Bureau for International Language Co-ordination

Self - Assessment Checklist
for
STANAG 6001 Testing Programs

1. Is there a formal testing policy based on a needs analysis?

- Was a needs analysis conducted to determine the approximate number of personnel to be tested and their anticipated or required STANAG 6001 testing levels?
- Does the testing policy include a mission statement with:
  - the purpose of testing (i.e. to select personnel for training or job placement, to inventory skills of personnel for a database, etc.)
  - the standard to be used (STANAG 6001)
  - approximate numbers of people to be tested, frequency of testing, and when and where testing should take place?
- Has the policy been approved by the MoD or another higher authority?

2. What is the structure of the testing organization and how is it funded?

- Is it independent from the management of an instructional program?
- Is it a national, centralized testing organization?
- Is there a specific budget for the test development and administration?
- Is the testing organization given adequate resources?
  - List resources available (number of computers, Internet, software, printers, reference books, copiers, recording equipment, etc.)

3. How is the testing program staffed?

- List number of employees (indicate if full time or part time) and their duties.
- What are their job qualifications? (knowledge of the professional discipline of language testing, successful completion of BILC's Language Testing Seminar, etc.)
- Is there certification of the testing staff's own language proficiency levels?
- Does the chief of the testing department have expertise in testing?
- Is there a plan for continuing professional development, cross training, and renorming sessions for the testing personnel?

Bureau for International Language Co-ordination

4. **What are the established test development procedures?**

- Is there a clear rationale for single, bi- or multilevel tests?
- Is there a table of specifications which determines skills, levels measured, format, number of items, timing, and weighting, if any?
- Is the test construct clear to everyone involved? (STANAG 6001 test development requires that everyone understand the content, tasks, and accuracy statements underlying this scale.)
- Is the development process iterative? (write, review, revise)
- Are authentic texts used in the listening and reading tests?
- Are the texts kept for future use?
- Are the texts and tasks appropriate for the level being tested?
- Does the level of the each task match the level of the text?
- Is the number of tasks/items per text limited to no more than 3?
- Are the instructions for examinees clear and concise?
- Is there a native speaker who can review items for authenticity (validate) of language usage?
- How is the item pool maintained?
- What is the schedule for reusing items?
- What is the schedule for updating the item pool (adding or deleting)?
- Are procedures in place to ensure test security during and after item development?

5. **How is the test validated?**

- Is there a method established to relate the test results to the standards (content validity)?
- Is there a method of setting cut-off scores for each level?
- Is there a method in place for evaluating test reliability?
- Is there a method in place for evaluating the effectiveness of the test as a whole as well as the individual items?

6. **What are the established written test administration procedures?**

- Is there a familiarization guide for the test takers?
- Is there a familiarization guide for all test stakeholders (teachers, administrators, score users)?
- Is a formal schedule published with the time, date, and location of the testing sessions?

- Are the proctors given an official list of approved names prior to the testing session?
- Are the proctors committed to test security?
- Is there a clear policy to prevent and penalize cheating of any kind?
- Do proctors have clear instructions and carefully follow procedures?

7. **How are the tests of productive skills (speaking and writing) administered and rated?**

- How are the raters selected? (Ideally, raters of speaking and writing tests should not be part of the instructional staff. As a minimum, teachers should not test their own students.)
- What training /refresher training is provided?
- To avoid tester fatigue, are limits placed on the number of speaking tests raters conduct each day?
- Do speaking tests elicit samples of unrehearsed proficiency?
- Are the speaking tests recorded?
- What is the policy as to whether ratings are done independently or in conference?
- Is there a quality control system for productive skills testing? (norming, calibrating, or periodically re-scoring tests)

8. **How are the STANAG 6001 test results reported?**

- Is there a formal system in place for reporting and recording scores?
- What are the procedures for inputting scores into a database?
- How are scores reported to the individuals?
- Are the scores included in personnel records?

9. **How are the testing materials and reports archived?**

- Is at least one copy of all test material kept indefinitely?
- Is there a policy for maintaining examinee records, for releasing them, and for declaring them inactive?
- Is all data related to tests taken documented in the examinees' record (test forms, dates, rater, etc.)?

10. **What are the testing conditions?**

- Is test administration fair? (no distinction made between categories ranks of examinees)

- Is test administration controlled? (same procedures every time, every place)
- Is the room where the test is administered pleasant, quiet, well-lit, well-ventilated, and is the room temperature comfortable?
- Do the examinees sit far enough apart to prevent distraction or cheating?

11. Miscellaneous administrative issues

- Is there a policy on length of time scores can be considered valid? (higher proficiency scores level 3 and above, are usually sustained longer than lower scores)
- How many alternate, equivalent forms of the test are there? (two or more forms of each reading and listening test are needed to prevent test compromise and to facilitate re-testing when needed; a validation process must ensure the forms are equivalent)
- What is the policy for recognizing certificates issued by other organizations or testing centers?
- Is there an established system for dealing with grievances or requests for retesting?
- What is the procedure for dealing with compromised tests?
- How frequently may individuals test?
- Is there a plan for dealing with emergencies or unexpected situations which may occur during a testing session, such as a power shortage, a fire alarm, etc.?