SHARED ITEM BANK WORKING GROUP

...a(ny) team member

- History
- Mission
- Countries
- Initiatives
- Priorities
- Way ahead

HISTORY

- Stating with 2012, in the wrap-up sessions of the ALTS candidates constantly asked for "relevant samples /level to use " and " learning about other nations best test development practices"

- Way ahead for STANAG 6001 testing community as expressed in Tours,France,2019

- Initiative took off in February 2020

MISSION

- TO STRENGTHEN LINGUISTIC INTEROPERABILITY

- To help testing organization in the test development process

- To help testing organization in an attempt of making a case in front of the decision makers

Countries

- 1. Canada
- 2. Croatia
- 3. Estonia
- 4. Latvia
- 5. Romania
- 6. Slovakia
- 7. Slovenia

INITIATIVES...

➢ Establish the ability to start the item banking activity . (start with Reading since it has not been addressed so far in any of the previous working groups .)

➢ Testing services form different countries should join this initiative voluntarily (the more the better outcome in terms of standardization among nations).

➢ All work can/should be done via Internet. We can create an email address that'll administered by one (or more) proctors.

➢ Based on the number of volunteer- countries, a number of items/level should be designated for each of the countries. (I am thinking levels 1,2 and 3 all together)

➢ Nations should send the moderated items to a proctor who will arrange the items for the piloting.

➢ Items should be piloted in a number of countries (the more, the merrier; a number of 10 candidates/country ? to end up with a sufficient number of candidates to run statistics).

➢ A proctor to gather the info after the piloting and with the help of somebody (more) friendly with the stats, should run the stats.

➢ Item revision.

➢ Piloting and stats.

➢ Final revision and setting up an item bank for Reading with access …. OR putting together a test booklet with access for the countries interested.

➢ If successful, another ability should follow (probably Listening)

▶ A sample test booklet to be posted on the BILC website for those interested in taking a STANAG 6001 test. One observation made by a WG member was that the sample test booklet format might not match the national STANAG 6001 test formats, and that this might cause some problems for the testing teams in those countries.

▶ An item bank with restricted access and available only to national STANAG 6001 test developers. The items in the item bank could be used as anchor items in national STANAG 6001 tests. Of course this means that decisions will have to be made about who is responsible for monitoring the item bank, as well as the terms for using the items in the national tests.
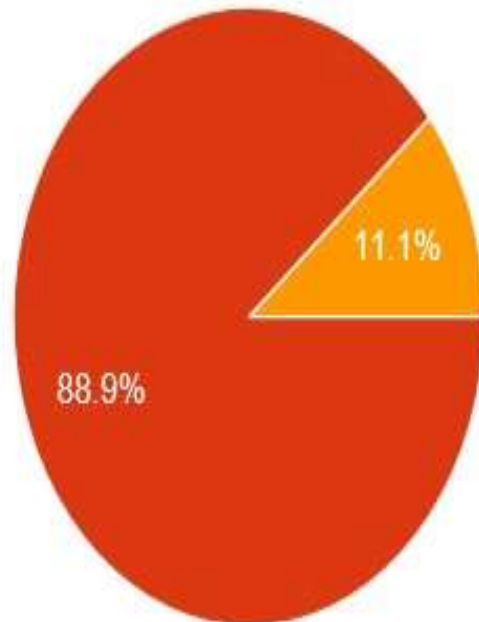
More Questions

▶ Should this project be open to all BILC-member nations or just a smaller group, at least initially?

▶ How many reading items should each nation contribute?

▶ How will the multinational moderation be accomplished?

Report on Questionnaire

What do you see as the primary final product of a shared item bank project?

9 responses

- A sample test booklet for the BILC website for those interested in taking a STANAG 6001 test.
- A restricted access bank of items only for national STANAG 6001 test developers.
- Other (if you select this option, please comment/elaborate below)

11.1%

88.9%

Comments on final product of project

5 responses

if the number of items (after moderation and piloting) allows, a combination of both would be desirable. A test booklet for those interested (no matter the format one country uses) could discourage sites that pretend to post "genuine" stanag tests. On the other hand, items only for national test developers could be used for anchoring with an emphasis on "could" since there isn't really a mean to control if any(one) particular country is actually doing it.
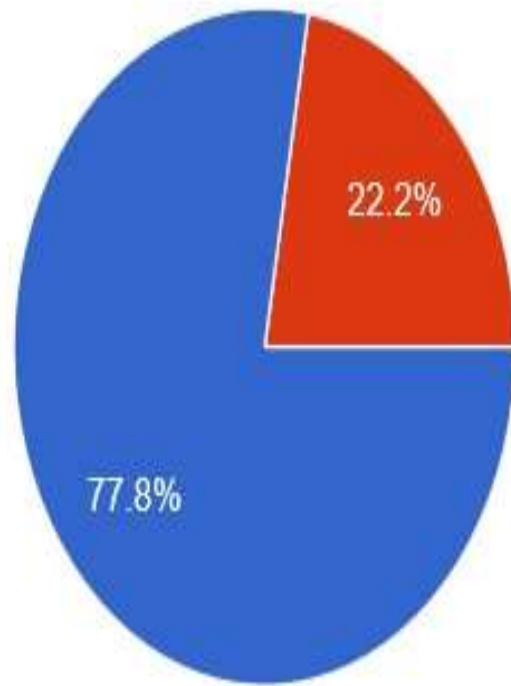
Anchor items to be used national tests.

If candidates take STANAG tests in other countries, then we have to be sure that our tests do not have the same items from the item bank.

Only participating nations should have access to the developed and approved items.

When it comes to the bank of items, we suggest a restricted access under certain terms and conditions (written document / agreement / set of rules or regulations) with the main focus on the item bank security on national as well as international level, i.e. in order to eliminate the risk of the item misuse, we suggest appointing one person in charge (a proctor) per each participating country.

Comments on participation in project

6 responses

countries that express interest

I think it initially might be a good idea to invite nations whose BAT2 results aligned well with results on their national test to develop and contribute items to the project.

All nations must follow agreed procedures. Estonian testers are willing to take responsibility for managing some stages of the project.

I think that those who want to benefit from the shared item bank should also contribute items.

A deadline should be negotiated for all nations to confirm or reject participation.

Although all the testing organizations should be invited to participate,
active contribution of involved countries should be prerequisite for the access to the item bank.
We agree with Corina s suggestion No. 6 – items should be piloted beforehand in a number of countries with a certain number of candidates to be able to run statistics.
Prior to piloting and pretesting of the items, documents setting all the procedures and conditions of piloting / pretesting need to be designed, so that each item undergoes the same procedure before its use.

How many items should each participating nation produce to submit for moderation?

9 responses
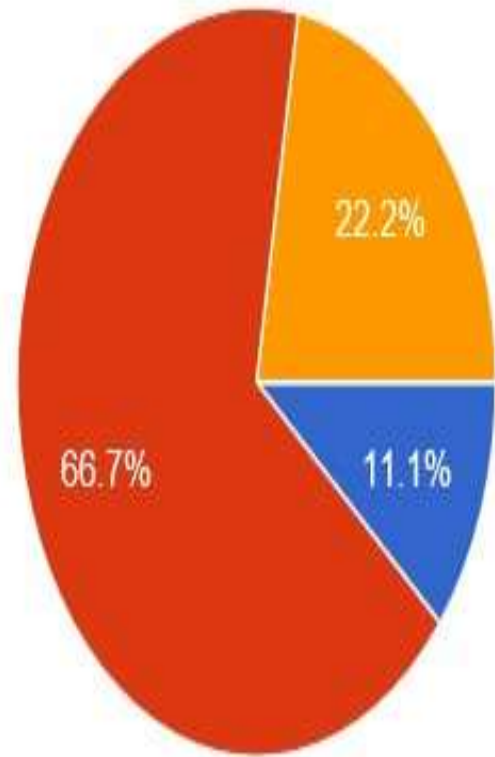
- 🔵 1 each at levels 1, 2, and 3
- 🔴 2 each at levels 1, 2, and 3
- 🟠 A total of 3 or 4 at any level
- 🟢 Other (please comment/elaborate below.)

Pie chart segments: 66.7%, 22.2%, 11.1%

Comment on number of items to contribute

4 responses

If a nation does not test at L1, they should they should focus on Levels 2 and 3 that they are more familiar with.

I think this is a good place to start, but it will also depend on how many nations participate.

Because of the computer delivered testing, the Bulgarian team will develop only MCQ items.

We understand that not all the countries test SLP 1 anymore and thus they may not see the need of having SLP1 items in the bank. However (sadly enough), in our testing context the majority of the candidates are SLP1. Maybe the willigness of the countries who do not test at SLP1 should be taken into consideration as they could be open to the idea to compensate for level 1 items with those at levels SLP2 or SLP3...

Feedback on the reading item specifications. How many responses should there be?

9 responses
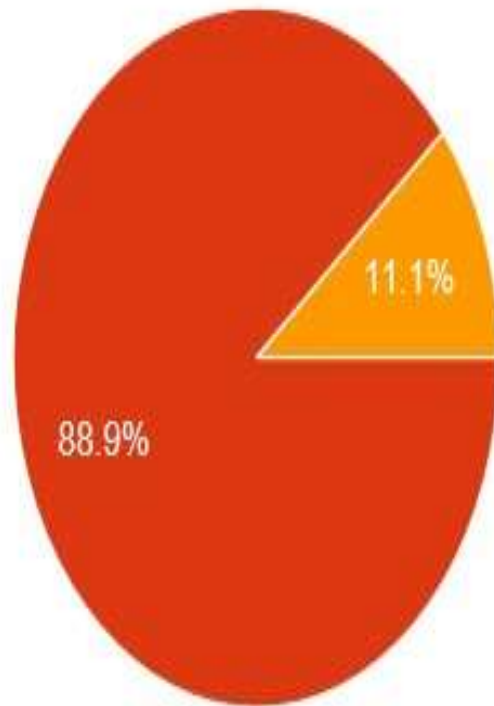
- 3 (including key)
- 4 (including key)
- Other (please comment/elaborate below)

11.1%

88.9%

Comment on number of distractors

4 responses

I think it could be either 3 or 4, depending on the quality of the distractors. If one distractor immediately can be ruled out, the test taker will in reality be left with three to choose between anyway.

I say 4 options only because that is what we are used to. If there is a consensus on having only 3 options, then we will follow that.

The version, offering 3 responses (including key), is irrelevant to the format used in Bulgaria.

There should be 4 options (including the key).
(Though we understand the reasons of the countries using 3 options only).

Feedback on the reading item specifications. Which metadata table should become part of the specifications?

9 responses
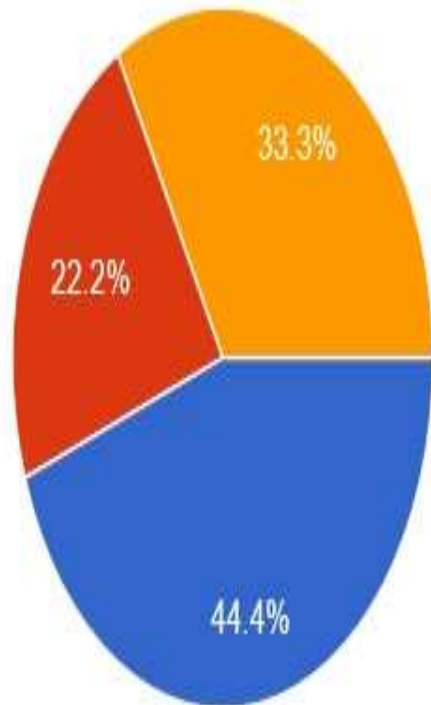
- Canada's
- Croatia's
- Other (please comment below)

44.4%
22.2%
33.3%

Comment on metadata in specs.

6 responses

a moderated combination of both

We could create a table that could be a combination of the two mentioned models. The Croatian model was offered to provoke discussion.

If necessary, we can still add to it.

Though we contributed our item spec form, we found that Croatia's was more user-friendly.

We suggest adding Title and Topic area to the Canadian metadata table.

We would like to use the combination of both. Please, find the attachment with our proposal of the metadata table in the email that follows.

Should there be a specific format for the submitted items?

9 responses
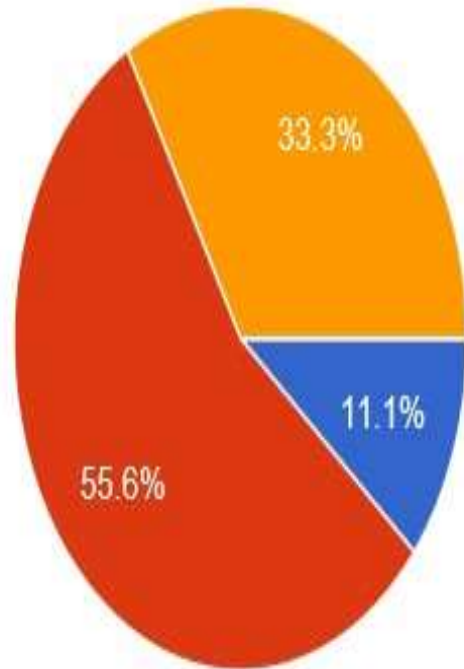
- Yes. The format should be outlined in the specifications for clarity and to limit extra documents. I/my nation can writ...
- Yes. The format should be illustrated in a sample that is attached to the specifications. I/my nation can prepar...
- Yes, formatting guidance should be included somehow, but I/my nation ca...
- No - standard formatting is not important.

Comments on formatting of items.

4 responses

I think guidance on the formatting of items will be crucial, especially if all NATO member and partner nations are to be invited to contribute items.

It will be easier to moderate if all items are formatted the same way. I don't mind sending a sample of how we format our items and if others do the same, then you can decide which formatting you prefer. I do believe, though, that they should all be the same.

The negotiated formatting reqirements should be obligatory for all participating nations.

We are open to any good practices and/or ideas the other countries may use/have..

Should the project include compiling a list of acceptable sources to get texts from?

9 responses
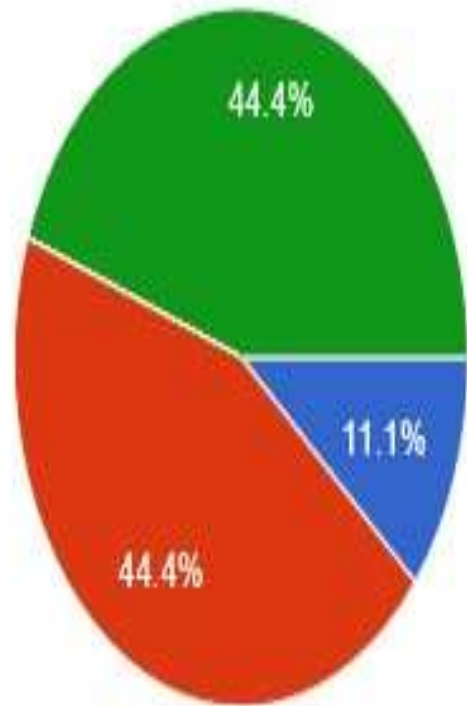
- Yes, and I/my nation would be willing to compile and maintain the list.
- Yes, but I/my nation cannot compile and maintain the list.
- No
- Other (please comment/elaborate below.)

44.4%

11.1%

44.4%

Comments on compiling a list of acceptable sources

7 responses

The list of sources might be accompanied with general guidelines on what kind of sources to use or not to use in terms of authenticity (e.g. websites for educational purposes, contrived material, materials for non-natives, etc.)

Perhaps a list of "suggested sources", rather than "acceptable" ones? What if a testing team comes across something brilliant but feel they cannot use it, because it was not on the list?

This should be work in progress, information to be added to the metadata.

I think that if we share some acceptable sources it will be a good starting point for some who aren't too sure where to go.

We believe compiling of a list is not necessary but we can contribute to this effort.

Sources should comply with BILC descriptors and subject areas, however, copyright restrictions should be observed.

In order to be able to fully answer this question, what exactly do you mean by acceptable sources, will the list be edited regularly or how limited or flexible will we be?

Thank you for your expertise, consideration and time. If there is anything else you would like to add, please do so below before you submit your answers.

3 responses

Should the item bank become a resource for various nations' test developers, it is important to stress that even though the items have been moderated, pre-tested, piloted, tried, and validated, nations will still have to do a main trial of their new test, every time they put different items together in different constellations. According to our Lancaster tutors, items can behave differently and contribute to test reliability (influence Cronbach's Alpha) to different degrees - depending on the other items that surround them. For instance, an item you answered on page 2 might make answering another item on page 4 easier - even if the two items may not be directly related to each other.

I think that this is a very interesting project and am very happy to be a part of it.

We appreciate all the effort of everyone involved so far, and we are eager to participate in this great project
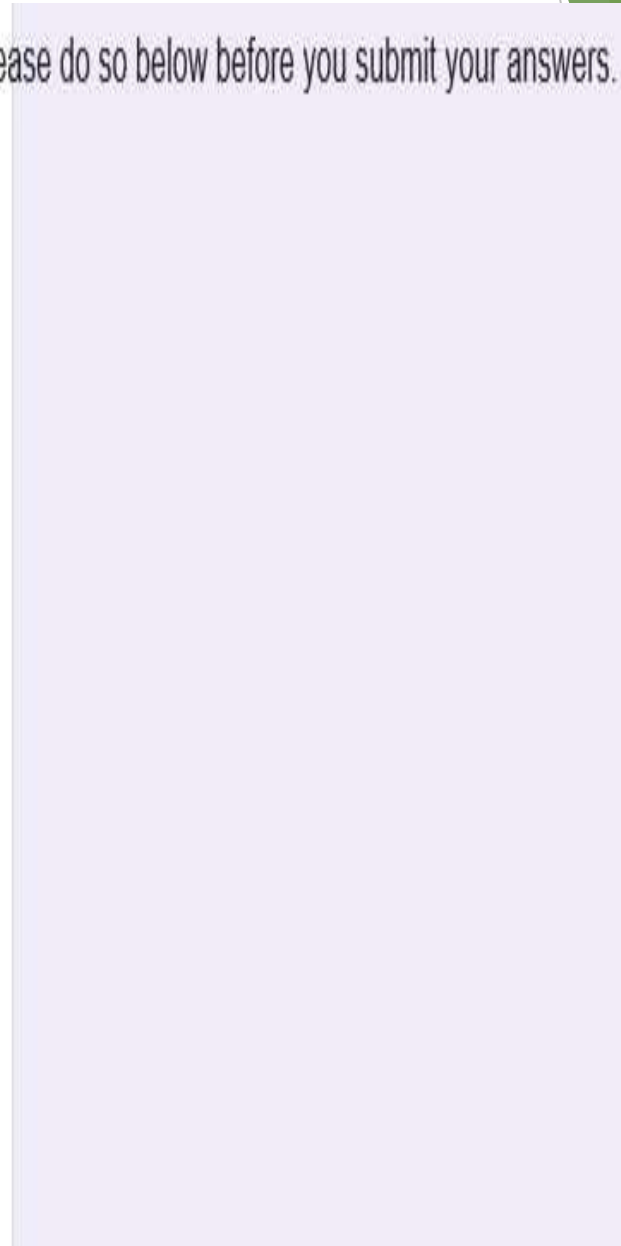
So far……

- Drafting specifications (Estonia & Latvia- January)

	STANAG 6001 Level 1	STANAG 6001 Level 2	STANAG 6001 Level 3
Input text	Texts from quality sources within the target culture, avoiding culturally sensitive material. (1) Texts should represent a balance of the varieties of English used in NATO context.		
	Authentic and semi-authentic	Authentic	Authentic
Text type & content	Very simple announcements, ads, personal notes; highly predictable descriptions of people, places and things.	News reports, magazine articles, short stories, human interest features, instructional descriptive materials, routine workplace correspondence.	Multiple-paragraph prose on a variety of professional or abstract subjects such as found in editorials, formal papers and professional writing.
Author purpose	To orient	To instruct	To evaluate
Text length	Up to 60 words	Up to 150 words	Up to 300 words
Number of questions per text	1 question per text	1 or 2 questions per text	1 or 2 questions per text depending on the length of text
Text organisation and structure	Loosely organized simple sentences.	Connected discourse within and across paragraphs, compound and some complex sentences.	Well organized cohesive discourse, complex structures.
Lexical properties(2)	Concrete simple vocabulary	Concrete vocabulary encountered in work and daily life situations	Abstract and concrete vocabulary
Editing	Ideally very little or no editing is required. If editing is necessary, then follow the guidelines in Herzog (2013).		
Tasks/subskills tested	Understand the basic meaning; find some specific details.	Locate and understand the main ideas and details.	Understand hypothesis, supported opinion, argumentation. Understand implicit information. Recognize humor, emotional overtones, and subtleties.
	Ensure representative sampling of tasks/subskills tested.		
MCQ item (3)	For item development and review, follow BILC Item Review Checklist. Language of the orientation, stem and options is English.		
Orientation	Each item begins with orientation: the context presented to provide a fuller understanding of the situation.		
Stem	Either a partial sentence to be completed or a question.		
Options	3 or 4 options per item.		
Key	There is only one correct answer.		
Test administration	Paper and pencil or computer-delivered.		

Feedback on the reading item specifications. Do you have other feedback on the reading item specs?

9 responses

Well done!

yes. establishing the cut-off scores/level (non-compensatory). 70%? should we include/discuss the issue of scoring plus levels as part of the specs?

Not at present.

Nothing more to add. I think the Estonians did a great job. If, for whatever reason, we need to amend it as we move forward in the project, then we can.

The reading item specs are very similar to our national specs. We are satisfied.

They are clear and acceptable for the Bulgarian team.

We would suggest more detailed version of text type and content, and subskills – please find attached our proposal in the email. We also believe it is a good idea to include the statistical requirements (FV, DI) each item should meet in order to be included in the bank.

Metatable

Reading Item Metadata					

Country:		**Skill:**	Reading	**Level:**	SLP1
Item Name/Code:		**Retrieved from (Source):**			
		Retrieved on (Date):			
Item Type:	MCQ	**Sub-skill Tested** (see item specs):			

FV:		**DI:**	

No. of Words:		**Input:**	authentic	**Register:**	informal
Editing:	no	**Content:**	civilian		
Text Type (see item specs):		**Topic** (see item specs):			

ITEM: