

## Paper session 5B | Validating teaching quality frameworks

Room: Skiftnyckel

Chair: Michael Tengberg, Karlstad University

### Observations of classroom instruction and rater error: towards better practices

**Mark White**

University of Oslo

Rater error is a serious threat to conclusions from observational research on teaching quality. This is especially true for cross-country research where raters are often nested within countries, making it difficult to use standard approaches for examining rater agreement. This paper introduces a new approach to examine rater error in cases where raters are nested within country. This approach focuses on incorporating uncertainty due to rater error into study conclusions, allowing for an examination of how robust study conclusions are to rater error. The approach creates a linking data set by “master scoring” a set of videos from each country. The linking data set is used to calibrate a hierarchical rater model that nests raters as items within observation rubric

dimension scores. The hierarchical rater model is then applied to the full data set to estimate scores that are adjusted for rater effects or the model is embedded within a larger model to examine research questions. I demonstrate an application of this approach to the LISA Nordic data set, which collected videos of 390 lessons across the five Nordic countries. This application shows the importance of accounting for rater error.