

Developing and Maintaining Reliability on the Environment Rating Scales

Debby Cryer

Environment Rating Scales Institute

Many states and other groups that are using the Environment Rating Scales (ERS) need to train large numbers of people to reliability. However, training a large group to reliability is a substantial task, requiring days of work, and a commitment of staff as well as resources. The following represents the plan to be used in “high stakes” observations (those observations where ERS scores are used to determine a program’s status or funding).

We suggest training a core group to reliability with the authors of the scales and their associates. Often this might be a group of 6-8 people, but in some cases, is a larger group of up to 25 people. Other less costly (in terms of commitment and resources) training, for very large groups of people who do not do high stakes observations, for example, those who deliver technical assistance on the scales, can also be provided, but this can not substitute for the training to reliability procedure that is described below.

Note: This plan must be implemented separately for each of the different scales in the ERS series. Reliability on one scale is not an indication of reliability on any other of the scales.

Training to Reliability

1. Initial training of a core of assessors, who will later be able to train others

A period of 5 days training by at least one author, with the author’s associates, is required. The first day of training is classroom lecture in which the scale is introduced, and includes the following topics:

- What the scale is measuring
- How the scale measure the three basic components of quality
- The meaning of the scale scores
- Understanding reliability and validity of the scales
- Accurate scoring procedures
- Introduction to basic interpretations of terms used throughout the scale
- Observation procedures that are required for reliable use of the scale

Days 2 through 5 consist of mornings spent in doing independent observations with a small group of participants, led by a group leader with proven reliability and experience in training others to reliability. The afternoons are spent in debriefing sessions in which participants compare scores, and work to come to a correct consensus score through explanations of interpretation by the group leader as well as consideration of all evidence observed in the morning’s observation.

As a result of the five-day training, all participants are expected to have improved reliability scores, with at least one person usually found to be reliable.

- Acceptable reliability is defined as having 85% agreement (within one point) with the consensus scores.

- Reliability for any one observation is calculated by dividing the number of correct (within one point) scores by the number of items completed during the observation. Scores of N/A are counted in the calculation of reliability.
- An individual's reliability is based on the average of the three most recent reliability scores received. For example, if a person scored reliabilities of 75, 85, and 95, across the last three observations, the reliability would be the average of the three, or 85%. As a newer reliability score is added to the person's reliability history, the oldest score is dropped, and the newest added, so that the last three reliability scores are always used.

No official assessments should be completed by an observer who does not have a reliability average of 85%. However, new people can be trained (or a non-reliable assessor can be retrained) by a reliable assessor during official assessments.

2. Completing reliability training for all participants

Since acceptable reliability will generally not have been attained by all participants in the 5-day training, it is necessary to continue training. This is done by having the observer (or observers) who reached the acceptable reliability level continue to do observations and debriefing sessions with the remaining observers. This is continued until all participants have reached the acceptable level of reliability.

3. Establishing one or more "state anchors" for establishing statewide reliability and completing reliability checks across the state

Usually, the state anchor for reliability is the most consistently reliable observer. It is the responsibility of the state anchor to:

- Communicate with the ERS authors for clarification on interpretation when needed
- Communicate clarifications to all assessors in a state
- Complete reliability checks on observers throughout the state. One check is completed every 10th observation on a scale, until the observer being checked has reliability scores consistently, of 90% or above. High reliability observers require less frequent checks. Those who fall below the level of 85% require more frequent checks, until their reliability is consistently at or above 85%.

4. Maintaining reliability, expanding the reliability check system, and training new observers to meet acceptable reliability standards (See figure 1)

- Observers who have consistently high reliability scores of 90% or above are considered to be Level 1 observers. Level 1 observers can take on the responsibility of checking the reliability of others who have lower reliability scores. Those with lower reliability scores or those being trained are called Level 2 observers. In addition, they can provide the training during practice observations followed by debriefing.

- The anchor then takes on the task of checking Level 1 Observers, and checking a sample of level 2 observers. .

Figure 1: Roles and tasks of observers in ERS reliability system

