ENGLISH LANGUAGE PROFICIENCY ASSESSMENT IN THE MILITARY CONTEXT: TEST DEVELOPMENT AND NATO STANAG 6001

Zverev Ilya
Uzbekistan Ministry of Defence
Partnership for Peace Training Centre
English Language
domusauri@gmail.com

Abstract. The principal focus of the article is the development of the English language proficiency test items for NATO nations military and civilian employees of international staffs using NATO STANAG 6001 Agreement.

Key words: NATO, standardization, STANAG 6001, assessment, measurement, evaluation, testbaselevel, threshold level, scale, trisection.

The Presidential Decree of 10 December 2012 "On the measures to further develop foreign language learning system" accompanied by a respective Decree by the Uzbekistan Cabinet of Ministers on 8 May 2013 adopting the new State Educational Standard for Foreign languages caused a significant rise of interest among both professionals and lay people alike to the Common European Framework of References for Languages, a comprehensive document that now virtually underlies the entire foreign language teaching and learning system of the Republic of Uzbekistan.

There is, however, a general lack of recognition of the fact that instruments provided by the CEFR are by no means unique or universally applicable. The absolute majority of European militaries, in fact, does not use CEFR-based levels of foreign language proficiency examinations delivered to their service members or civilian employees. CEFR textbooks cannot, therefore, be exclusively used for foreign language classes delivered to soldiers, commissioned and non-commissioned officers. The absolute majority of European nations being members of NATO, the principal document guiding their approaches for foreign language training of their military personnel is NATO STANAG 6001.

North Atlantic Treaty Organization (NATO) is a military alliance of 28 independent nations founded upon the principle of collective defense, in accordance with which an attack against one member of the Alliance will be regarded as an attack against them all. Since the very beginning of its existence, one of the major problems that NATO members faced was that of *interoperability*, i.e. effective employment of military forces provided by two (or more) of the Alliance members in order to achieve the effect sought. In order to overcome this hindrance, NATO launched the process of **standardization** (or normalization), which is officially defined as "the development and implementation of concepts, doctrines, procedures and designs in order to achieve and maintain the compatibility, interchangeability or commonality which are necessary to attain the required level of interoperability, or to optimize the use of resources, in the fields of operations, materiel and administration"[6].

Since the biggest hindrance to actually achieving the interoperability desired was the lack of language skills among representatives of various NATO nations, in 1966, as a scion of the standardization process, NATO Bureau for International Language Coordination (NATO BICL), which was "to promote and foster interoperability among NATO and PfP nations by furthering standardization of language training and testing, and harmonizing language policy"(1), was established. During 1973 through 1975 BILC developed the first draft of **language proficiency standards**, which came to be known as STANAG 6001 (or Standardization Agreement 6001).

NATO Standardization Agreement 6001 (currently in its 5th edition) provides the user with "definitions of standardized proficiency levels" in 4 commonly recognized macro language skills (listening, speaking, reading, writing) that fall into two major proficiencies (oral and written respectively). NATO STANAG 6001 was adopted in order to ease communication between participating nations on language requirements for international staff appointments, to record and report (in international correspondence) language proficiency measures, as well as to compare national standards through a commonly-agreed standardized table.

STANAG 6001 describes six **base language proficiency levels in between** 0 (no proficiency) and 5 (highly articulate native speaker). Each of those levels is a **threshold**. Later on the BILC Working Group on Testing and Assessment developed intermediary proficiency level descriptors that came to be known as **plus levels**. Those plus levels (e.g. 2+) indicate that a candidate's ability does exceed the requirements of the base level, but does not "fully or consistently meet all of the criteria for the next higher base level" [6]. Plus levels might also be regarded as an indication of the fact that candidate managed to demonstrate the language characteristic of the next level, but could not sustain at that level.

There is a set of separate description of performance at each level provided for each skill area. When those are taken together in the order *Listening – Speaking – Reading – Writing*, the 4 digits will form what is usually referred to as **Standardized Language Profile**. For example, SLP 3212 means level 3 in listening, level 2 in speaking, level 1 in reading and level 2 in writing.

Similar to the Common European Framework of Reference, NATO STANAG 6001 is not a language test, but rather an attempt made by the language experts to create a **language proficiency scale**. Since language proficiency scales do usually describe what is often termed "typically occurring language behavior", their application for construction of language proficiency tests requires judgment on part of both the language designer and test administrator. It should also be noted that NATO STANAG 6001 does not fit neatly into any one of the four generally recognized types of scales (nominal, ordinal, interval, ratio). There is no equidistance (a principal characteristics of the interval scale) between the language proficiency levels as presented in STANAG 6001: the scope of the intervals increases in geometric progression between the Levels 0 – 5. Ergo, this scale can be characterized as "geometric" or "expanding". Furthermore, a STANAG 6001 Proficiency Level represents a range: there might be a great number of various combinations of strengths and weaknesses within a given level (two candidates that might strike one as fundamentally different in their language ability can be assigned the same STANAG 6001 Level. Finally, seem as it may that STANAG 6001 represents a ratio scale, it cannot be reasonably claimed that a candidate with Level 4 has two times as much language proficiency as one with Level 2.

Having thus established the fact of STANAG's being an instrument, which can be applied to design and develop language proficiency test, it behooves us to provide a characterization of any language test and to further proceed to discuss some peculiarities of applying STANAG 6001 to fit one's specific set of requirements.

The term language test is often invoked with a number of other terms: assessment, measurement and evaluation.

Assessment is the "process of collecting information about a given object of interest according to procedures that are *systematic* and *substantially grounded*" (3, p. 7). In terms of language proficiency assessment, the object is often a certain aspect of this proficiency. Systematicity of an assessment is achieved via its being conducted in accordance with strict and clearly defined procedures guaranteeing its potential replicability, whereas its substantial grounding is an outcome of an assessment's being based upon "a widely accepted theory about the nature of language ability, language use or language learning"(3, p.7).

Measurement is the "process of quantifying the characteristics of an object of interest according to explicit procedures and rules" (2, p. 18), i.e. measurement is a

type of assessment based upon assigning numerical values to certain traits of elicited behavior.

Evaluation is one "possible use of assessment" involving "making value judgments and decisions" (3, p. 9).

A test is a "procedure designed to elicit certain behavior from which one can make certain inferences about certain characteristics of an individual" (5, p. 48). Any test is a method of measurement, which in its turn is a method of assessment. Results of the test can be used for evaluation of the potential candidate in terms of his/her fitness for a particular position. Any test must be valid and reliable. Validity of a test reflects its measuring of what it is supposed to measure, whilst reliability signifies the consistency of this measurement over time. A test may be valid but not reliable or it can be reliable but not valid.

It has been already mentioned that STANAG 6001 does not constitute a test per se, but rather a set of instruments that can be used to construct a number of test items that can be later united under the umbrella of a language proficiency test. The issues pertaining to test development in general being beyond the scope of the article, let us consider a way of applying STANAG 6001 descriptors for the purpose of constructing of test items.

Descriptor of each of the skills for a particular level in STANAG 6001 might seem to be daunting and hardly comprehensible at first sight, but in fact each of them can be easily parsed in a number of easily-used constituents: content, accuracy, task (CAT). Those 3 components for level *trisection* for a particular skill.

Consider, for example, the following descriptor of listening skill for STANAG Proficiency Level 1: "[c]an understand common familiar phrases and short simple sentences about everyday needs related to personal and survival areas such as minimum courtesy, travel and workplace requirements when the communication situation is clear and supported by context. Can understand concrete utterances, simple questions and answers, and very simple conversations. Topics include basic needs such as meals, lodging, transportation, time, simple directions and instructions. Even native speakers used to speaking with non-natives must speak slowly and repeat and reword frequently. There are many misunderstandings of both the main idea and supporting facts. Can only understand spoken language from the media or among native speakers if content is unambiguous and predictable."

Trisection for descriptor above would look like this:

Listening – Level 1 - Survival		
Content	Tasks	Accuracy
• Familiar phrases and short simple sentences; • Everyday needs such as	• Understand the main idea	• Even native speakers used to speaking with non-natives must speak slowly and repeat or

minimum courtesy, travel, and workplace requirements; • Concrete utterances, simple questions and answers, and very simple conversations; • Topics such as meals, lodging, transportation, time, simple directions and instructions.		reword frequently; • There are many misunderstandings of both the main idea and supporting facts; • Can only understand speech from the media or among native speakers if the content is completely unambiguous and predictable.

Having thus parsed the descriptor, one may commence his/her search for appropriate listening materials (dealing with the content specified) and develop relevant test items, which would be combined into a STANAG Level 1 Listening test.

There is no overarching STANAG-6001-based test delivered by some higher authority (e.g. BILC). Every nation that wishes to adopt it must design and develop their own tests for establishing foreign language proficiency levels for their personnel. Those tests would have to be in compliance with the fundamental principles of testing, i.e. they would have to be valid and reliable, but in the end there will never be any silver bullet applicable to all the possible range of testing problems and real-world situations.

In conclusion, STANAG 6001 is a language proficiency scale developed by NATO nations in order to increase their interoperability. STANAG 6001 distinguishes between 6 base proficiency levels accompanied by plus levels. STANAG 6001 descriptors are easily passed along the Content-Accuracy-Task line for construction of reliable and valid test items. Every nation considering applying STANAG 6001 is responsible for the development of its own set of tests based on this scale: there is no body, specifically assigned with the design and delivery of STANAG 6001 tests.

REFERENCES

1. "NATO BILC - About Us." Bureau for International Language Coordination. Accessed January 14, 2017. NATO BILC. (n.d.). Retrieved January 14, 2017, from http://www.natobilc.org/en/info/about-us/.
2. Bachman, Lyle F. Fundamental Consideration in Language Testing. Oxford. Oxford University Press, 1990.
3. Bachman, Lyle F. Statistical Analyses for Language Assessment. The Cambridge Language Assessment Series. Cambridge University Press, 2005.
4. Bloom, Paul. Why are People Different?: Differences.. Open Yale Course - Psychology. 07 Mar. 2007. Lecture.
5. Carrol, J.B. The psychology of language testing in A.Davies (ed.), Language testing symposium: a psycholinguistic approach (pp. 46 – 69). London. Oxford University Press, 1968.
6. NATO Allied Publication – AAP6 – NATO Glossary of Terms and Definitions (English and French) (2008).
7. NATO Standardization Agreement (STANAG) 6001 (Edition 4) (2010).