Best Practices
in
Virtual Multinational Norming Sessions
for Testing Speaking

Introduction

Norming is essential for the interoperability of nations. When members of different national testing teams are normed, candidates get the same rating no matter where they are tested or who tests them. It's important to have norming sessions on a regular basis to avoid drift. Drift occurs when the interpretation and the application of the standard changes over time to become stricter or more lenient.

Norming usually happens in face-to-face settings. *Virtual norming,* however, is a practice where members from several different nations meet **online** to discuss speaking tests for the purpose of norming in a multi-national setting. In the case of the BILC testing community, the objective of a norming session is to ensure that all testers interpret and apply the STANAG 6001 level descriptors in the same way. A standardized interpretation and application of the rating criteria leads to increased reliability and validity.

Before you begin planning your virtual norming session, make sure that you and your organization have a platform and the network capability to allow the online connection of participants from multiple countries. The network must also have sufficient bandwidth to support audiovisual capability.

This is important because participants need to be able to hear the audio files of recorded speaking tests easily and clearly. Constant breaks in connection or lag in sound or poor sound quality can affect the rating given by participants. It is also key for participants to be able to see and hear each other easily as this makes it easier to build a sense of comfort and community during the virtual norming session. This facilitates open, productive discussions among participants. Keep the group small (around 8-10 people) to ensure maximum connectivity and participation.

Virtual Norming Challenges

➢ There can be problems with technology that lead to interruptions, a participant's failure to connect, lagging, poor sound quality, etc.

➢ Turn taking can be more difficult virtually. People sometimes speak over each other, which then cuts off both speakers' voices.

➢ It is more difficult for facilitators to create an environment where people feel comfortable expressing their opinions because participants do not have the opportunity to make interpersonal connections among themselves (for example, no coffee breaks so no opportunities to chat and to get to know one another). Therefore, it is important for facilitators to set rules that foster collaborative and respectful discussions.

➢ Many people find that focusing online can be fatiguing. Taking this into account, facilitators may have to schedule more breaks than they would if participants were meeting in person.

➢ Scheduling may also become an issue when there is a multinational group because of different time zones.

****

Three Stages

There are special considerations and steps to take to prepare for a virtual norming session, conduct a session, and complete the virtual norming.

Before the session	During the session	After the session

Facilitators must:

- ✓ confirm that participants have access to the chosen platform and an appropriate connection with adequate bandwidth.
- ✓ make sure all participants have been trained in the BILC speaking test format.
- ✓ make sure all participants have access to the STANAG 6001 level descriptors (pdf copies, a link to the BILC website, etc).
- ✓ send the rules for the online session (see Ground Rules for Participants in a Virtual Norming Session, on p.4).
- ✓ give participants access to the chosen speaking tests a week or two in advance so that they can listen to and rate them before the session begins. Participants must be ready to justify their ratings during the norming session.
- ✓ set a deadline for participants to submit their ratings a few days before the session begins.
- ✓ compile the ratings and note any discrepancies.

Preparation is key for facilitators:

- ✓ Review STANAG 6001 level descriptors. It is important for facilitators to have a firm grasp of the rating criteria (Content/Task /Accuracy statements).
- ✓ They must be ready to explain why performances were rated at a given level.
- ✓ Determine which tests will serve as benchmarks. They must be clear examples of a base/plus level based on the criteria in the STANAG 6001 level descriptors. These can be chosen from your own team's previous speaking tests, if they are done in accordance with the BILC format or from benchmark BILC training interviews.
  - o Select audio files that have high sound quality and are easy to listen to (e.g. no background noise, no mumbling, no interrupted connection, etc).
  - o Make sure the structure of each interview corresponds to the final rating in that the floor and ceiling are clearly established. This makes the test easy to rate.
  - o Choose several base level tests (at least two examples per base level).
  - o Choose several plus level tests (at least two examples per plus level).
  - o Ensure that variations within a level are represented in the chosen test samples.

- ✓ Facilitators may choose several borderline interviews. Select tests that generated a lot of discussion in prior rating sessions. Raters may have disagreed on the final ratings.

Before the session

During the session

- ✓ Together as a group, review the STANAG 6001 level descriptors and the BILC speaking test format. Make sure participants understand the importance of establishing a clear floor and ceiling.

- ✓ Begin the session by listening to base level benchmark tests. Facilitators may choose to begin the session with an interview that had 100% agreement, i.e., all participants are on the same page.

- ✓ Facilitators can use two different approaches:
  - o Listen to the test completely without stopping. Participants then discuss the performance and structure of the interview. This approach may be tiring for some participants.
  - o Listen to the warm up and stop the recording. Discuss the performance and ask participants at which level they would begin the body of the interview. Continue to listen and stop the recording after each task to discuss performance and interview structure.

- ✓ If there are any discrepancies in the rating of the chosen interviews, facilitators must explain why the performance is at a certain level, based on the level descriptors, so that all participants have a common understanding of how to interpret the level descriptors.

- ✓ Repeat the same procedure for each base level test before going on to plus levels.

- ✓ It is important to discuss the ratings without pinpointing anyone in particular.

- ✓ Once participants have a clear idea of base and plus level performances, borderline tests can be introduced.

- ✓ Facilitators should call on participants by name and acknowledge each person's comments.

- ✓ Facilitators should keep the discussion moving at an acceptable pace.

- ✓ If discussions get intense, facilitators should be ready to intervene and guide participants back to a more productive discussion.

- ✓ The aim of a norming session is to reach consensus on the ratings after discussion, which is not always easy to achieve. If participants disagree on the ratings of multiple interviews, more frequent norming sessions may be necessary.

After the session

✓ Allow participants the chance to give feedback on certain aspects of the session such as

      o   Were the objectives clear?

      o   Were the norming activities helpful?

      o   Were the samples appropriate/relevant?

      o   How could the session be improved?

****

Ground Rules for Participants in a virtual norming session:

1. Keep microphones off unless you are speaking.
2. Keep cameras on to help create a sense of participation and group cohesion.
3. Raise your hand when you have something to say and wait to be called on by the facilitators.
4. Be respectful of others' opinions.
5. Do not monopolize speaking time.
6. The environment should be quiet, free from background noise.

Remember:

A norming session is not about winning or losing. Rather, it is a chance for participants to objectively discuss a candidate's performance and judge that performance against clearly stated criteria. Use this discussion as a means to identify and internalize which aspects of the performance warrant the rating in order to achieve a common interpretation of the STANAG 6001 Language Proficiency Levels.