

How to check your data with interobserver agreement.

What it is. **Interobserver agreement (IOA)** is what you get when two people watch the same session, each record the same behavior on their own, and then compare.

The result is a percentage: how often they agreed.

Why it matters. Every decision on a graph assumes the dots are true. If two trained people watch the same child and write down 12 and 7, the graph is showing the observers, not the child. IOA is the cheapest way to find that out before a plan changes because of it.

THE THREE YOU WILL USE MOST

Pick the formula that matches how you took data.

TOTAL COUNT

For counting (frequency) data. $\text{Smaller count} \div \text{larger count} \times 100$. Fast, but loose: 5 and 5 agree 100% even if the two observers counted different moments.

INTERVAL-BY-INTERVAL

For interval data: the session split into boxes (say, 10 seconds each), each marked yes or no. $\text{Intervals where both agree} \div \text{all intervals} \times 100$. Stricter, because it checks *when*, not just *how many*.

DURATION

For how-long data. Shorter duration ÷ longer duration × 100. Same idea as total count, in seconds.

The usual bar. Aim for at least 80% agreement, checked in at least 20% of sessions, spread across conditions and observers. Those are common research conventions, not a law; your payer or program may set its own.

WORKED EXAMPLE

Same session, two answers.

Made-up example. The names and numbers below are invented to show the method.

A technician and a supervisor watch Maya for ten 10-second intervals and mark each one where she leaves her seat. Each of them marks 5.

INTERVAL	1	2	3	4	5	6	7	8	9	10
Technician	X	X	X			X	X			
Supervisor	X		X	X		X		X		

Total count says 100%. Interval-by-interval says 60%. Both marked 5, but they matched in only 6 of 10 intervals (1, 3 and 6 both marked; 5, 9 and 10 both blank). The loose formula hid the problem; the strict one found it. The fix is a clearer definition of “leaves her seat,” a short retraining, then check again.

MISTAKES TO AVOID

Ways IOA looks good and means nothing.

- Only using total count. It can read 100% when the observers never saw the same moment.
- Letting the second observer see the first one's sheet. Independent means independent.
- Checking only the easy sessions. Spread checks across people, settings and phases.
- Treating low IOA as someone's fault. It usually means the definition is fuzzy. Rewrite it so a stranger could score it.
- Fixing the data after the comparison. Record the IOA you got, then retrain.

TRY IT

One question, then the answer.

Two observers time a tantrum. One writes 90 seconds, the other 120. What is the duration IOA, and does it meet the usual bar?

▼ Show the answer, and why

75%, below the usual 80%. $90 \div 120 = 0.75$. The likely cause: they disagreed on when the tantrum started or ended. Write an onset and offset rule (for example, "ends after 10 seconds of quiet") and check again.

SOMETHING TO KEEP

IOA check sheet.

DRAGEN ACADEMY · IOA CHECK SHEET

· DRAGEN.CLOUD

BEHAVIOR DEFINITION (SO A STRANGER COULD SCORE IT):

--

INTERVAL	1	2	3	4	5	6	7	8	9	10
Observer 1										
Observer 2										
Agree?										

FORMULA	WORKING	RESULT
Total count: smaller ÷ larger × 100		
Interval: agreements ÷ intervals × 100		
Below 80%? What we will change in the definition:		

WHERE DRAGEN HELPS

The professional version of this page.

Trustworthy data starts with a definition a stranger could score and ends with a graph that shows what the data really say. DRAGEN teaches both next.

SOURCE

Where this comes from.

Cooper, Heron & Heward, *Applied Behavior Analysis* (3rd ed., 2020), chapter 5, on assessing the quality of measurement. Kazdin, *Single-Case Research Designs* (2nd ed., 2011), on agreement conventions.

Examples on this page are made up. They contain no client, student or family information. This lesson teaches a method; it does not replace your BCBA, your program, or your payer's written rules.