

Classroom Quality in Numbers: Using Multiple Measures of Classroom Quality and the Effect on Reliability and Validity

Cassandra Rowand Meagher and Natalya Verbitsky-Savitz

BACKGROUND

As a part of the FACES study*, Mathematica Policy Research designed and conducted a study to assess the extent to which using different observation measures and protocols to evaluate classroom quality may affect (1) the resulting scores and (2) the reliability of the measures.

The measures examined include the standard Early Childhood Environment Rating Scale-R (ECERS-R), the Classroom Assessment Scoring System (CLASS), and a shortened version of the ECERS-R. The short ECERS-R consists of 21 items (two factors) derived from the National Center for Early Development and Learning's (NCEDL's) multistate Study of Prekindergarten.

RESEARCH QUESTIONS

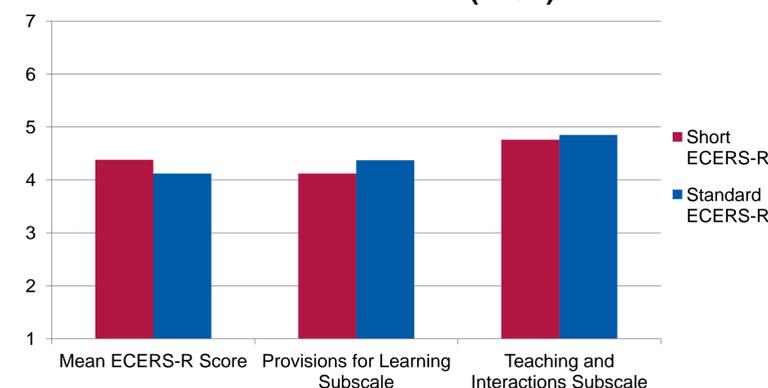
1. Do estimates of classroom quality differ based on whether observers use the short ECERS-R versus the standard ECERS-R?
2. Can a single observer be trained to score classrooms reliably using the standard ECERS-R one day and the CLASS on the following day? Are the results the same as two observers using the ECERS-R one day and the CLASS on the following day?
3. Does completing the short ECERS-R and CLASS on the same day affect the ratings of either, compared to completing the forms on different days?

METHODS

- All observers were trained to administer the CLASS; half were also trained on the short ECERS-R and half on the standard ECERS-R. All observers met or exceeded the publisher's minimum reliability criteria during training.
- A correlation of 0.7 or above was set a priori for acceptable reliability on the measures.
- Observations were conducted in 75 classrooms across five sites; both Head Start and community-based child care classrooms were included.
- 15 classrooms in each site were randomly assigned to combinations of four experimental conditions.

FINDINGS

Figure 1. No Mean Differences Were Found Between the Short and Standard ECERS-R Scores, Either Overall or on the Two NCEDL Factor Scores (RQ 1).



Note: The mean ECERS-R scores are based on an average across all available items. For the short ECERS-R, the score is based on 21 items. For the standard ECERS-R, the score is based on 37 items.

FINDINGS (CONT'D.)

Figure 2. The Inter-observer Reliability Correlations Between Standard ECERS-R Scores for Two Observers Completing the Standard ECERS-R on the Same Day Were Lower than the Acceptable Level Set A Priori (RQ 2).

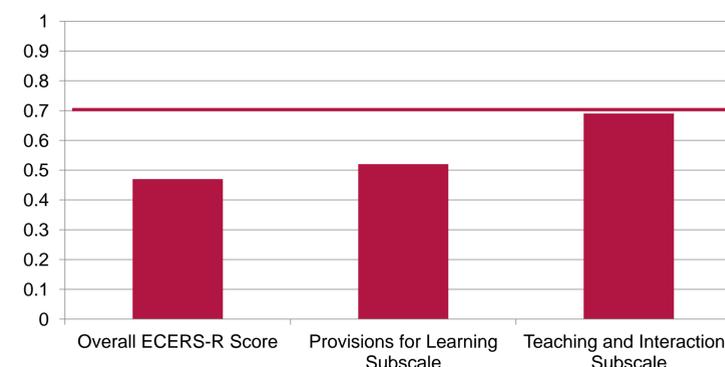
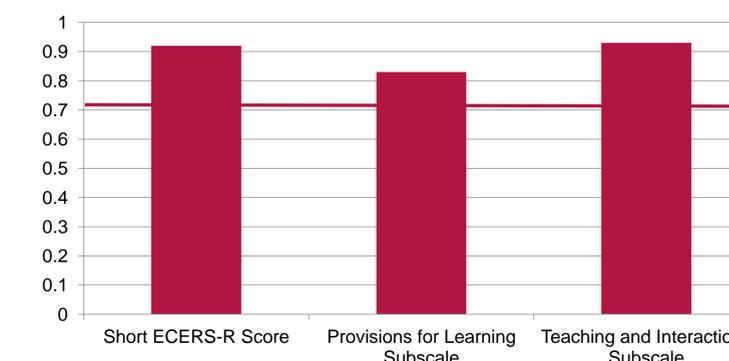


Figure 3. The Inter-rater Reliability Correlations Between Short ECERS-R Scores from One Observer Who Completed the Short ECERS-R and Another Who Completed the Short ECERS-R and the CLASS Concurrently were generally high (RQ 3).



CONCLUSIONS AND LIMITATIONS

- The short ECERS-R appears to be more reliable than the standard ECERS-R. A single observer can be trained to use the short ECERS-R and CLASS and will obtain more and higher-quality data than he or she would with the standard ECERS-R.
- Gaps in coverage within the short ECERS-R may need to be filled by other measures. For example, only one item from the "personal care routines" subscale on the standard ECERS-R is included in the short ECERS-R.
- Care should be exercised when interpreting a total ECERS-R score based on the short version; the short ECERS-R score assigns greater importance to the quality of teacher-child interactions, activities, and program structure than does the standard ECERS-R.
- Findings are based on a small and nonrepresentative sample of classrooms. Further study is recommended.