Rater Agreement in Washington State's Evaluation System

Contents

Rater Agreement Background

The new Washington State Evaluation System for teachers and principals will require new learning and ongoing support to increase the effectiveness of implementation. The four framework authors would like to underscore the importance of the shared experience between evaluator and evaluatee.
To that end, the new rules (WAC 392-191A-050) support this intent:

The purposes of evaluations of certificated classroom teachers, certificated principals, and assistant principals shall be, at a minimum:

(1) To acknowledge the critical importance of teacher and leadership quality and support professional learning as the underpinning of the new evaluation system.

(2) To identify, in consultation with classroom teachers, principals, and assistant principals, particular areas in which the professional performance is distinguished, proficient, basic or unsatisfactory, and particular areas in which the classroom teacher, principal, or assistant principal needs to improve his or her performance.

(3) To assist classroom teachers and certificated principals and assistant principals, who have identified areas needing improvement, in making those improvements.

Overarching Roles and Responsibilities: District Office, Principals, and Teachers

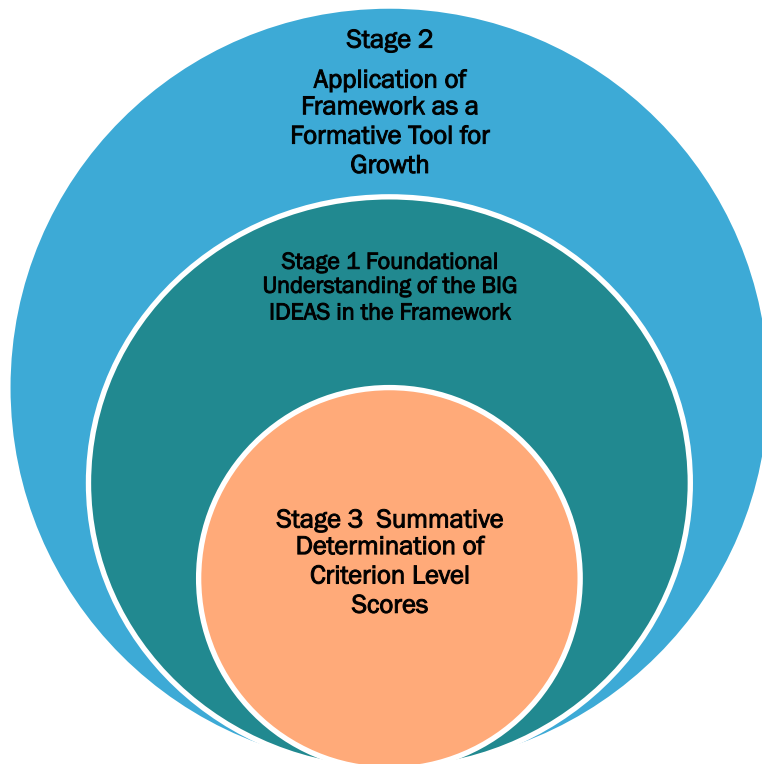|	District Office	Principal	Teacher
Rater Agreement Implementation Responsibilities	• Leads the oversight of evaluator training and ensures capacity building across a district. • Moves evaluators toward maximizing rater agreement and Stage 3 summative scoring. • Provides the structures (e.g., principal PLC time) and resources for teachers and principals to learn, implement, and sustain growth-oriented evaluation.	• Leads the implementation of the growth-oriented teacher evaluation processes. • Takes a collaborative learning stance with teachers in the evaluation process. • Takes responsibility for learning and sustaining their own rater agreement. • Provides the structures and resources for building level staff to learn, implement, and sustain the growth-oriented evaluation system.	• Engages in the implementation of the growth-oriented evaluation processes. • Takes a collaborative learning stance with the principal and colleagues in the evaluation process. • Takes responsibility for learning and applying the frameworks and rubrics to continually improve practice.
Communication	• Provides communication and sponsorship for immediate and long-term implementation.	• Provides communication and feedback to support immediate and long-term implementation	• Provides communication and feedback to support immediate and long-term implementation.

Rater Agreement Definition and Stages

The TPEP project is guided by the RCW 28A.405.100 and WAC 392-191A. The law requires that evaluators of both teachers and principals **"must engage in professional development designed to implement the revised systems and maximize rater agreement"** (RCW 28A.405.120 and RCW 28A.405.130**).**

Rater Agreement

The extent to which the scores between the raters have consistency and accuracy against predetermined standards. The predetermined standards are the instructional and leadership frameworks and rubrics that define the basis for summative criterion level scores.

The research and practical application of implementing this law is of primary concern for the Office of Superintendent of Public Instruction (OSPI), TPEP Steering Committee organizations, and the instructional and leadership framework authors.

In partnership with the three instructional framework authors and the Association of Washington School Principals (AWSP), OSPI will use the following working definition and steps to guide the 2013–14 implementation year:

Stage 2
Application of
Framework as a
Formative Tool for
Growth

Stage 1 Foundational
Understanding of the BIG
IDEAS in the Framework

Stage 3 Summative
Determination of
Criterion Level
Scores

Rater Agreement is the extent to which the scores between the raters have consistency and accuracy against predetermined standards. The predetermined standards are the instructional and leadership frameworks and rubrics that define the basis for summative criterion level scores.

Stage 1
OSPI, through the services of the Instructional Criteria Framework Feedback Specialists, will provide a two-day Stage 1 training for all who evaluate classroom teachers. This two-day overview provides an understanding of the "Big Ideas" of the Instructional or Leadership Frameworks and the inter-dependency of the frameworks, rubrics, and state criteria (dates pending for principal evaluator training).

Evaluators should know and be able to:	*Understand* the Big Ideas of the Instructional or Leadership Frameworks and the inter-dependency of the frameworks, rubrics, and state criteria. Accuracy in scoring depends on this foundational level. This foundational level understanding is the key to future calibration and application. ***It is recommended that this training occur prior to evaluating teachers or principals.***
Success Indicators:	Understanding of: • The appropriate use of the instructional language and framework vocabulary • The alignment of framework to state criteria • The interdependence of Dimensions/Indicators (CEL), Domains/Components (Danielson & Marzano) • The key ideas in each framework and what evidence would look like/sound like • The definition of essential aspects of the framework • The possible evidence for aspects of the framework
Districts should be creating plans to:	Provide the Stage I training to all evaluators before evaluating staff.

Stage 2
OSPI, through the services of the Instructional Criteria Framework Feedback Specialists, will provide up to 30 hours* of ongoing training on the frameworks for all who evaluate classroom teachers, principals, and assistant principals.

Evaluators should know and be able to:	*Apply* the framework and rubric in a ***formative*** process. This includes the capacity to give feedback, provide immediate support for implementing the feedback, and provide long-term professional development. Real life observation must be included in the application of this stage. ***It is recommended that this training occur during the year-long evaluation cycle.***

Success Indicators:	Formative Application of: • The appropriate use of language of instruction and framework vocabulary • The interdependence of dimensions/domains and indicators/components • The key ideas in each criteria indicators/components and what evidence would look like/sound like • The essential aspects of the framework • The evidence for all aspects of the framework • Develop expertise to provide formative feedback through professional conversations/coaching on the framework and rubrics to foster teacher growth development and implementation of short-term professional development plans
Districts should be creating plans to:	Move evaluators to accuracy and consistency through ongoing calibration conversations involving real-life or video-based observation. Consistency among raters is possible at this level.
Stage 3	
OSPI does not provide training for this stage although portions of Stage 2 training may have applicability toward Stage 3. Stage 3 should be integrated into the evaluation of principals (Criterion 5) and district administrators.	
Evaluators should know and be able to:	*Utilize* the rubric for summative purposes. This includes assessing artifacts, assessing observation data, and scoring of the Washington State Criteria. At this level evaluators should provide immediate support for implementing feedback and developing long-term professional development plans for teachers.
Success Indicators:	Summative Application of: • Appropriately use language of instruction and framework vocabulary • The interdependence of dimensions/domains and indicators/components • The key ideas in each criteria/indicators/components and what evidence would look like/sound like • The essential aspects of the framework • The evidence for all aspects of the framework and ability to determine summative criterion level scores using that evidence.
Districts should be creating plans to:	Utilize the criterion scoring document, guiding principles, and matrix. Utilize Summative Scoring Methodology from WAC 392-191A. Develop and sustain rater agreement. This includes summative scoring against a pre-determined standard.

Rater Agreement Glossary

ICFFs/LCFFs (Instructional or Leadership Criteria and Framework Feedback Specialist): A team of Washington State practitioners who are trained by framework author designees to provide facilitation around Stage 1 and 2 on the specific instructional or leadership framework chosen by OSPI as the "preferred instructional or leadership frameworks." These practitioners applied and were selected by a state-level committee through a rigorous process. **The ICFFs and LCFFs are not experts in the new evaluation law. They are gaining expertise around the frameworks by the framework authors as the foundational tool for the new evaluation systems for teachers and principals.**

FITs (Framework Introductory Trainers): A district or regionally-based team trained by framework author designees to provide a 1-day overview of the Instructional or Leadership Framework to evaluatees of the new teacher or principal evaluation system. The framework authors and the TPEP steering committee recognize that in order to create a shared experience, evaluatees should receive training on their framework. These practitioners were identified by either their own district or ESD to attend the training. The FITs are not **ICFFs and LCFFs** specialist, but are able to provide an overview of one of the instructional or leadership frameworks. The district-based FITs do not have responsibilities outside of providing training to teachers (instructional framework) or principals (leadership framework) within their own district.

Accuracy: A measure of observer data quality indicating the extent to which an observer is assigning scores that agree with scores assigned to the same observation by an expert rater; the extent to which rater's scores agree with the true or "correct" score for the performance.

Consistency: A measure of observer data quality indicating the extent to which an observer is assigning scores that agree with scores assigned to the same observation of practice by another typical observer. Consistency among the untrained is not what we are looking for. A goal of rater agreement is to ensure both accuracy and consistency.

Calibration: A process by which the regular practice of an observer's scoring is monitored and verified that the observer is still scoring accurately and consistently according to the standards and definitions of the framework/rubrics.

Artifact: Observed practice, products, or results of a certificated classroom teacher or certificated principal's work.

Evidence: Observed practice, products, or results of a certificated classroom teacher or certificated principal's work that demonstrates knowledge and skills of the educator with respect to the four-level rating system. (392-191A-030)

Feedback: Information aligned with a rubric provided to reduce discrepancies between current performance and desired performance. Effective feedback answers three questions:

- Where am I? (What are the performance goals based on a self-assessment of the rubrics?)
- Where am I going? (How is my performance related to the rubrics?)
- Where to next? (What actions do I need to take next to increase my performance?)

Criterion: The standards for teaching and school leadership as defined by RCW 28A.405.100.

Instructional framework: One of the preferred instructional frameworks adopted by the Office of Superintendent of Public Instruction to support the new evaluation system pursuant to RCW 28A.405.100.

The preferred instructional frameworks chosen by OSPI:

- CEL's 5D+ Teacher Evaluation Rubric 2.0
- Charlotte Danielson's Framework for Teaching (2011)
- Robert Marzano's Teacher Evaluation Model

Leadership framework: One of the preferred leadership frameworks adopted by the superintendent of public instruction to support the new evaluation system pursuant to RCW 28A.405.100.
(392-191A-030)

The preferred leadership frameworks chosen by OSPI:

- The AWSP Leadership Framework
- Robert Marzano's School Leadership Evaluation Model

Rubrics: The progression/description of practice used during an observation to capture evidence and classify teaching or leadership practice into differentiated aspects and performance levels. Typically consists of:

- Several Scales (components, domains, dimensions, indicators – there are numerous terms) See Architecture and Alignment Documents on TPEP website.
- A set of score levels applied within each scale to classify performance. The score levels are described in Washington State as: Distinguished (4), Proficient (3), Basic (2), Unsatisfactory (1)

Observe or observation: The gathering of evidence made through classroom or worksite visits for the purpose of examining evidence over time against the instructional or leadership framework rubrics. (392-191A-030)

Summative Criterion Scoring: Rating given to performance based on Washington State Criterion (see definition above). These scores will be based on an ongoing and varied process using a preponderance of evidence to determine final summative scores that promotes and recognizes growth.
This process is determined at the district level; guidance is available at http://tpep-wa.org/trainingpd/tpep-training-modules/.

Final Summative Scoring: Aggregation of the summative criterion scores. Final Summative Scoring Methodology approved by OSPI to determine final summative ratings. Raw Score Model
This process is determined at the state level; guidance written in WAC 392-191A.