

## **BENCHMARK ADVISORY TEST (BAT)**

### **History**

The idea to create a Benchmark Advisory Test (BAT) was born in 2003 at the BILC Conference in Harrogate, UK after a delegate indicated that the NATO STANAG 6001 interpretations seemed to vary significantly, and that it appeared that PfP nations were expected to adhere to the standard more stringently than some older NATO nations.

During the ensuing discussion, the need to develop a common test that would be of advisory nature was confirmed by a subsequent survey conducted among nations which revealed that most countries were interested in having a test available against which they could compare the results attained on their national tests. The test results would be of advisory nature **only** and would not replace the national score results. Based on the survey results and the interest expressed by the nations, BILC began (in 2005) to develop test specifications and test items by requesting that BILC member nations donate test reading and listening items for the test item pool (voluntary national contributions), and by tasking the BILC WG on testing and assessment, that had been formed in the meantime, to review these items for acceptability.

At first, the project was very limited in scope: the BILC WG worked together to create tests in listening and reading against which nations could calibrate their national tests. This initial small team of professionals collaborated through e-mail exchanges and met once or twice a year at the annual BILC conference or seminar to discuss progress and validate the test items. Eventually the team expanded to 13 individuals from 8 nations.

To accelerate the development process and to expand the number of skills tested, in December 2006, NATO ACT awarded a contract to the American Council for the Teaching of Foreign Languages (ACTFL). Under the contract, in addition to administering the BAT reading and writing tests, ACTFL was to develop and administer tests in the speaking and writing skills. The contract also stipulated that 200 BAT allocations would be distributed among nations.

The BAT test development process concluded in December 2008 and raters/testers were trained in the administration and scoring of the speaking and writing tests. The benchmarking process commenced in the late spring of 2009.

Eleven nations participated in the project. The tests in the listening, reading and writing skills were administered via the Internet, while the speaking test was administered via the telephone.

The benchmarking process was expected to be finalized by the end of October 2009; however, due to some delays, it was completed in January 2010.

## **BENCHMARKING RESULTS**

Even though the sample obtained may be considered to be statistically insignificant due to the limited number of candidates who took the test in each nation, the results provided an indication that there was a lack of alignment between the BAT and national test results. Although some nations were aligned in some skills, in an overwhelming number of cases, the results shown on the national tests were higher than those achieved on the BAT.

This discrepancy could be ascribed to a variety of factors, such as: test purpose, testing method, alignment of author purpose, text type, and reader/listener task, inadequate tester/rater norming (productive skills), inconsistencies in the interpretation of STANAG 6001, cut-off score setting, etc.

## **CONCLUSION**

The BAT project was the result of concerted efforts invested by many individuals and organizations. BILC considers this project a successful one and a great example of the positive outcomes produced through collaboration among BILC members, NATO authorities (ACT) and a service provider (ACTFL). The BAT project is another significant step toward language testing standardization and enhancement of interoperability.