

# ***Assisting Novice Raters in Addressing the In-Between Scores When Rating Writing***

Brittney Greer

A project submitted to the faculty of  
Brigham Young University  
in partial fulfillment of the requirements for the degree of  
Master of Arts

Norman W. Evans, Chair  
Wendy Baker Smemoe  
Troy Cox

## ABSTRACT

In the discussion of rating ESL writing assessments, borderline writing samples are mentioned, but a solution has yet to be addressed. Borderline samples are the problem raters face when a writing sample does not perfectly fit a set level within the rubric, but instead has characteristics from multiple levels.

The aim of this thesis is to provide an improved training module in the setting of an Intensive English Program by exposing new raters to borderline samples and rating rationale from experienced raters. The purpose of this training is to increase the confidence, consistency and accuracy of novice raters when rating borderline samples of writing. The training comes in the form of a workbook with a rubric and instructions for use, benchmark examples of writing, borderline examples of writing with comments from experienced raters defending the established scores, then a variety of writing samples for practice.

The selection of the benchmark and the borderline examples of writing was based mainly on infit statistics. Eight experienced raters participated in providing rationale based upon the rubric explaining why each borderline sample was rated with its established score, and describing why the sample could be considered at a different level.

In order to assess the effectiveness of the training workbook, 10 novice raters were selected to pilot the training workbook. Following the pilot, data was collected both statistically, based on the ratings they assigned to each sample they were asked to rate, and qualitatively through the means of a survey.

Results of the survey showed that rater confidence increased following the training, but that they needed more time with the training materials to utilize them properly. The statistical analyses showed insignificant changes, which could be due to the limitations of the data collection. Further research regarding the effectiveness of this training workbook is necessary, as well as an increased discussion in the field regarding the prevalent issue of rating borderline samples of writing.