If you want people to trust the results, what should be the goal? To achieve Inter-Rater Agreement.

{03 - SCHOOL/DISTRICT-LEVEL REFERENCE}
It’s the degree to which two (or more) people using the same scale get the same result in a similar situation.

Why is it important?
Because a good performance is a good performance – no matter who is judging it.

Signs that you have achieved Inter-Rater Agreement:
- The evaluation criteria are clearly defined.
- The evaluators have learned to see the same thing in the same situation before they cast official judgment.
- There is consistent judgment across people.
- The results are explainable.

> ADMINISTRATIVE COUNCIL ACTIVITY

*Develop a training plan to determine:*
- Content – topics to cover
- Timeline – process, not event
- Audiences – various groups
- Format – online v. face-to-face, video v. live
- Resources – including guide book
- Materials – sample videos, artifacts, evidence
- Communication/Promotion
- Assessment of effectiveness
- Sustainability plan

> THE WIFM FACTOR

*What’s In It For Me?*
- Credibility
- Buy-in
- A foundation for trust
- Predictability
- And so much more ...

ANY EVALUATION THAT HAS IRA …

<table>
<thead>
<tr>
<th>IS</th>
<th>IS NOT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent</td>
<td>Irregular</td>
</tr>
<tr>
<td>Objective</td>
<td>Subjective</td>
</tr>
<tr>
<td>Reliable</td>
<td>Iffy</td>
</tr>
<tr>
<td>Fair</td>
<td>Biased</td>
</tr>
<tr>
<td>Explainable</td>
<td>Indecipherable</td>
</tr>
<tr>
<td>Transparent</td>
<td>Mysterious</td>
</tr>
<tr>
<td>Defensible</td>
<td>Unjustifiable</td>
</tr>
<tr>
<td>Credible</td>
<td>Questionable</td>
</tr>
</tbody>
</table>
YOU’RE EVALUATING TEACHERS, TEXTBOOKS (ANYTHING, REALLY)

Start by building a common frame of reference so that evaluation is consistent.

Phase I: Distribute the Instructional Practice Protocol (IPP) framework

Phase II: Learn to apply the IPP framework
- As a team, watch 1-2 minute videos of exemplary performances of core practices.
- Discuss and determine what each core practice looks like at each performance level.
- Use longer videos to help identify and discuss rubric elements in the specific core practice.
- Practice scoring full-length master-scored classroom videos. Then discuss scoring decisions.

Phase III: Calibrate the way the team rates performances.
- Assess activities to demonstrate evaluators’ mastery of skills and agreement
- Recalibrate and reassess as needed
- Recalibrate periodically to retain accuracy and reliability

What is calibration?
Practice. Think of it as a dress rehearsal where you determine if everybody can hit the same mark at the same time. Begin with the administrative team.

1. Choose a focus. Target one area of instructional practice, such as student engagement, content knowledge or classroom environment.
2. Build shared understanding. As a team, watch videos of teaching performances in the targeted area. Then debrief as a team, using the relevant section of the district’s IPP instrument. Ask: What does each criterion mean? Does anybody disagree? If so, why? Make sure every participant speaks about each criterion.
3. Calibrate the team’s scoring. First, have all evaluators use the rubric to rate the performance in the video. Then debrief until the group reaches consensus on what constitutes each rating.

Repeat to calibrate team ratings of artifacts and evidence.