Appendix C

Calculation of IRR in CAT

The calculation of ICCs (and their variants) assume that ratings from multiple observers for a set of targets (i.e., observation intervals) are composed of a true score component and measurement error component. This can be rewritten from equation 1 in the form

$$X_{ij} = \mu + r_i + c_j + rc_{ij} + e_{ij}$$

where X_{ij} is the frequency of a behavior code for a time interval i provided by observer j, μ is the mean of the true score for this variable X, r_i is the deviation of the true score from the mean, c_j estiamtes any systematic deviation of observer j, rc_{ij} represents the interaction between observer and deviation for each interval, and e_{ij} is the measurement error. Hence, the session-based ICCs for each code is calculated based on a “time interval” x “observer” matrix. The observation interval length is determined by the researcher.

To illustrate: A researcher wants to know the inter-rater reliability for a code A based on 5-minute resolution for a session with a 50 min. length. The matrix for the ICC calculation is based on ten repeated observations (10 rows (intervals) x 2 observer matrix). If observer 1 has coded 3 instances of behavior A during the first 5 min interval and observer 2 has coded behaviour A 5 times (during the same interval), the observed frequencies for A for the first interval are 3 and 5 for observer 1 and 2, respectively.

CAT provides single-measure ICCs (i.e., ICC(A,1) and ICC(C,1)). The single-measure ICCs are more conservative reliability estimates as they tell the researcher if the coded behaviours from a single observers can be used for further analyses (the average-measure ICC(k) tend to be higher than single-measures ICC(1), Hallgren et al., 2012).