

MOST (Missouri Observation Simulation Tool)

Training and Calibration Description for Educators

Background

The Missouri Observation Simulation Tool (MOST) is intended to give administrators and teachers assistance in **calibrating observations** for the purpose of **inter-rater reliability** regarding **classroom observations**. Not all of the selected videos are exemplary examples, but rather give a realistic practice session to educators for calibrating purposes.

Two of the Department of Elementary and Secondary Education's seven principles from the Missouri Educator Evaluation System state that all Missouri districts will incorporate an evaluation system that:

- Provides ongoing, timely, deliberate and meaningful feedback on performance relative to research-based targets and
- Requires standardized, initial and periodic training for evaluators to ensure reliability and accuracy

MOST assists districts with these two principles. This document provides information about how videos were selected, scored and viewed by a group of Missouri Educators. It also provides strategies to use for calibration activities. In MOST, you will see sample scores and observation forms to be used as a guideline and point of discussion. Finally, this document gives educators specific support by providing examples of feedback that could be used with teachers following classroom observations.

Purpose

The overall purpose of observing and giving feedback is to affirm best practices used and help improve teaching quality. The purpose of MOST is to give evaluators practice in articulating strategies being used and the effectiveness of those strategies. It also gives evaluators practice in providing meaningful feedback to help teachers improve their practice. MOST gives administrators common language and consistency in teacher observations in a particular district.

Process of Preparing MOST for Missouri Educators

Missouri educators, including teachers, administrators, consultants, association representatives and retired educators gathered to view and score videos in an exercise to increase inter-rater reliability. This group is referred to as the "master coder". Scores from the "master coder" and observation forms were completed to provide a common baseline for discussion and calibration purposes. While those using MOST may vary in their ratings from the "master coder", they should consider the justification, comments and feedback provided by this group. It should be noted that MOST is a training tool, and its main purpose is to provide some structure for educators to have common discussion about classroom observations and research-based strategies. The final outcome of the completed observation form, the score and the feedback

to teachers may not necessarily be the same, depending upon the conversation had in various districts among different educators.

Strategies for Calibration Activities – MOST Observation Form

A variety of inter-rater reliability activities can be used with a group of educators, to come to consensus on articulating what is observed and providing feedback relative to research-based targets. Depending on the size and make-up of the group, educators may use any of the following strategies or others to facilitate discussion which will lead to calibration.

- Hand out MOST Observation Form Reference Sheet and discuss definitions of Strategies and Delivery Methods. Reflective questions to facilitate discussion about the definitions could include:
 - What are some of the strategies and delivery methods that might look similar to an observer? What are some key differences in those (i.e. group work and cooperative learning)?
 - What is an example of something that could be going on in a lesson which would involve more than one strategy and/or delivery method?
- Point out that the intent is not to find all of the strategies or delivery methods you can, rather to **indicate the most prevailing ones observed** in that lesson. If a delivery method is being used for only one minute of a twenty-minute lesson, you would not indicate that on the observation form, as that is not a prevailing strategy being used by the teacher.
- For each strategy or delivery method selected, first discuss Level of Engagement definitions and what that means. Discuss what it would look like in an observation if there was “high” engagement, “moderate,” “low,” and “disengaged.”
- After selecting a level of engagement then discuss Depth of Knowledge. Educators may want to have more training on DOK levels at this time or at another time. The more observers and teachers understand different levels, and how to move students to higher levels, the more impactful lessons will be.
- Show MOST video in its entirety. (No video is more than 15 minutes in length and videos are numbered to help with organization.)
- At the end of the video, participants should indicate which strategies **they saw**. If there is disagreement, conversation should continue about which are most prevalent.
- Participants then should indicate the level of engagement for each strategy. This could be done using clickers, post-it notes or other methods, but participants should have the opportunity to share why they chose the level of engagement they did. After discussing it, the group may want to “re-vote” for their level of engagement. The benefit is in the discussion and conversation, as this is where calibration and inter-rater reliability occur.
- The same process (clickers, etc.) should be used to determine DOK level. Again, conversation should occur so that participants can hear other points of view, and eventually inter-rater reliability occurs.
- Finally, discussion should be held about what comments will be made in the observations section of the form and the feedback section of the form. Observers

should continue to focus on feedback that is based on research-based practice. This can be done by pointing out certain research based strategies that are being used, and then providing other research based strategies which might increase the level of engagement and DOK level.

Strategies for Calibration Activities – Growth Guides

Growth guides on specific indicators may be used to open discussion about certain strands of an observation. For example, if the teacher is working on indicator 1.1 (content knowledge) for his/her growth plan, educators may want to calibrate on this specific strand. **Typically, a growth guide itself would not be used as an observation form, since the growth plans are meant to provide direction over a period of time, and with several forms of evidence including but not limited to classroom observations, lesson plans, data teams, student growth, etc.** That being said, it can be used as a calibration tool among educators in the same building or district. The purpose is to have conversation about specific strands of teacher performance.

If growth guides are used to calibrate observations, participants should focus **primarily on the “practice” portion of the growth guide**, since “commitment” and even “impact” may be difficult to determine in a video and/or in one observation.

Depending on the size and make-up of the group, educators may use any of the following strategies or others to facilitate discussion which will lead to calibration.

- Hand out and study the growth guide for a particular indicator. (Note some will be more appropriate than others for this inter-rater reliability exercise. In MOST, you will find the list of indicators which may be most suitable to the different videos.)
- Focus mostly on the “practice” portion of the growth guide. You may also want to discuss how you might be able to see “impact” in a classroom observation.
- Show MOST video in its entirety. (No video is more than 15 minutes in length and videos are numbered to help with organization.)
- At the conclusion of the video, have participants rate the score based on what they saw. Using clickers, post-it’s or other methods, determine the ratings amount the group. Facilitate discussion on the range of different ratings and “re-vote” after discussion occurs. The intent is to get as close to the same number score as possible in that group.
- Master coder scores are presented on MOST and should be used for groups to see what another group of educators in Missouri scored videos. Please refer to the comments and feedback section to determine justification for score.