

1056-62-814 **Loree Kim Killebrew*** (lkkilleb@bsc.edu), Box 540374, 900 Arkadelphia Road, Birmingham, AL 35254, and **Xinyan Yan** (xyan@bsc.edu), Box 540756, 900 Arkadelphia Road, Birmingham, AL 35254. *Rubric-scored Assessments: A Study of Inter-rater Reliability*. Preliminary report.

This study is primarily focused on inter-rater reliability. The subjects of the study are three university faculty members, two undergraduate students, and 44 participants in a mathematics content course for teachers entitled Numerical Reasoning. This course is offered by the Greater Birmingham Mathematics Partnership (GBMP), an NSF-funded Math Science Partnership (award #0632522). A series of rubric training sessions was given by the three faculty members who were familiar with the rubric and the course itself. Using 44 matched pre- and post-test completed by the teachers in the 2008 GBMP course, the undergraduate student-raters independently scored the tests on five different dimensions: Accuracy, Conceptual Understanding, Processes and Strategies, Verification, and Communication. Descriptive statistics were computed for the responses of the three questions on the pre-tests and post-tests. Kappa scores were used to determine inter-rater reliability. A minimum Kappa score of 0.400 was hypothesized as indicative of an acceptable level of inter-rater reliability. This paper reports efforts to achieve such a level of inter-rater reliability and the methods of training. The shortcomings of the training and suggestions for future study will be discussed. (Received September 17, 2009)