *Developing Rating Criteria*

The rating of any writing sample should be holistic by using the STANAG 6001, Ed. 4 level writing descriptors as a rating criterion. To facilitate the rating process it is recommendable to develop a rating factor grid, extracted from the Level 4 writing descriptor and capturing the factors contributing to the final rating. The rating factor grid should reflect the handling of topics and tasks, lexical and structural control, orthography, socio-linguistic appropriateness, discourse organization, etc. All these factors will contribute to the final rating; however, some of them may play a more important role than others.

- *Sample response to the above Level 4 writing prompt*

The following response meets the threshold requirements of the Level 4 descriptor.

By making long-range budgetary commitments to either educational or training programmes, NATO is effectively shaping its very own future. The decision whether to allocate more funds to the one or the other will not only have a profound effect on the Alliance's military effectiveness, but also on the question of its very existence. This will become increasingly clear if we discard the arguments most commonly associated with the topic of education versus training in military circles.

The discussion about whether to favour education or training in the individual development of the modern day soldier is, in essence, a conflict about which of the two better contributes to his or her military effectiveness. Although they reach completely different if not contradictory conclusions in placing that effectiveness at the centre of their attention, both sides basically share the same perspective. NATO, being faced with the question of how to make most of the funds at its disposal, should choose an entirely different perspective.

A small and easily accessible example illustrates quite well why a change in perspectives is necessary and what perspective we should change to. If I were asked the question of where to allocate NATO's funds to, first as a NATO staff

member and then as a *German Air Force officer*, I would have to give not only different, but also contradictory answers. The interests of my country do not necessarily align with those of the Alliance.

If you understand education not only as a means of neutrally improving an individual's ability to autonomously analyse a given situation, to react flexibly and to find creative solutions, but also as an opportunity to shape and change a person's views and opinions on social, political and/or other issues, you come to realise how powerful a tool education is. As a *German Air Force officer* I would therefore, try to hold on to that powerful tool and insist that education remain a national privilege at least in the early stages of an officer's career. I would look to NATO for advanced military training in order to improve the interoperability and effectiveness of my forces and to benefit from the lessons learned by NATO forces in ongoing missions all over the world. To put it simply, I would allow NATO to teach *German soldiers* how to fight better while reserving the right to give their social, political and moral guidance to the *German Armed Forces* exclusively. I would prefer my soldiers to always represent *German interests* first. As a NATO staff member I would, in contrast, be interested in creating soldiers that, at least as long as they are on a NATO assignment, understand themselves to be NATO soldiers first. The interests of NATO might differ from those of each individual member state, but NATO has to rely especially on its staff members to solely represent the Alliance's interests, nevertheless. Questions of loyalty and allegiance are at the heart of the matter.

The unique system of staffing headquarters and facilities NATO has adopted adds to the problem. Staff members rotate into a post after they had been trained and educated in a strictly national environment for years. It takes time for them to adjust and to adopt not only the procedures of NATO but also its interests. Just as the transformation is completed and the officer starts to consider the Alliance's interests before promoting those of the member state he belongs to, he will rotate out of the post and leave NATO. The transition phases in and out of NATO are crucial. Ideally, an officer not only adapts quickly, but also, and that is of even bigger importance, continues to represent NATO's interests after he has been reassigned to a national staff. This is exactly what NATO should aim for officers that have adopted a NATO mindset not only for the duration of this NATO assignment but also beyond.

By deciding whether to allocate funds to education or training, NATO chooses between either investing in the effectiveness of the armed services of its member states or the dissemination of its ideas. The former will render NATO's fighting force more capable of confronting an armed threat, whereas the latter will secure its continued existence. While investments in training will have a palpable effect almost immediately, the benefits of investment in education will only become visible in the long term. Since NATO is about to make what are going to be long-range budgetary commitments, its own interests should have priority over the interests of its individual member states and short-term effects should yield to long-term benefits. Let us invest into the future of NATO, let us invest in NATO's own educational programmes and facilities.

4.5 Pre-testing and Test Administration

- *Challenges of Pre-testing Level 4 Tests*

Any test development project must include pre-testing items on candidates. Pre-testing is conducted mainly to collect data on the validity (meaningfulness of scores) of the test before its actual live administration (Bachman & Palmer, 1996). Ideally, items will first be piloted on a small number of candidates – preferably, native speakers – to collect comments that would help improve the quality of the items. Subsequently, items should be trialled on a larger scale under actual administration procedures on a representative group of candidates.

The information collected during large scale pre-testing may reveal problems with the items themselves and how well they discriminate between high and low achievers. They may also reveal issues with the actual testing environment in terms of the quality of lighting, noise at test site, etc. Most importantly, pre-testing will provide information regarding appropriate time allocations, that is, test administration time. What language testing experts generally agree on (*cf.* Bachman & Palmer, 1996; Alderson *et al.*, 2001; Bachman & Kunnan, 2005) is that this information will be meaningful only if the pre-test population resembles as closely as possible the actual test population with regards to its background and language proficiency.

The scope and the stakes involved will determine how extensively the items or the test should be trialled. Although no definite number has been determined for statistically significant information to be yielded, it is understood that quantitative evidence will reveal trends or issues only on a sufficiently large number of candidates of similar background and proficiency. According to some sources, 30 candidates per level tested would be a minimum; however, if that is not feasible, the rule of thumb regarding the number of candidates in trialling would be ‘the more the better’.

Indeed, this aspect poses a real challenge when pre-testing at Level 4. As aforementioned, Level 4 language users can be found in taxing academic and professional settings. Even in these settings where performance rather than proficiency is involved, candidates may have uneven profiles, i.e. Level 4 only in some skills s/he uses more extensively.

Another issue to consider is the external measurement tool used to determine candidates’ proficiency at level to qualify as ‘guinea pigs’ and against which to compare results. If there is no other validated test available, screening candidates for eligibility to participate in the trialling will be problematic. In that case, a screening tool will have to be developed to discriminate between Level 3 candidates and those that appear to be higher than Level 3 (see the section on pre-screening Level 4 candidates below) .

Pre-testing results will provide data regarding the discrimination factor of test items if a sufficient number of true Level 4 test takers are identified . Even though Level 4 texts and passages contain highly individualistic qualities in which meaning is personal and the receiver’s framework of world background and extra-linguistic knowledge may vary, true

Level 4 language users can produce and comprehend Level 4 documents in many non-specialties.

Careful planning and use of available resources in terms of personnel qualified for the initial, informal piloting and a sufficient number of candidates available for large scale trialling represent a significant challenge. Most nations do not have native speakers available nor a sufficient number of high level candidates. Provided that among the personnel possessing high level proficiency a sufficient number were to be found as suitable candidates for the test trialling, those same candidates would most likely be the ones who would take the same tests once they become official, and that would pose a sincere threat to the security of the test itself (test compromise). This problem would become even more significant if Level 4 tests are needed in languages other than English, especially less commonly spoken languages. Although networking with other countries is still an option, a large number of administrations of Level 4 testing is not foreseen to justify the costs of such an endeavor.

- *Standardizing Test Administration Procedures*

Standardizing test administration procedures is another contributor to test reliability. This is especially important if an organization administers tests in several languages. From one language to another, tests should be administered in the same way and according to the same protocols. Allotting sufficient time to tester training would ensure that the OPI at Level 4 is administered in a standardized fashion and all phases of the interview covered. Test rating procedures would also follow standardized and pre-defined protocols. If one or more skills is tested in conjunction with another skill – for example, if speaking is tested partly through reading – it is very important to develop testing and rating protocols reflecting that. If the protocols are to be applied to many different languages, suitability for each language must be reviewed.

- *Tester Qualifications and Tester/Rater Training*

Training testers/raters for oral proficiency test administration at any level requires significant resources in terms of time required for initial training as well as subsequent standardization sessions. It is imperative that Level 4 oral interviews be conducted by testers who are not only fully trained, but also whose own oral proficiency is at least at Level 4. Even if an oral proficiency test were to be fully scripted (which would not be advisable in the first place), testers would need to have the ability to rate responses and distinguish the examinees performing at Level 4 from those below that level. Rating itself would require Level 4 proficiency.

The higher the tested level, the more demands are placed on testers not only in the linguistic sense, but also in terms of extra-linguistic knowledge that is required to successfully move from one high level, abstract topic to another, as well as to make appropriate shifts in registers reflecting the socio-linguistic and cultural norms of the target language. There is anecdotal evidence of cases in which face validity of oral

proficiency tests was threatened after some test takers had noticed that the testers' proficiency seemed to be lower than their own.

Achieving and maintaining inter-rater reliability (consistency in rating between raters) will also be of critical importance, especially in the cases in which oral proficiency is assessed by using reading passages as prompts. Decisions have to be made whether the spoken responses have to demonstrate clearly all the features of the speaking skill at Level 4 plus the correct interpretation of the meaning found in the text itself. In order to enhance test reliability it is recommended that Level 4 oral proficiency interviews are conducted by two testers. Some testing organizations have extensive quality control programs. Not only are there two testers, but also randomly selected and specifically targeted ratings by a third rater to maintain an on-going quality control system.

- *Pre-screening Level 4 Candidates*

Since testing at Level 4 is even more challenging and labour-intensive than at lower proficiency levels, testing should preferably be restricted to candidates who are potentially at this level. One way of pre-screening candidates for Level 4 testing in the listening and reading skills is to have them take a Level 3 tests. As all levels are ranges, it is expected that Level 4 candidates would achieve a (near-)perfect score on the Level 3 tests, indicating thus that their performance significantly exceeds the threshold level. One important consideration in identifying accurately Level 4 candidates is to ensure that Level 3 tests contain items that are of various difficulty levels within the Level 3 range itself. If they are all of threshold Level 3 difficulty, an individual attaining a (near-)perfect score may not necessarily have Level 4 proficiency and may therefore not be eligible for taking a Level 4 test. Validated Level 3 tests should also be used for identifying potential Level 4 candidates on whom newly developed Level 4 tests could be pre-tested. To ensure proper identification, a (near-)perfect score should be expected for trialling selection purposes.

The same would apply to the speaking and writing tests at Level 3. Candidates displaying an exceptionally strong performance at Level 3 could be given the opportunity to be tested at Level 4 for official purposes. Some institutions administer adaptive, multi-level OPIs and the testers automatically raise the level of the test if they have enough evidence that the candidate is fully successful at Level 3. In case the OPI does not go beyond Level 3 and a candidate shows that s/he might be able to satisfy the requirements of Level 4, a separate Level 4 OPI could then be administered.

References

- Alderson, J.C., Clapham, C. and Wall, D. (2001). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Bachman, L.F. and Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L.F. and Kunnan, A.J. (2005). *Statistical Analyses for Language Assessment (Workbook and CD-ROM)*. Cambridge: Cambridge University Press.
- 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, Clifford, R., and Lowe, P. Jr. (1993). Proficiency and performance in language testing. *Applied Language Learning* 4 (1&2), 19-54.
- Clymer, T. (1968). What is reading?: Some current concepts. *67th Yearbook of National Society for the Study of Education*, 7-29.
- Flege, J.E. and Liu, S. (2001). The effect of experience on adult's acquisition of a second language. *Studies in Second Language Acquisition*, 23, 527-552.
- Holtgraves, Th.M. and Kashima, Y. (2007). Language, meaning, and social cognition. *Personality and Social Psychology Review*, 12, 73-94.
- Leaver, B.L. and Shekhtman, B. (eds.) (2002). *Developing professional-level language proficiency*. Cambridge: Cambridge University Press.
- Lye, J. (1996). *The problem of meaning in literature*. Retrieved 27 August 2011 from <http://www.brocku.ca/english/jlye/meaning.php>
- Snow, C. (2002). *Reading for understanding: Towards an R&D program in reading comprehension*. Santa Monica, CA: RAND

Working Group on Level 4 Proficiency - *members and contributors*

Working Group Leaders: Ms. Jana Vasilj-Begovic, BILC Canada
Mr. Gerard Seinhorst, Netherlands

Special Consultant: Dr. Martha Herzog, USA

Members and Contributors: Ms. Emilija Nesheva, Bulgaria
Ms. Ulla Gudnason, Denmark
Dr. Dugald Sturges, Germany
Ms. Mary Jo Di Biase, Italy
Mr. Michael Adubato, SHAPE
Ms. Annette Nolan, Sweden
Ms. Sabine Atwell, USA
Mr. James Dirgin, USA
Mr. David Oglesby, USA
Ms. Peggy Garza, BILC USA

Appendix A

Trisections

All proficiency level descriptors in STANAG 6001 can be divided into three areas describing the content, tasks and accuracy of language use. Content refers to topical domains one can deal with when using the foreign language. Tasks refer to what one can do with the language, and accuracy to how well one uses the language. The breakdown below illustrates these three aspects of the Level 4 descriptors for listening, speaking, reading, and writing.

Listening Level 4

CONTENT:	<p>All forms and styles of speech used for professional purposes</p> <p>Highly sophisticated language including most matters of interest to well-educated native speakers</p> <p>Language used in representation of official policies, lectures, negotiations</p> <p>Language tailored for various audiences, including persuasion, representation, and counselling</p> <p>Highly abstract concepts</p>
TASKS:	<p>Adjust to shifts of subject matter and tone</p> <p>Follow unpredictable turns of thought in both formal and informal speech on any subject matter addressed to the general listener</p> <p>Recognize nuances of meaning and stylistic levels, irony, humour</p>
ACCURACY:	<p>Readily understands language in media and in conversations among native speakers, both globally and in detail</p> <p>Generally comprehends regionalisms and dialects</p>

Speaking Level 4

CONTENT:	<p>Matters of interest to well-educated native speakers</p> <p>Highly abstract concepts in such areas as economics, culture, science, technology, and his/her own professional field</p>
TASKS:	<p>Tailor speech to a specific audience</p> <p>Counsel</p> <p>Persuade</p> <p>Advocate a position at length</p> <p>Communicate with native speakers at conferences, negotiations, lectures, briefings, debates</p> <p>Shift and adjust to shifts of subject matter and tone</p>

ACCURACY: Uses the language with great precision, accuracy and fluency
Can express nuances and make culturally appropriate references
Speaks effortlessly and smoothly, with a firm grasp of various levels of style, but seldom to be perceived as a native speaker
Any shortcomings, such as non-native pronunciation, do not interfere with intelligibility

Reading Level 4

CONTENT: 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 on topics such economics, culture, science and technology
Highly abstract concepts

TASKS: 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

ACCURACY: Can relate a specific text to other written materials in the culture
Demonstrates a firm grasp of stylistic nuances, irony, and humour
Understands reasonably legible handwriting

Writing Level 4

CONTENT: All professional purposes including the representation of an official policy or point of view
Private letters, job-related texts, reports, position papers, the final draft of a variety of papers

TASKS: Use the written language to persuade others and to elaborate on abstract concepts

ACCURACY: Can write precisely and accurately for all professional purposes
Organizes extended texts well, conveys meaning effectively, and uses stylistically appropriate prose
Shows a firm grasp of various levels of style and can express nuances and shades of meaning.

Appendix B

Sample Level 4 Reading Text

The article below is taken from the magazine *Philosophy Now*⁵. Even though the text is of a philosophical nature, it is written by a columnist whose regular readers would be familiar with his general ideas. The author's purpose of the article is to update his readers on some of his new insights into morality.

To comprehend the text, the reader must be able to read "beyond the lines", i.e. interpret the text in its wider context and understand the complex ideas expressed by the author.

AN AMORAL MANIFESTO (Part I)

A special extended column from our (erstwhile) Moral Moments columnist **Joel Marks**.

I. Hard Atheism or What Shall I Name This Column?

Hold onto your hats, folks. Although it is perhaps fitting that the actual day on which I sit here at my computer writing this column is April 1st, let me assure you that I do not intend this as a joke. For the last couple of years I have been reflecting on and experimenting with a new ethics, and as a result I have thrown over my previous commitment to Kantianism. In fact, I have given up morality altogether! This has certainly come as a shock to *me* (and also a disappointment, to put it mildly). I think the time has come, therefore, to reveal it to the world, and in particular to you, Dear Reader, who have patiently considered my defenses of a particular sort of moral theory for the last ten years. In a word, this philosopher has long been laboring under an unexamined assumption, namely, that there is such a thing as right and wrong. I now believe there isn't.

How I arrived at this conclusion is the subject of a book I have written during this recent period (tentatively titled *Bad Faith: A Personal Memoir on Atheism, Amoralism, and Animals*). The long and the short of it is that I became convinced that atheism implies amorality; and since I am an atheist, I must therefore embrace amorality. I call the premise of this argument 'hard atheism' because it is analogous to a thesis in philosophy known as 'hard determinism.' The latter holds that if metaphysical determinism is true, then there is no such thing as free will. Thus, a 'soft determinist' believes that, even if your reading of this column right now has followed by causal necessity from the Big Bang fourteen billion years ago, you can still meaningfully be said to have freely chosen to read it. Analogously, a 'soft atheist' would hold that one could be an atheist and still believe in morality. And indeed, the whole crop of 'New Atheists' (see [Issue 78](#)) are softies of this kind. So was I, until I experienced my shocking epiphany that the religious fundamentalists are correct: without God, there is no morality. But they are incorrect, I still believe, about there being a God. Hence, I believe, there is no morality.

Why do I now accept hard atheism? I was struck by salient parallels between religion and morality, especially that both avail themselves of imperatives or commands, which are intended to apply universally. In the case of religion, and most obviously theism, these commands emanate from a Commander; "and this all people call God," as Aquinas might have put it. The problem with theism is of course the shaky grounds for believing in God. But the problem with morality, I now maintain, is that it is in even worse shape than religion in this regard; for if there were a God, His issuing commands would make some kind of sense. But if there is no God, as of

⁵ Retrieved 06 April 2012 from http://philosophynow.org/issues/80/An_Amoral_Manifesto_Part_I.

course atheists assert, then what sense could be made of there being commands of this sort? In sum, while theists take the obvious existence of moral commands to be a kind of proof of the existence of a Commander, i.e., God, I now take the non-existence of a Commander as a kind of proof that there are no Commands, i.e., morality.

Note the analogy to Darwinism. It used to be a standard argument for God's existence that the obvious and abundant design of the universe, as manifested particularly in the elegant fit of organisms to their environments, indicated the existence of a divine designer. Now we know that biological evolution can account for this fit perfectly without recourse to God. Hence, no Designer, no Design; there is only the appearance of design in nature (excepting such artifacts as beaver dams, bird nests, and architects' blueprints). Just so, there are no moral commands but only the appearance of them, which can be explained by selection (by the natural environment, culture, family, etc.) of behavior and motives ('moral intuitions' or 'conscience') that best promote survival of the organism. There need be no recourse to Morality any more than to God to account for these phenomena.

I cannot hope to make all of that convincing to my readers in the short space of a column: hence the book I have written. But even in the book I am not attempting so much to give a rigorous proof as to consider the *aftereffects* of my counter-conversion (to apply William James's term for the loss of religious belief to my loss of moral belief). What is it like to live in a world without morality? Is such a life even viable? This is what I had to discover before I could so much as walk out my front door! That is why the first draft of my book was written in an urgent rush, almost without my leaving the house. (Fortunately I am retired and *sans famille*.) I was reeling – much as, I imagine, a religious believer whose whole life has been based on a fervent belief in the Almighty, would find herself without bearings or even any ground to stand on if suddenly that belief were to vanish, no matter whether by proof or just by *poof!* Just so, morality has been the essence of my existence, both personally and professionally. Now it is no more.

Does this mean, among other things, that this column will end? I hope not! The book is only the beginning. I must learn how to live life all over again, like a child learning to walk. And just as a child growing up discovers one fascinating thing after another about the 'new' world, so the floodgates have been opened for me from a sea of possibilities. For, yes Virginia, there is life after morality, and I would like to report back to you as I experience it.

There is just one thing, though: I might have to change the name of my column. 'Moral Moments' now seems problematical, to say the least. 'Amoral Moments' would be closer to the mark (and to Marks). One thing that hasn't changed, however, as you can see, is that my writing is still filled with similes, allusions, mixed metaphors, and bad puns. Fortunately I can now rest assured that in persisting with these I am doing nothing *wrong*.

II. In the Mode of Morality

I have relinquished the mantle of the moralist since I no longer believe there even is such a thing as morality. How, then, shall one live? One thing to note is that in asking that question I am able to retain the title of *ethicist*, for ethics is just the inquiry into how to live. This suggests a new name for my column, namely, 'Ethical and other Episodes', in which I hope in due course to articulate my answer in full. But I would also like to suggest at the outset of this undertaking that, even though an amoralist, I can still engage in moral argumentation ... and in good conscience (so to speak!).

Consider that for the foreseeable future I will be living in a society that continues to pay homage to morality and believe in its reality implicitly. So I am likely to be confronted time and again by a question like, "Do you believe x is wrong?" It would usually be hopeless to attempt to refashion the question into an amoralist mode of speaking; at the very least this would change the subject from the particular issue under discussion, say, vivisection, to an abstract issue in meta-ethics, namely, whether there is such a thing as wrongness. But there is still a way I could answer the question both honestly and effectively. Thus, I could reply, "Vivisection is wrong

according to morality as I conceive it.” For that reply is *not* asserting that vivisection is wrong, only that, according to morality (as I conceive morality) it is wrong. In the abstract this has no more force than if one were to say, “Unicorns are a type of horse (according to the common conception of unicorns).” In other words, there is no implication that unicorns actually exist, nor, all the more, that, say, a person could possibly find one for the purpose of trying to ride her.

Note further that it is possible to *argue* about these things whose existence is not being asserted. Thus, I could say, “Vivisection is wrong (in my conception of morality) because it involves treating sentient beings merely as means.” This is of course a kind of Kantian justification for my claim. And I would offer it as an argument that I believe to be perfectly sound because (1) it articulates the analysis of morality that I consider to be the correct one, namely, Kant’s categorical imperative (suitably modified to accommodate nonhuman animals), (2) it characterizes vivisection in a way that I consider to be correct, namely, as violating the Kantian imperative, and (3) it logically draws its conclusion therefrom. Again this would be just as if I had argued, “Santa Claus could not possibly be mistaken for Popeye because Santa Claus has a big beard while Popeye is barefaced.”

Thus, I am become like the father in this joke – courtesy of my attorney’s rabbi – about a Jewish boy from a liberal family who attends the neighborhood parochial (Christian) school:

One day Isaac comes home in great puzzlement about what he had been taught in school that day; so he goes to his father and asks him about it.

“Father, I learned that God is a Trinity. But how can there be three Gods?”

“Now get this straight, Son: We’re Jewish. So there is only one God... and we don’t believe in Him!”

Just so, I no longer believe in morality (like God in the joke), but I would still insist that the nature of morality is Kantian (monotheism in the joke) rather than utilitarian (Trinitarianism in the joke).

Now, if I were to employ this technique without elaboration, it could easily be part of a deceptive strategy, since it is likely that people would assume I was defending something outright rather than only hypothetically. A statement like “If *anything* is wrong, *this* is” is naturally interpreted as a rhetorical emphasis of just how wrong the speaker considers *this* to be. But if I, as an amoralist, were to say “If anything is wrong, vivisection is,” I would mean it literally, not rhetorically; that is, the ‘if’ would have real force for me, even suggesting that I do *not* believe that anything is wrong (since morality does not exist): all the more, that I do not believe that vivisection is wrong. (Of course that does not mean I think vivisection is *right* or even *permissible*, since those are moral notions also. I just *don’t like* vivisection.) So my intention in making the utterance would be at variance with the impression it would leave in my listener’s mind; and knowing this, I would be a deceiver.

However, if I were only trying to persuade a Kantian vivisectionist of the error of her ways, its usage, it seems to me, would pass muster even morally. I would be using reasoning to show my interlocutor that what she was doing violated *her own* moral/theoretical commitments. My own view of morality itself would be irrelevant; my interlocutor can assume what she likes about my meta-ethics. It would be exactly as if I were talking with a religious believer about the proper treatment of other animals: whether or not the believer knew I was an atheist, it would be perfectly proper for me to try to convince her that there is Biblical support for a benign ‘stewardship’ of other animals – would it not? I need not believe in the concept of stewardship myself, nor in its divine sanction, in order to invoke it undeceivingly when arguing with someone who does. Just so, it seems to me, morality.

Rather aptly, I now realize, I have been led to a sort of Socratic mode of moral argumentation. Socrates was notorious for interrogating his interlocutors rather than asserting and defending theses himself. Similarly, I am suggesting, I will continue to be able to hold forth as a critical moral reasoner, even though I no longer believe in morality, so long as I confine myself to questioning the inferences of others (and gingerly deflect their questions about my own moral

commitments by speaking *in the mode of* morality, as above). It is true that I would thereby fail to be completely forthcoming about my own meta-ethics whenever doing so would be disruptive to the dialogue; but I do not think I would be doing anything that is considered unkosher even when moralists are arguing among themselves. After all, my meta-ethics could be mistaken; maybe there is such a thing as morality. So my 'suspension of disbelief' could be conceived as an expression of intellectual humility, and my arguments considered in themselves by the intellectual light of my interlocutor.

The bottom line for me, as both a philosopher and the possessor of a particular personality, is that I do not 'suffer fools gladly.' This has always been true of me, but it used to be supplemented by a belief (or assumption) in morality. Now that I have turned the philosophic eye on my own largely unexamined assumption that morality exists, I see that I have been a moral fool. But I retain my belief (or assumption) in Truth as such, as well as my pig-headed allegiance to it. Thus, I shall henceforth apply a skeptical scalpel to the moral arguments of all, unsparing even of the ones I have been sympathetic to as a moralist, since all of them, I now believe, are premised on a bogus metaphysics. For it is intellectual dishonesty or *naïveté* that I am most temperamentally disposed to dislike, even as I retain my passionate preferences for certain 'causes,' such as animal liberation.

III. Desirism

I have explained how an amoralist, such as I have become, could still continue to argue in the mode of morality. Although this risks being deceptive and hypocritical, it can also be done aboveboard because the amoralist could be appealing to his or her interlocutor's (or reader's) moralism. This is analogous to how a native speaker of English might nonetheless, with some knowledge of other languages, be able to point out a grammatical mistake being made by someone speaking in French. Thus for example, if I were conversing with someone who believed that meat-eating is morally good because it promotes the greatest good of the greatest number, I could point out that this utilitarian credo is supposed to apply to all sentient beings and not only to human beings; so that if one tallied up the net pleasure and pain being experienced not only by the human meat-eaters but also by the animals being bred and slaughtered for eating under the current regime of factory farming, one would likely conclude that eating meat does *not* lead to the greatest good and hence is wrong. Meanwhile, I myself, as an amoralist, believe meat-eating is neither right nor wrong; but I would have done nothing dishonest in convincing my interlocutor that it is wrong, that is, by her lights.

But why would I even care whether I was being honest or not? Isn't that, again, something an amoralist would be indifferent to? Strictly speaking, yes. But an amoralist still has a compass, a 'guide to life', an *ethics*, or so I would argue; and it can be a match for anybody's morality. Thus, consider that in purely practical terms, honesty may still be the best policy. A reputation for truth-telling will likely make one a more attractive person to do (literal or figurative) business with, which will enable one to thrive relative to one's less scrupulous competitors. Thus, 'survival of the fittest' could naturally promote honesty as a prevalent trait even in the absence of any moral concern.

There I am, then, honestly discussing particular issues with opponents, and justifying my positions to them by *their* moral lights. But how do I justify them *to myself*, since I have no moral lights anymore? For example, on what basis would *I myself* be a vegetarian? The answer, in a word, is desire. I *want* animals, human or otherwise, not to suffer or to die prematurely for purposes that I consider trivial, not to mention counterproductive of human happiness. For the vast majority of human beings in the world today, meat-eating is a mere luxury or habit of taste, while at the same time it promotes animal cruelty and slaughter, environmental degradation, global warming, human disease, and even human starvation (the latter due to the highly inefficient conversion of plant protein to animal protein for human consumption). For whatever reason or reasons, or even no reason, these things matter to me. Therefore I am motivated to act on the relevant desires.

But if I were conversing with another amoralist, how would I convince *her* of the rightness of my desires? Well, of course, I wouldn't even try, since neither of us believes in right, or wrong. What I could do is take her through the same considerations that have *moved* me to my position and hope that her heartstrings were tuned in harmony with mine. If the two of us have grown up in the same culture, we will certainly have many desires in common. For example, we may both be averse to animal suffering and cruelty to animals. But even within the same society, there can be large differences in *knowledge*. I speak from personal experience regarding even my own knowledge, for, to stay with my example, I was blissfully unaware of the horrors of factory farming until only a few years ago. Most people in our society continue to be, even though the practice has been prevalent for the last fifty years. Thus, there is a good chance that I would be able to influence my interlocutor's carnivorous desire and behavior simply by introducing her to the relevant facts. The absence of a moral context, therefore, need not be harmful to my hitherto-moral project of honestly promoting vegetarianism.

But what if my amoral interlocutor were just as versed in the facts of factory farming as I but still did not care about animal suffering, or simply loved eating meat *more than* she loved animals? At this point the dialogue might serve no purpose. But that certainly would not mean that I had no further recourse, even honest recourse. For example, I could try to bring around as many *other* people as possible to my way of seeing (and feeling) things, so that ultimately by sheer force of numbers we might reduce animal suffering and exploitation by our purchasing practices and voting choices. In this effort I could join with others to employ standard methods of 'marketing,' such as advertising campaigns and celebrity endorsements. These things are not inherently dishonest simply in virtue of being strategic. (And of course if I did not value honesty, additional tactics would become available to me.)

I conclude that morality is largely superfluous in daily life, so its removal – once the initial shock had subsided – would at worst make no difference in the world. (I happen to believe – or just hope? – that its removal would make the world a *better* place, that is, more to our individual and collective liking. That would constitute an argument for amorality that has more going for it than simply conceptual housekeeping. But the thesis – call it 'The Joy of Amorality' – is an empirical one, so I would rely on more than just philosophy to defend it.)

A helpful analogy, at least for the atheist, is sin. Even though words like 'sinful' and 'evil' come naturally to the tongue as a description of, say, child-molesting, they do not describe any actual properties of anything. There are no literal sins in the world because there is no literal God and hence the whole religious superstructure that would include such categories as sin and evil. Just so, I now maintain, nothing is literally right or wrong because there is no Morality. Yet, as with the non-existence of God, we human beings can still discover plenty of completely-naturally-explainable internal resources for motivating certain preferences. Thus, enough of us are sufficiently averse to the molesting of children, and would likely continue to be so if fully informed, to put it on the books as prohibited and punishable by our society.

Joel Marks is Professor Emeritus of Philosophy at the University of New Haven in West Haven, Connecticut.

[page intentionally left blank]