



INTER-RATER AGREEMENT or RELIABILITY

Principal/School Leader Quick Guide

HOW CAN SCHOOL LEADERS ESTABLISH CLASSROOM OBSERVATION INTER-RATER AGREEMENT AND RELIABILITY?

In statistics and research, inter-rater reliability refers to the degree of agreement among raters and is represented as a score measuring how much consensus two or more judges or evaluators have in their ratings. For school leaders, the inter-rater reliability for the Teacher Assessment on Performance Standards (TAPS) does not need to be a score of how closely related two evaluators are in their observations, but rather a consistent understanding, agreement, and process for applying the TAPS rubric in teacher assessments. Evaluators that achieve consensus using the TAPS rubric have arrived at the same rating when observing the same teacher conducting the same lesson.

To ensure that all observers are on the same page in their ability to identify and rate observations with consistency...observers take time to view teacher practice together and then discuss what each person observed.

Marzano, R.J. (2007) *Art and Science of Teaching*

Credibility in an educator evaluation system is essential and requires (1) a consistent definition of good teaching, (2) a shared understanding of the definition, and (3) skilled evaluators (Danielson, 2011). Evaluators must be able to consistently assess teachers accurately so teachers accept the judgments as valid and have confidence in the results. While it is important to achieve inter-rater agreement among school leaders across the state and within a district, it is critical to establish inter-rater reliability for school leaders in the same building. Teachers expect school leaders to fairly and consistently assess classroom practice and provide the same level of low inference feedback to all teachers. Educators need observers to have high levels of understanding about the TKES process, performance standards, and rubrics and be able to recognize various classroom examples of the different components of practice. Establishing inter-rater agreement confirms multiple evaluators will consistently interpret evidence against specific levels of performance (i.e., furnish accurate assessment/observation, provide meaningful feedback, and engage teachers in productive conversations about practice).

HOW TO CREATE INTER-RATER AGREEMENT OR RELIABILITY:

Establishing inter-rater reliability among evaluators helps to address these common differences:

- Level of compliance with the scoring rubric,
- Degree of fairness and consistency exhibited when scoring examinee performance,
- Understanding and use of rating scale categories.

Inter-rater reliability should be established across all evaluators in a school before classroom assessments begin and should be verified throughout the school year. To create inter-rater reliability, school leaders should:

- meet with all school leaders after attending TKES credentialing/training to review the training and the TAPS rubric.
- develop a consistent definition of good teaching using the TAPS rubric.
- develop a common process for implementing TKES, monitor it for consistency, and improve it systematically.
- observe and rate the same videos of classroom instruction.
- view commentary on videos that have been categorized as proficient or exemplary.
- use instructional "rounds" to establish early inter-rater reliability on specific standards. Rather than look for all 10 performance standards, identify two or three that all evaluators will focus on.
- conduct joint classroom observations and compare ratings immediately after the observation. Discuss any observation differences and explain the rationale for the ratings.
- discuss feedback that each observer would provide to the teacher. Practice writing feedback together as a group.
- identify examples/indicators of various standards within the school.