

Read before evaluating a comparison

Target: Decide whether a comparison isolates sleeve performance. Prerequisites: read units/time, distinguish starts and changes, subtract whole numbers. If needed use Kit 1 or the support below: 10 → 14°C is 4°C warming; 30 → 26°C is 4°C cooling.

First name the question. To compare complete sleeve designs, keep other relevant conditions matched. To investigate a single sleeve feature, vary that feature while keeping the rest of the sleeve design matched too. A confound is another relevant difference that prevents assigning the result to the intended factor alone.

Data P · Imagined setup cards for a paper comparison of complete sleeve designs. No results supplied.

Planned comparison conditions

Constructed teaching data: these values were invented for this activity, not collected in an experiment.

Condition	Sleeve A	Sleeve B
Water mass	100 g	100 g
Start	10°C	10°C
Cup and lid	Cup C + lid L	Cup C + lid L
Time / room	20 min / 22°C	20 min / 22°C
Reading method	Same method	Same method

K2-P1. What is being compared? What would be measured? Name two conditions kept the same.

K2-P2. Would changing both the sleeve and water mass still isolate sleeve performance? Explain.

Model · Two things changed

Data M · Room 22°C, same 15-minute exposure, matching cups/lids and thermometer procedure. Two complete sleeve designs; water masses and starts differ.

A flawed sleeve comparison

Constructed teaching data: these values were invented for this activity, not collected in an experiment.

Setup	Water mass (g)	Start (°C)	After 15 min (°C)
Sleeve A	100	8	14
Sleeve B	200	10	13

K2-M1. Find both temperature changes. Can the difference be assigned to the sleeve alone? Explain.

K2-M2. Name the confounds. Propose actual matching conditions for a better sleeve comparison.

K2-M3. What descriptive conclusion is still possible? Does comparing changes repair the confounds?

Guided practice · Repair setup cards

Data G · Paper plans, not collected readings. The question asks which complete sleeve better limits warming.
Two proposed setups

Constructed teaching data: these values were invented for this activity, not collected in an experiment.

Condition	Sleeve C	Sleeve D
Water mass / start	100 g / 10°C	100 g / 10°C
Cup	Cup C	Cup D
Lid	Lid L	No lid
Time / room	20 min / 22°C	20 min / 22°C
Reading method	Same method	Same method

K2-G1. Identify what differs besides sleeve design. Explain why that matters.

K2-G2. Rewrite D’s setup to match C’s relevant conditions while still comparing sleeves C and D.

K2-G3. Give a consistent measurement and recording plan, including repeated trials. If the question were about one sleeve material, what else would need matching?

Independent task · Fresh critique

Data I · Both setups are in a 24°C room. Cups, lids, exposure position and reading method match. The sleeve design, mass and duration differ.

New imagined comparison

Constructed teaching data: these values were invented for this activity, not collected in an experiment.

Setup	Mass (g)	Start (°C)	Time (min)	Final (°C)
Sleeve E	80	12	10	18
Sleeve F	120	12	20	16

K2-I1. Calculate each change. A fictional answer says “F works better because its water warmed less.” What is supported and what is too strong?

K2-I2. Identify the confounds and specify a repaired comparison with actual masses, starts and duration.

K2-I3. Describe how to repeat and record that comparison. State one limit even after repair.

Reteach · Subtraction does not fix a confound

Data A · Same 20 minutes in a 22°C room; cups, lids and method match. Sleeve and mass differ.

Equal starts, unequal masses

Constructed teaching data: these values were invented for this activity, not collected in an experiment.

Setup	Mass (g)	Start (°C)	Final (°C)
Sleeve H	60	10	17
Sleeve J	120	10	15

K2-A1. Find both changes. Can the lower increase isolate a sleeve effect? Repair the comparison.

Fresh check

Data Ac · New equal samples in a 23°C room, same cups/lids, method and 15-minute exposure. This is one matched trial.

A matched comparison

Constructed teaching data: these values were invented for this activity, not collected in an experiment.

Setup	Mass (g)	Start (°C)	Final (°C)
Sleeve U	100	12	17
Sleeve V	100	12	16

K2-A2. What bounded sleeve comparison is supported? What is not established?

Reteach · Specify a usable plan

Data B · Three separate trials of fresh 100 g samples starting at 10°C in a 22°C room. Cups/lids, position, 20-minute interval and reading procedure match within and across trials.

Matched final readings

Constructed teaching data: these values were invented for this activity, not collected in an experiment.

Trial	Sleeve J final (°C)	Sleeve K final (°C)
1	15	16
2	16	17
3	15	16

K2-B1. Which sleeve shows less warming across these trials? Does this prove a particular material property?

K2-B2. Write a paper retest plan: compared factor, matched conditions, measurement method, number of fresh trials, and fields to record. Do not invent new readings.

Fresh check

Data Bc · Fresh equal 100 g water, 23°C room, matching cups/lids and reading method. Times differ.

Check another plan

Constructed teaching data: these values were invented for this activity, not collected in an experiment.

Setup	Start (°C)	Time (min)	Final (°C)
Sleeve R	11	20	16
Sleeve S	11	10	15

K2-B3. What needs repairing before attributing the result to the sleeve alone?

Later check · Cooling and a repaired test

Data T · Paper cooling task. Both in a 22°C room for 15 minutes; cups, lids and method match. Masses differ.

First cooling comparison

Constructed teaching data: these values were invented for this activity, not collected in an experiment.

Setup	Mass (g)	Start (°C)	Final (°C)
Sleeve P	100	36	29
Sleeve Q	150	36	32

K2-T1. Find both falls. Explain why “Q better limits cooling because its final value is higher” does not isolate a sleeve effect.

Data Tr · Repaired paper record: fresh 100 g samples all begin at 36°C in a 22°C room. Cups/lids, position, method and 15-minute interval match within and across trials.

New matched trials after repair

Constructed teaching data: these values were invented for this activity, not collected in an experiment.

Trial	Sleeve P final (°C)	Sleeve Q final (°C)
1	30	32
2	31	32
3	30	33

K2-T2. What comparison is now supported across trials? Include one meaningful limit.

K2-T3. Does this paper record show that you constructed and tested a device? Explain.