DEPARTMENT of EDUCATION
louisiana Believes

Early Childhood

Guide for Shadow Scoring *CLASS*® Observations

What is Shadow Scoring?
Shadow scoring is when two reliable CLASS® observers conduct a CLASS® observation at the same time in order to sharpen their CLASS® observation and coding skills. It is an avenue for maintaining inter-rater reliability and fidelity to the system.

Note: Shadow scoring will be used for the entire CLASS®observation (all four cycles) as seen in the example below.

Are Shadow Scoring and Double Coding the same thing?
Yes, shadow scoring and double coding are terms that are used interchangeably.

What is Inter-rater Reliability?
Inter-rater reliability occurs when two reliable observers produce consistent observation results for the same classroom at the same time. (This means that their results are at least 80% reliable overall.)

Does Bulletin 140 address Inter-rater Reliability?
Yes, Bulletin 140 states that each Early Childhood Community Network Lead Agency will conduct inter-rater reliability checks for each observer every year and 10% of all observations, with at least one shadow score in each age band.

What is the procedure for a Shadow Scoring observation?

Step 1	Step 2	Step 3	Step 4	Step 5
One person is assigned as the observer and one person is assigned as the shadow scorer.	Arrive at the site at the same time and begin and end Cycle 1 at the same time – should last 20 minutes (observe and record notes separately).	Take at least 10 minutes to score/code Cycle 1 independently without discussing.	Once scoring is completed for Cycle 1, review scores and discuss any Dimension scores that are off by 2 or more. Determine who has the supporting evidence most closely aligned with the CLASS® manual and come to a consensus on the score for the Dimension.	Repeat steps 2-4 for each remaining cycle. Count to determine the number of scores that were off by 2 or more across all 4 cycles. Subtract this number from the total possible scores. Then divide result by the number of total scores to determine the percentage of reliability overall. See Pre-K example below.

Example: The observer's scores are the blue circles. The shadow scorer's scores are the orange squares.

Cycle 1	Observer notes are not included in this example.						
Positive Climate	1	2	3	4	(5)	[6]	7
Negative Climate	(1)	2	3	4	5	6	7
Teacher Sensitivity	1	2	3	4	([5])	6	7
Regard for Student Perspectives	1	2	3	(4)	5	6	7
Behavior Management	1	2	(3)	[4]	5	6	7
Productivity	1	2	3	4	(5)	[6]	7
Instructional Learning Formats	1	2	3	(4)	5	6	7
Concept Development	1	(2)	[3]	4	5	6	7
Quality of Feedback →	1	(2)	3	[4]	5	6	7
Language Modeling	1	2	(3)	4	5	6	7

Legend: blue circle = observer score (shown as (n)); orange square = shadow score (shown as [n])

Cycle 2
Observer notes are not included in this example.

Dimension	1	2	3	4	5	6	7
Positive Climate	1	2	3	4	5	(6)	[7]
Negative Climate	([1])	2	3	4	5	6	7
Teacher Sensitivity	1	2	3	4	(5)	6	7
Regard for Student Perspectives	1	2	3	4	5	(6)	[7]
Behavior Management	1	2	3	(4)	5	6	7
Productivity	1	2	3	4	[5]	(6)	7
Instructional Learning Formats →	1	2	[3]	4	(5)	6	7
Concept Development	1	[2]	(3)	4	5	6	7
Quality of Feedback	1	(2)	[3]	4	5	6	7
Language Modeling	1	2	(3)	[4]	5	6	7

Cycle 3
Observer notes are not included in this example.

Dimension	1	2	3	4	5	6	7
Positive Climate	1	2	3	4	5	([6])	7
Negative Climate	([1])	2	3	4	5	6	7
Teacher Sensitivity →	1	2	3	[4]	5	(6)	7
Regard for Student Perspectives	1	2	3	(4)	[5]	6	7
Behavior Management	1	2	3	4	(5)	6	7
Productivity	1	2	3	4	(5)	[6]	7
Instructional Learning Formats	1	2	3	(4)	[5]	6	7
Concept Development	1	([2])	3	4	5	6	7
Quality of Feedback	(1)	[2]	3	4	5	6	7
Language Modeling	1	(2)	[3]	4	5	6	7

Cycle 4
Observer notes are not included in this example.

Dimension	1	2	3	4	5	6	7
Positive Climate	1	2	3	4	(5)	[6]	7
Negative Climate	[1]	(2)	3	4	5	6	7
Teacher Sensitivity	1	2	3	([4])	5	6	7
Regard for Student Perspectives	1	2	([3])	4	5	6	7
Behavior Management	1	2	3	(4)	[5]	6	7
Productivity	1	2	3	4	(5)	6	[7]
Instructional Learning Formats	1	2	3	(4)	5	6	7
Concept Development	1	(2)	[3]	4	5	6	7
Quality of Feedback	1	(2)	3	4	5	6	7
Language Modeling →	1	2	(3)	4	[5]	6	7

The **red** arrows identify the four Dimensions that have scores that are off by 2 or more. As mentioned in Step 4 above, the observer and shadow scorer will discuss these scores to determine the consensus score for each of these four Dimensions. **The observer will make note of the consensus score which will be used when the observer enters the scores into the *CLASS®* System. The observer will also enter his/her scores which are the blue** circles. *Note: The observer will be the only one to enter scores into the CLASS® System for this observation. No shadow scores are used.*

> **Number of Dimension scores that were off by 2 or more across all 4 cycles: 4**
> **Subtract from the total possible scores (40): 40 – 4 = 36**
> **Divide result by the number of total scores to get an overall percentage: 36 divided by 40 equals 90%**
> **The overall inter-rater reliability for this observation is 90% which meets the 80% or higher threshold.**