

ILR TESTING COMMITTEE MINUTES

1. Purpose: to provide an account of the ILR Testing Committee meeting held 25 January 2013, at the Foreign Service Institute (FSI) at the George P. Shultz National Foreign Affairs Training Center, Arlington, VA.
2. Minutes:
 - a. Christina Hoffman started the meeting by introducing the presenters and the topic – testing writing. As a side note, she asked the attendees how long the scores were valid in their agencies. Patrick Wright, ILI representative, informed that the score of 5/5/5 was valid for 6 years in his agency. If somebody scores 5/5/5 twice in a row, they do not need to test ever again. Cynthia Taskesen from DLNSEO commented that there was no policy regarding speaking tests but there was policy regarding DLPT. Maria Manfre from FBI, said that there were many factors involved such as the purpose of the test, that determined the longevity of the scores' validity. Elvira Swender, from ACTFL, informed that their speaking test score was valid for two years. The agencies representatives agreed to send the detailed information on this to Christina in writing.
 - b. Elvira Swender, from ACTFL and Bill Hindle from LTI, delivered a presentation on *ACTFL Writing Proficiency Tests*. The ACTFL presentation:
 - provided a historic overview of the development of the ACTFL writing tests starting with the Academic Writing Test for College Students in 1989 (developed by Ray Clifford and Pardee Lowe) and finishing with the ACTFL “iWPT” in 2012 and AAPPL Presentational Writing Component in 2013. It was noted the results of the 4-skill NATO Benchmark Advisory Test (BAT) scores revealed that writing was surprisingly the weakest skill and that writing scores were consistently equal to or lower than reading scores. In a 2007 study, when writing was at the 1+ level, reading was better for 60% of the test takers; when the writing level was level 2, 75% had better reading skills; and finally, when test takers were rated level 3 in writing, 100% of them scored level 3 in reading; This prompted the comment that writing scores could serve as a screener for predicting minimal reading levels;
 - mentioned that the ACTFL WPT and BWT can be rated using ILR Skill Level Descriptions
 - provided a detailed description of ACTFL WPT which was described as a test of writing proficiency (real-world context and content with a variety of writing tasks) that measured how well a person could write spontaneously without resources and with limited time for revision. Major level and sublevel are assigned;
 - presented the writing assessment criteria for each level;
 - outlined the four requests for a variety of the writing tasks: practical, social and professional needs (e.g., descriptive, informative, narrative and persuasive writing);
 - described the prompts and test instructions. All of them are written in English with the context, intended reader, tasks, content, and suggested response length and time specified;

- presented descriptions and examples of the writing skills at ILR levels 1, 2 and 3;
 - outlined three delivery options for WPT: booklets, internet-delivered fixed form, and iWPT which was customized to the individual test taker. It was noted that the test was predominately delivered through Internet;
 - demonstrated the major features of iWPT such as technology system check, background survey, self assessment, and keyboard options to support different scripts;
 - listed the languages that could be tested through WPT and additional options for the tests such as the option to customize to lower or higher levels, or to target the specific contexts and content area.
- Bill Hindle noted that the main clients for the ACTFLWPT were universities, commercial entities. A new user is the American Translators Association (ATA) which uses the test to screen for the ATA certification test. Bill added that given the expense involved in developing reading tests, the possibility of using a score on a writing test to screen for a minimal level of reading ability may be a sensible predictor. James Dirgin commented that the approach described had been used in translation for a long time: writing was the minimal score screening test. Deborah Kennedy posed a possible research project that might show that if the writing skills improved, so would the listening skills.
- Elvira Swender in reply to the question about the rating procedure, informed the group that the writing sample was rated based on the ACTFL Guidelines and the assessment criteria - that no special, prompt-specific notes were provided for raters. The raters are trained to look at the functional criteria first, then at the accuracy by level or ACTFL sublevel. Two raters rate independently, and in a case of a disagreement, the sample is arbitrated. James Dirgin commented that in the absence of the protocol, the assessors needed to know the scale very well. He added that one of the problems in training the assessors would be what mistakes were acceptable. Elvira clarified that only trained and certified OPI testers rate ACTFL writing tests and are skilled in recognizing the level-specific accuracy expectations. a
- c. Deborah Kennedy (*Center for Applied Linguistics (CAL)*) delivered a presentation on *Assessing English Language Writing Proficiency* for the English for Heritage Language Speakers Program. The presentation
- described the program fundamentals: its origin, purpose, nature and program partners. The program's goal is to recruit the native speakers of critical languages who enter the program at ILR level 2 in English and should exit the program with ILR level 3 in English. It is the scholarship program that is coordinated and funded through the DoD National Security Education Program, managed by the Center for Applied Linguistics and delivered by Georgetown University;
 - mentioned that based on 2005 feasibility study that surveyed a variety of agencies within DoD, the Department of State and the Intelligence Community about English language needs, writing received the lowest priority among the four modalities. She added that the results were most likely the reflection of the fact that the people were not asked to write because their writing was poor;

- informed that the early writing assessment for the program (2006 – 2008), was based on the writing assessment from DLI-ELC and consisted of one essay question scored with a rubric. A and B versions were created for entry and exit testing. The writing test rating factors included global tasks and functions, lexical control, structural control, sociolinguistic competence/style, orthography and text produced. This approach had a few major challenges, namely it was hard to assess writing based on only one writing sample, assign the rating that reflected the increased proficiency within a short period of time (6 months) using a rubric that did not reflect subdivisions within the levels and to relate the orientation of assessment (proficiency) to orientation of instruction (performance);
 - described the current writing assessment that was based on the model for the Royal Society of Arts Examination Board (1981) for testing communicative language use. This model consists of three constructed response items (level 2, 2+ and 3), as well as two forms (entry/exit). The test items do not appear in the sequence based on their levels. One hour is given for the whole test without specifying how much time should be spent on each task. It is a paper and pencil test. There is no special protocol but the raters take extensive notes. The scoring method requires that each item and each factor is rated individually. The final rating for the test is equal to the lowest item rating, while the overall rating for each item is equal to the lowest individual factor rating. Plus and minus scores are assigned on a percentage basis;
 - concluded by emphasizing that the new assessment protocol helped with addressing two out of the three challenges described earlier in the presentation. One challenge still remains: how to measure performance orientated instruction with the test that aims at measuring proficiency.
-
- In reply to the question about the breakdown approach not being used, Deborah Kennedy commented that they almost always saw breakdown at level 3 but no breakdown below level 2.
 - Cynthia Taskesen asked how the current rating procedure would deal with the test taker who did well at level 3 but failed level 2+. The reply was that level 3 writers exhibited level 3 features such as a good socio-cultural control and lexicon at levels below. James Dirgin commented that the problem of relating the task level with the performance level could be solved by giving the test taker a choice in the tasks. He added that the America Council asked the examinee to choose two tasks out of five offered (they could choose a topic and a level).
 - Deborah Kennedy concluded the presentation by expressing the desire to externally validate the test that she presented.
- d. Don Smith (FBI) delivered a presentation on *Testing Writing Skills at FBI*. The presentation
- informed that English was the only language for which the writing skill was tested in the agency. The test is administered to linguist applicants if no translation test is available, as well as to English Monitor applicants, Rater applicants and FBI National Academy applicants (English is the lingua franca of the instruction in the Academy that is attended by the representatives of different countries);

- reported that 213 tests were administered in 2013 which corresponded to only 2% of annual testing volume;
 - described the testing format which came in two forms: one-sentence prompt and two “descriptive prelude” prompts. The common features of both forms are: one hour is allocated to complete each form and the expected outcome is a 150-word essay on a topic of a general interest. The one-sentence prompt format provides no choice for the topic while the “descriptive prelude” form allows choosing one topic out of the two presented. The “descriptive preludes” provide the context. The rating is done based on the ILR Writing Skill Level Descriptions. There is also an error notation system in place;
 - Don Smith concluded the presentation by emphasizing that FBI did not administer the writing test to show the progress in language acquisition; rather it is used to see how well someone expressed themselves in English.
3. Point of contact: Christina Hoffman, HoffmanCN@state.gov; Inna Sabia, inna.sabia.civ@mail.mil