

Table 1: Are we interpreting the tasks, rubrics, and student work the same way?

	Task 1					Task 2				
Rater	Student 1	Student 2	Student 3	Student 4	Student 5	Student 1	Student 2	Student 3	Student 4	Student 5
Mode										
# of disagreements with mode										
Consensus score										

Table 2: Do consensus scores match our intuitions about differences and similarities in task difficulty and student performance?

	Student 1	Student 2	Student 3	Student 4	Student 5	Mean Scores
Task 1						
Task 2						
Total						