

TOOLS FOR

MAKING IT  
REALPre-Scoring Video to Clarify  
Expectations for Effective TeachingDCPS Advice  
for Facilitating  
Reconciliation

This document for facilitators gives tips for reaching agreement on correct ratings when pre-scoring video of teaching.

A key part of pre-scoring video for observer training and assessment is the reconciliation of different ratings provided by different expert observers (or “master coders”) for the same video. Examples produced from pre-scoring are more likely to show the right way to apply a rubric when multiple master coders agree on what the ratings should be and on the relevant evidence from the video to support those ratings.

District of Columbia Public Schools (DCPS) gives this advice to central office administrators who guide discussion among master coders, called “anchor raters” in DCPS. Individual anchor raters review and rate each video segment themselves, after which they submit their ratings and rating justifications. Pairs of anchor raters who reviewed the same segments then speak by phone with a facilitator to try to reach agreement on any differences. Lindsay Wallace Michael, a former member of the DCPS evaluator support team, created this document to help facilitators lead productive reconciliation discussions.

Included in these pages is a script to open the call with norms for discussion and a reminder of the goal to determine the best scores based on the district’s observation rubric, the Teaching and Learning Framework (TLF). Other advice suggests an order for comparing the ratings on individual components in the TLF (called Teach standards) and how to consider evidence of the performance indicators (called rows) for those components.

Facilitators do not themselves decide what should be the correct score. Rather, their role is to guide anchor raters in considering the evidence and how it aligns to the rubric. As noted in the advice, this may include reminding raters of a scoring policy in the rubric (e.g., if one indicator should be given more weight than others).



This is one in a set of tools cited in the MET project practice guide *Making It Real: Pre-Scoring Video to Clarify Expectations for Effective Teaching*. Publications under the banner of *Building Trust in Observations* offer practical advice to states, districts, and technical assistance providers on how to build and improve a trustworthy observation system. All MET project resources are at: [www.metproject.org](http://www.metproject.org).

## Open the session by reminding raters of the meeting’s purpose

“Thanks so much for the time you took to submit scores for this video, and for setting aside time this afternoon for our call. It’s exciting for me to have the opportunity to discuss this video with two really smart evaluators. Before we dive into our discussion, I’d like to quickly touch on a couple of norms for the call that will help our conversation go smoothly. First and foremost, remember that we have a common goal today - to come to agreement about what the best scores for this video are – and it’s OK if those scores are different from the ones you initially assigned it. In addition to creating a powerful tool for the district, we hope that the experience of participating in score reconciliation will be a rich professional development experience for all of our anchor raters – so, in that spirit, please be willing to challenge your colleague respectfully to consider the evidence from a different perspective, and also be willing to be challenged yourself so that together the two of you can come to the best final score. Let’s also agree to hold each other accountable to maintaining a rigorous focus on evidence and the TLF standards as we work through scores for this video, and to apply the scoring policies when relevant. “

## Roadmap: Provide an overview of the work to be done

“We’ll go through the evidence for each Teach standard one at a time. We have several scores with agreement, and we have disagreement for Teach standards 2, 3, 5, 6, and 7. We need to ensure that we have enough evidence to fully support all of the scores, but we’ll spend the majority of our time reconciling the standards for which we’re starting with disparate scores.”

It might be helpful to set time limits. Share the time limits with the raters at the beginning of the meeting, and monitor the time/provide time cues throughout the call.

## Summarize Teach standards with agreement

“You both scored Teach 1 as a 2, and your evidence was similar ...”

When raters agreed on the score – and perhaps much of the evidence, as well – start by summarizing what they agreed on, then focus on ensuring that the evidence fully supports the selected score. Evidence should be collected and discussed for every row of every Teach standard, including standards on which the raters agreed from the start.

## Guide the conversation on Teach standards with discrepancies

Provide an overview: “For Teach 2, we have a disagreement in scores – a 1 and a 3. Heidi, could you please start by sharing the evidence you saw for this Teach? Matt, after Heidi shares her rationale, you can share your evidence.” Try to let the first rater finish his/her thoughts before the other rater jumps in.

OR, if the evidence raters submitted allows: “For Teach 2, we have a disagreement in scores – a 1 and a 3. It looks like you both noted that the teacher did (fill in the blank), and so it appears that the difference in scores might be due to different perspectives on the evidence for (insert relevant rows). Heidi, could you please share about the evidence you saw for (insert relevant rows)?”

Raters do not need to agree on every row score, as long as they agree on the overall score for the standard. However, evidence should be discussed and recorded for every row of every Teach standard.

## Remind raters of the scoring policies when necessary

If raters are leaning toward a score that diverges from a scoring policy, gently remind them: “This is one of the scenarios for which we have a scoring policy. Since, the teacher had a 1 in (insert relevant row), s/he cannot receive an overall 4 for the Teach standard, unless the weight of the evidence provides a compelling reason to deviate from this policy. Do we have compelling evidence in this video?” Require the raters to articulate the

evidence – if evidence cannot be clearly and convincingly articulated, the evidence is not compelling enough to deviate from the policy.

### **When raters don't come to agreement**

When raters are having a difficult time coming to agreement, remind them that it is not necessary to agree on every row score – overall agreement for the standard is the goal. Provide a quick summary and then highlight the area of disagreement: “You’ve both noted that the teacher did (fill in the blank), and so it appears that the difference in scores is due to different perspectives on the evidence for (insert relevant rows). Let’s focus in on these points. Heidi, could you please share about the evidence you saw for (insert relevant rows)?”

If difficulty reconciling persists, guide the raters to move on to a different Teach standard and then come back to the difficult one later. If raters still are unable to come to agreement, consider the reason. Is it because of evidence collection challenges that are unique to video (e.g., audio/visual quality issues)? If so, this video may not be useful for training or calibration for the Teach standard in question and no more time should be spent attempting to reconcile the score.

### **When raters agree on a questionable score**

If you have serious reservations about a score that the raters agree on, please note this in the comment box when submitting final scores. Please be as specific as possible when explaining your concerns so that the next pair of anchor raters can pay particular attention to this area.

### **When scores don't feel quite right**

Sometimes the final scores might not feel quite right. For example, it might seem that if the rater had been in the classroom, s/he would have been able to see important evidence that could have led to a different score. However, because the video is the common experience/data set for every viewer, it is important to base scores only on the evidence that we have access to.

Remember that our goal is to create replicable, gold standard scores. To accomplish this, our video scores and evidence must align as precisely as possible to the language of the TLF rubric and the observable evidence in the video. If a score doesn't feel quite right, but the evidence matches the description of that score in the rubric, support raters in resisting the urge to change the score.