

Inter-Rater Agreement

Cohen's κ

Two raters

Fleiss' κ

Multiple raters

ICC

Continuous scores

Krippendorff's

α

Any data type

Kappa Interpretation Scale



$\kappa < 0.20$

Poor agreement

Unacceptable



$\kappa = 0.21-0.40$

Fair agreement

Needs work



$\kappa = 0.41-0.60$

Moderate agreement

Acceptable



$\kappa = 0.61-0.80$

Substantial agreement

Good



$\kappa > 0.80$

Almost perfect

Excellent

Disagreement Detected



Consensus Discussion



Third Expert

Disagreement resolution: Consensus discussion or third expert adjudication