

BILC
NATO Benchmark Advisory Test (BAT)
Listening Test

Instructions for Contributing Nations

Voluntary Contributions

Each nation is invited to contribute **at least 2 items** for each level.

Submitting nations should send **3 files** for **each item**:

- 1- An audio file, which can be played with Media Windows Player (version 7 and above) format
- 2- The item (including the stem and the 4 options, indicating the correct answer) in a Word file format
- 3- The orientation, a verbatim transcript of the text, and an indication of the original source of the text (i.e. broadcast source and date if possible) in a Word file format.

Labeling

Each file must be named to include the name of the country, level, and a unique title.
Example: If Canada submits a level 2 text about a train accident, it should be labeled:
CAN_L2_TRAIN_ACCIDENT

Submissions

The link to the server for uploading the test items is:

<http://www.natoproject.com>

When you access the link for the first time, you will be asked to create a user ID and a password, which you will be using to submit materials later on.

More detailed instructions on the uploading procedures, guidelines and other information will follow.

Special Notes

Items that have been previously used or otherwise exposed cannot be submitted.

In order to maintain test security, once submitted, the items cannot be used for any other purpose.

Thank you

BILC Secretariat

BILC
NATO Benchmark Advisory Test (BAT)
Listening Test Specifications

GENERAL TEST OVERVIEW

Purpose of the BAT

The purpose of the BAT is to provide nations with an external criterion for validation of their general proficiency tests based on NATO STANAG 6001. This test is designed for NATO military and civilian personnel who are non-native speakers of English. The BAT assesses English language proficiency regardless of how it was acquired.

SPECIFICATIONS FOR THE LISTENING TEST

Construct Definition

One of the definitions of proficiency relates to a general type of knowledge or competence in the use of a language, regardless of how, where or under what circumstances it has been acquired. Proficiency is conceptualized here as a global construct that transfers across contexts, tasks, and events, while proficiency tests attempt to sample the underlying competence by eliciting behaviors on tests that generalize to domains of interest. Proficiency can also be defined as the language knowledge that is needed to function in a future situation. The performance elicited in a proficiency test is usually measured or judged against a set of criteria, represented in a rating scale.

The BAT will measure proficiency in listening comprehension in accordance with the STANAG 6001 (Edition 2).

Delivery of the BAT Listening Test

Computer-based (online) delivery: This test will be administered and scored online by using a tailored software product hosted internally on a server at Brigham Young University. Examinees will not be able to exercise control over the recordings.

Computer-adaptive delivery (in the future): It is foreseen that the second stage of the delivery will be in a computer-adaptive test format, as the item bank is expanded over time.

Test Format

This will be a multi-level test consisting of sixty items covering STANAG 6001 (Ed. 2) Levels 1 to 3 with 20 items at each level. The approximate length of the test will be 90 minutes. For ease of scoring, test items will be multiple-choice. Each text will be preceded by an orientation, spoken in English, and each item will consist of a stem and four options written in English. Each text will be heard only once.

After the orientation, there will be a 10-20 second pause (depending on the level of the text and the number of items) to allow the examinees to read the item(s) before they hear the text. Following the text, there will be another pause (10-20 seconds per item, depending on the level) to allow the examinees to respond.

The recommended text length for levels is as follows:

- L1= up to 45 seconds (*1 multiple-choice question per text*)
- L2= up to 60 seconds (*up to 2 multiple-choice questions per text*)
- L3= up to 90 seconds (*up to 2 multiple-choice questions per text*)

Spoken Texts

The BAT will focus on authentic texts (spoken by native speakers for communicating with native listeners), taken directly from American, British, Canadian, and other authentic sources, with good acoustical quality. The pronunciation of the native speakers needs to be representative of a mainstream variety of the aforementioned “Englishes.”

Texts will be selected to include a variety of speakers, talking at a rate that is normal and realistic for the text type. Monologues, dialogues, debates and interviews are some examples of text types.

Texts may include redundancies, false starts, fillers and other features particular to spoken language. Overlapping speech should be avoided at all levels.

A small number of semi-authentic texts may be used as needed. More specifically, level one texts may represent a combination of texts derived from authentic sources, as well as a number of dialogues based on semi-scripted topics.

All texts should be self-standing and representative of the target level.

Topical Content

Content areas for all listening items are targeted to the general listener. Topics at all levels may include texts that are relevant to NATO operations. Lower level content areas will include everyday survival and work-related topics. At higher levels, content areas may include:

- military and security issues
- economic and political matters
- scientific and technical issues
- cultural and social issues.

Tasks and Accuracy

Listening tasks and accuracy requirements are in accordance with NATO STANAG 6001.

Test Administration

- a) Test instructions will be in English.
- b) There will be a sample item to familiarize examinees with the test format.
- c) No aids will be allowed (such as note-taking, reference materials, dictionaries and electronic sources)

Validation

The initial item validation work and setting of cut scores will be accomplished using a modified Angoff rating procedure. The performance of the test items and the cut scores will be verified through collection of data in subsequent test administrations.

These specifications may be subject to modifications by the BAT Working Group during the test development process.