# **Introduction to Rater Training**

#### **COLORADO DEPARTMENT OF EDUCATION**

#### **COLORADO COMMUNITY COLLEGE SYSTEM**

## **Rater Training**

### • Why is rater training important?

- Promotes more accurate ratings
- Promotes higher inter-rater reliability
- Promotes higher intra-rater reliability
- We want to be able to meaningfully interpret and compare students' scores.
  - Test reliability and validity!
  - A student's scores should not depend on what rater he/she had.

## **Rater Training: Basic Steps**

Prior to rater training, tasks and rubrics are created, reviewed, and validated by experts. Rater training typically involves these steps:

1. **Task/Rubric**. Trainees review the task and scoring rubric.

2. **Sample Student Work**. Trainer guides trainees through samples of student work at various performance levels.

3. **Practice Scoring**. Trainees practice scoring sample student work. Trainer and trainees discuss scores. Trainees receive feedback on scores.

#### 4. Reliability Check

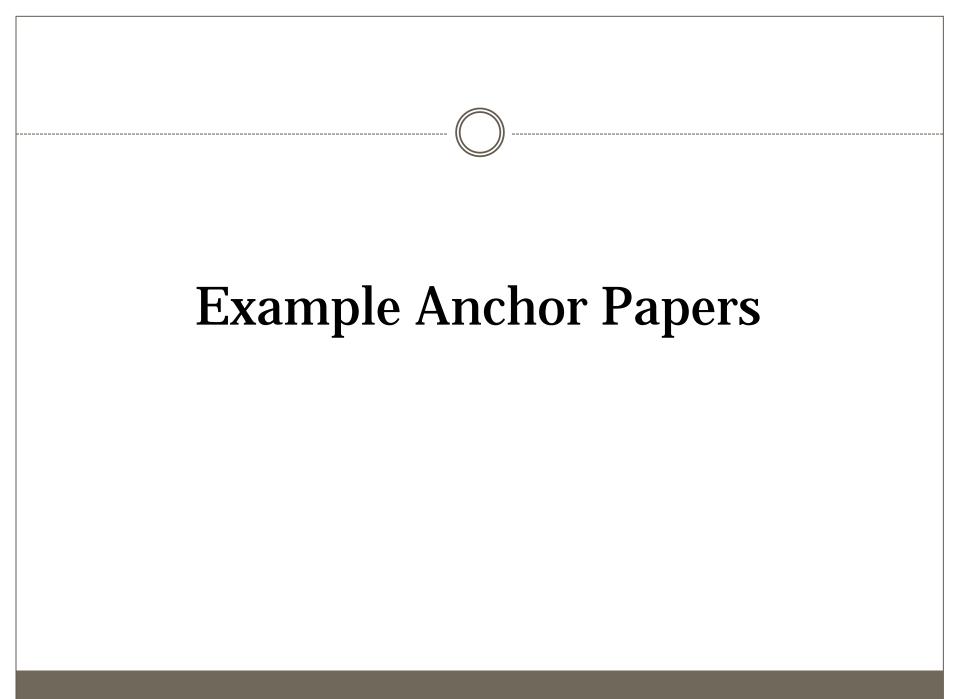
- Raters (trainees) are given information about:
  - The purpose of the assessment
  - The tasks in which students will engage
  - Rubric format (i.e., score points, layout)
  - Scoring criteria (with definitions of key words)
  - Rubric point descriptions

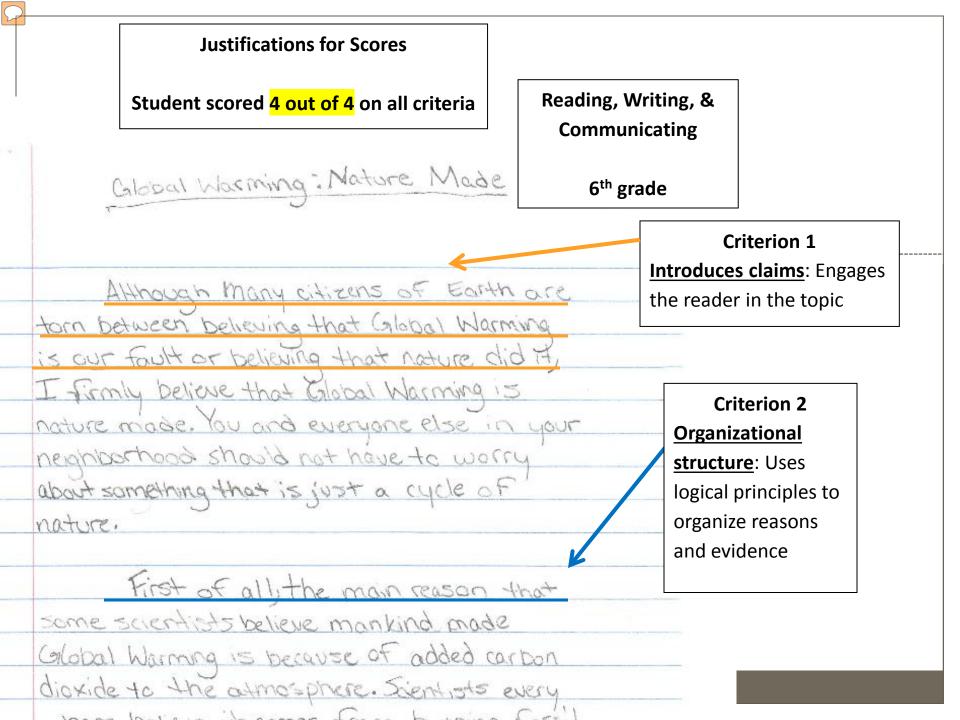
#### • Remind trainees that scores should be based on:

- Criteria in the rubric
- The behavioral descriptors (i.e., rubric point descriptions) with each criterion in the rubric

- Provide trainees with information about common rating errors:
  - o Halo
  - o Leniency
  - o Severity
  - Central tendency

- Trainees review "anchor papers" to help them envision what particular score points on particular criteria look like.
  - o "High" anchor paper
  - o "Medium" anchor paper
  - o "Low" anchor paper





**Justifications for Scores Criterion 1 Introductory claims:** Student scored 1 out of 4 on all criteria Personal opinions confuse rather than strengthen the claim. Global Warming is caused by natral interuptions. Those out that think it is human **Criterion 2** not a sientist. Sientist hove **Organizational** structure: The been and still are studying connection between reasons and evidence Global Warming. or the reasons and the central claim is unclear. Mow can be mans think its human caused bome of **Criterion 3** Supporting evidence: them barent even beren The writer alive For that long. The Sch demonstrates is abuard asshot as a green confusion about the house every ycor. Since IVe topic. study Global warming and

- Trainees score sample student work
  - Sample student work should represent various levels of performance.
  - Trainers might break up scoring each sample into sections.

#### • Purposes:

- Assess inter-rater agreement on the rubrics
- Identify potential issues with the tasks/rubrics

- Discuss scores and provide feedback to one another
  - Are there any inconsistencies in your scores?
    - **×** Talk about how each rater justifies his/her scores.
    - For major discrepancies/disagreements about scores...do the rubric point descriptions need to be re-written?
  - Is the rubric missing any important criteria?
  - Are any score points seldom-used? Why?
    Trial data will help with this discussion, as well.

## • Reliability Check

- Once you have agreed on scores to assign to the student work, your scores can be considered "expert" scores.
- If you want to train new raters, you can use the same sample work and your ratings to calibrate new raters' scores.

### Rater Calibration

• Raters who have not been 'calibrated' to a scoring rubric introduce error into scores.

## **Rating Practice Documents**

### Materials you will provide:

- o Student work samples
- **o** Rubrics

#### • Materials we have provided:

- Rating Practice Guidelines
- o Scoring Sheet
- Rater Calibration Tool (Excel doc)

All documents are on the website: http://www.coloradoplc.org/node/12765