

Why Interrater Reliability Matters

As an observation-based tool, the *Assessment, Evaluation, and Programming System for Infants and Children (AEPS®)* provides an authentic assessment of a child's skills and abilities and generates useful information for instruction. However, one challenge in using observation-based measures is the need to trust the judgements of professionals conducting the assessments. Monitoring reliability is very important.



What is reliability?

Reliability refers to the stability and accuracy of assessment results. It answers the question: Is the tool producing consistent information across different circumstances?

What is interrater reliability?

Interrater reliability (IRR) refers to whether different people administering the same assessment tool can do so in a consistent way. A rating of interrater reliability looks at the extent to which two observers score an observed situation the same.

Why check interrater reliability?

Early learning experiences guided by authentic, accurate assessment data are considered a best practice for children. And, an assessment cannot be considered authentic if it cannot be conducted and scored as intended. Knowing that individual assessors are reliable and maintaining their scoring reliability over time cannot be assured unless steps are taken to check their ability. Measures of IRR contribute to more meaningful assessments for younger children.

How does the AEPS IRR Certification help?

The **AEPS** IRR Certification modules have been designed to give **AEPS** users, and their supervisors, confidence in their ability to be reliable scorers. In addition to the certification test, the modules include an **AEPS** Scoring Guide presentation, a Practice Test, and access to other resources to expand learning. Users that achieve 80% overall agreement with the gold standard scoring receive an **AEPS** Interrater Reliability Certification that lasts for three years.

Sources: Grisham-Brown, J., Hallam, R., & Pretti-Frontczak, K. (2008). Preparing Head Start personnel to use a curriculum-based assessment: An innovative practice in the "age of accountability." *Journal of Early Childhood Intervention*, 30, 271-281. Hallgren, K.A. (2012) Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23-34.