

A COMPARATIVE ANALYSIS OF INTER-RATER RELIABILITY: ANALYTIC VS HOLISTIC WRITING ASSESSMENT METHODS

MIF. Kareema and USF. Jahan

Department of English Language Teaching, South Eastern University of Sri Lanka

Correspondence: mifkareema@seu.ac.lk

Abstract

This study explores the reliability and effectiveness of holistic and analytic rubrics in ESL writing assessment, focusing on how rater experience of novice and experienced raters' scoring system influences scoring consistency. Using quantitative methods and analyses like Intraclass Correlation Coefficient (ICC) and Mixed-effects model, the study compares inter-rater reliability across both rubric types, highlighting their strengths and limitations. Results show that analytic rubrics offer greater scoring stability due to their criterion-specific structure. Novice raters exhibited significant inconsistencies when using holistic rubrics. Experienced raters achieved higher reliability across both rubrics, showing systematic accuracy (ICC 0.72 to 0.80). These raters preferred analytic rubrics due to its structured approach. The study recommends the use of analytic rubrics for novice raters. It also emphasizes the need for training and calibration to improve novice raters' consistency. Despite limitations such as a small sample size, and convenience sampling, the study provides valuable insights for ESL learners, teachers, policymakers, and evaluators. Selecting the appropriate rubric based on rater experience, combined with proper training, can improve the fairness and reliability of ESL writing assessments, ensuring more consistent and accurate evaluations.

Keywords: inter-rater reliability, holistic rubric, analytic rubric, Intraclass Correlation Coefficient (ICC), mixed-effects models, Sri Lankan education