

GETTING RELIABLE ON OBSERVATIONAL MEASURES: *PICCOLO* OR THE *HOME VISIT RATING SCALES*

Adapted From: Roggman, L., Cook, G., Innocenti, M. (in development). *Home visit observation guide: Using the Home Visit Rating Scales (HOVRS)*. Baltimore, MD: Brookes; AND Roggman, L., et al., (2013). *Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO) User's Guide*. Baltimore, MD: Brookes.

Observational measures take practice to accurately score and agree with other observers. Observers must meet agreement criteria **on at least 4 observations** to be considered reliable.

Agreement criteria to establish on at least 4 observations:

PICCOLO: domain scores within 3 points, 3 or fewer items differing by 1 point or less on any domain.

HOVRS: scale ratings within 1 point, 3 or fewer items differing by 2 points or less in any scale (2 in Scale 7).

To learn to score observations reliably, **two or more people** should separately score the same observation and then compare their ratings, discuss any discrepancies, come to agreement on the ratings and their rationales, and continue until they have **4 observations meeting agreement criteria on the initial ratings**. Although two observers can become reliable by observing in person, we recommend using video when initially establishing reliability to allow repeated observations. When establishing reliability based on in-person observations, use the same agreement criteria but plan to make more observations to become reliable.

Videos make it easier to learn to use observational measures reliably. When possible, include videos similar to observations you will be scoring. With videos, two or more observers can establish inter-rater reliability by following these steps:

1. Observe one video at a time, score one or more domains or scales, and note your reasons. For beginners, it is helpful to work on one domain or scale at a time, then add another, then another, and so on. View the video with another observer or independently, but score independently and do not discuss your scores while scoring.
2. After scoring independently, compare your scores with another observers' ratings to identify items that do not meet agreement criteria. For each discrepancy, review the two sets of ratings, share your rationales, review the wording of items and scales or domains, re-reading any additional observation guidelines, such as in published *Guides*. Discuss these with another observer working on reliability.
3. Watch the same video again, pausing if needed to note specific observations. (Pausing the video when scoring is recommended only when learning to observe reliably.) Review each item that showed a discrepancy, rate it again, and write notes about specific behaviors you observed that provide a rationale for your rating.
4. Work with the other observer to **come to consensus** on any remaining discrepancies.
 - Discuss your differing ratings and your reasons or rationales for the differences.
 - **Do not vote or average**; listen to each other's reasons for assigning the score assigned.
 - Defend your score if you believe that what you observed meets the indicator criteria.
 - Together, decide the final rating and rationale for a score you truly agree on.
5. After coming to consensus, compare your ratings and rationales with the measurement authors (see *PICCOLO User's Guide*). Carefully read the authors' rationales for items where you disagree.
6. **Repeat with the same video until you meet the agreement criteria** on all domains or scales for that observation. Do not attempt to score another observation until you resolve all discrepancies in the previous observation. Continue until you meet the agreement criteria for at least 4 observations.