

Summary

Scoring validity refers to “test results that are *stable over time, consistent in terms of the content sampling and free from bias*” (Weir, 2005, p. 23). Scoring validity involves test scores that are obtained through “the application of appropriate and representative marking criteria, exhibit consensual agreement in the marking, are as free as possible from measurement error, stable over time, and engender confidence as reliable decision-making indicators” (Weir, Vidaković, & Galaczi, 2013, p. 195). Although variability in rating writing and speaking is an inextricable part of the process inherent to the subjectivity of a process involving the performance of individuals, random variability can lead to casualization of marker consistency (Barrett, 1999) and have devastating consequences for candidates. Training teachers to score performances is presented in the literature (Weigle, 1994, 1998) as the solution to reduce rater error while respecting the individuality of the rater, the candidate and the task at hand.

This workshop intends to introduce the principles of rater training by following a four-step process in which participants will deal with:

1. Familiarisation with rating scales and linking a scale to the CEFR
2. Identifying and avoiding rater errors
 - a. Rating writing tasks
 - b. Rating speaking tasks
3. Using the scale and identifying descriptors in written productions from candidates
4. Benchmarking productions and Gold Standards