

Score Resolution in Essay Grading:

A View from a Signal Detection Model of Rater Behavior

Lawrence T. DeCarlo and YoungKoung Kim

Teachers College, Columbia University

Abstract

Each essay obtained in a large scale assessment is typically scored by two raters. If the scores differ by more than 1, a third rater, or adjudicator, is sometimes used to resolve the difference. There are many questions, however, about the best way to arrive at a score for each essay. We examine this using simulations and analyses of real world data. For real world data, results for a latent class signal detection model suggest that the raters, including the adjudicator, discriminate the latent classes of essays about equally well. In addition, the percent of cases that needed to be adjudicated was consistent with that expected from the signal detection model, which suggests that the adjudicated cases simply reflect randomness in the raters' perceptions. The increase in classification accuracy that is obtained by using a third rater is also shown.