

Practical activity: Norming Speaking Assessment

Regular norming of raters is an important aspect of rater training. BILC offers two courses for NATO and partner countries' language testers where, among other issues related to language assessment training, participants are instructed on how to interpret STANAG 6001 level descriptors when rating speaking and writing performances. The Language Testing Seminar (LTS) is intended for novice testers and the Advanced Language Testing Seminar (ALTS) caters to the more experienced.

Participants of the first hybrid BILC testing workshop were invited to take part in a pre-workshop norming activity aimed at looking into how normed the rating of speaking performances among the testers is. Each participant was asked to listen to six interviews from the BAT2 project that were provided by BILC. They then had to rate and provide comments on their ratings using the relevant level descriptors IAW STANAG 6001. The ratings and the comments were submitted via Google Forms. In total, 50 testers from 14 nations took part in this activity. The results indicate some differences in the ratings and suggest there is a need to carry out regular norming activities among the STANAG 6001 testing community.

References:

Eckes, T. (2011). *Introduction to Many-Facet Rasch Measurement*. Bern, Switzerland: Peter Lang D.

Freelon, D. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, 5(1), 20-33.

Linacre, J. M. (2020) *Facets computer program for many-facet Rasch measurement*, version 3.83.4. Beaverton, Oregon: Winsteps.com

Mizumoto, A. (2015). *Langtest (Version 1.0) [Web application]*. Retrieved from <http://langtest.jp>

Myford, C. M., & Wolfe, E. W. (2003). Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part 1. *Journal of Applied Measurement*, 4(4), 386–422.