Cool-Water Design Investigation — teacher

Teacher reference · Classroom copy · 2 November 2026 · Resource page and current files

Adult preparation: Review the operational setup card and procurement specifications before offering the practical route. Cup capacity is nominal (250–300 mL); approximately 8 cm is the external cup height, not a cylindrical internal-volume specification. Check actual matching cups around 7.5 cm internal diameter with broad stable bases at least 6 cm across. Verify that 150 mL fits, one 26 × 6 cm dry sheet (at most 3 mm thick) wraps the original cup with overlap secured by the two supplied pre-cut strips, and the maker's sensing region is fully immersed with the tip 1.5 cm above the inner base. Keep the reference bare and cup bases and covers uncovered. A nominal size label does not prove fit; if any check fails, select suitable matched equipment or use supplied data. Do not cut or extend materials to force a fit. Keep each redesign within the original two-sheet, 30 cm tape, 12 cm footprint and 9 cm height limits. Supply pre-cut materials, check cup stability, probe fit and response, and verify all starting and comparison conditions. Students follow the numbered assembly steps; they do not cut, heat, puncture or modify equipment. If equipment or conditions do not qualify, use the complete supplied-data route. Retain all inconclusive or excluded measurements and their reasons.

Free instructional files. Resource downloads and print choices .

MENTOR TEACHING · THERMAL ENERGY | Classroom copy · 2 November 2026

One-page teaching card

Target/prerequisites: compare temperature rises, keep conditions matched, retain anomalies and revise a design claim. Learners need subtraction with decimals, °C/mL/min units, and distinction between temperature and energy. The investigation can be used independently; Design Brief supplies its prerequisite support.

For practical work, print Prepare and Run the First Comparison, then Redesign and Retest for the separate redesign phase. For supplied-data reasoning, first give Part 1: compare and propose . Collect or retain the learner's written or recorded design, prediction, controls and reference before giving Part 2: reassess the record . Those constructed retest values belong to a specified fictional redesign, not to the learner's unbuilt proposal. Keep the full packet, mixed archives, later checks and complete teacher answers with the teacher; give only the selected phase. Use the separate Later Reasoning Check after reflection.

Teach: model assembly and a consistency check; rehearse one paired reading; check each group before water use. Notice: mismatched starts, moving probes, confusing endpoints with rises, counting timepoints as trials or discarding surprising values. Next: a fresh later check, question 1 is a fresh short reasoning check after reflection; if already exposed, use it as practice and record the exposure rather than calling it a fresh check.

Timing estimate per phase: 15 min setup/plan + three 13 min fill/settle/run cycles + two 5 min resets + 10 min comparison = 74 min. These are planning estimates, not measured classroom timings. In a 60 min session, stop after two trials; store dry labeled assemblies, retain records, and run trial 3 in a supervised 20–30 min continuation with a new instrument check. Keep first and retest phases separate. No equipment homework.

Operational setup and quantities

For G groups of 3–4: 3 G identical rigid stable plastic cups with matching loose covers and pre-existing probe openings; 2 G suitable non-glass probes; G dry spill trays; G measures marked at least every 2 mL; G timers; G probe holders; 3 G clean pre-cut felt sheets 26×6 cm≤3 mm; 40 G cm pre-cut tape (10 cm initial original sleeve + ≤30 cm redesigned sleeve). Keep spare dry sleeves and cups available. Supply enough printed forms for each fresh replacement. bare cup and original sleeve remain labeled; redesigned sleeve is a third cup.

Water allocation for both phases: at least 2.1 G L common cool water (1.8 G L for six pairs plus reserve) and 3 G L room-temperature conditioning/check water. Six after-run resets of two cups using 150 mL per cup require 1.8 G L; the remaining 1.2 G L allows initial conditioning, instrument checks and reserve. Provide extra for replacements and cleanup. For 6 groups: 18 cups, 12 probes, 6 trays/measures/timers, 18 felt sheets, 240 cm tape, 12.6 L cool water and 18 L room-temperature water. One air thermometer at cup height serves all groups; record its ID/location and each run’s room readings.

Adult prepares common 12–14 °C water from clean cool liquid water, mixing with room water if necessary; no student heating or ice in test cups. Use level tray surfaces, well away from power outlets/electronics, direct sun and drafts. Verify measured 150 mL produces sufficient depth for the entire sensing region at the stated 1.5 cm tip clearance. If geometry/probe cannot fit, obtain suitable matched equipment or use fallback.

Instrument requirements: check maker instructions for immersed portion, immersion depth, water suitability, safe range, minimum response and cleaning. Battery displays stay dry unless explicitly rated otherwise. No glass thermometers. Do not invent accuracy. 0.1 °C displays make the matching/check criteria readable; other instruments require an explicitly adapted protocol or supplied data.

Allowed material substitution: teacher-supplied dry craft foam sheets with the same dimensions and≤3 mm thickness may replace felt for the whole series. Label material/version and any change; do not compare a foam series as if it were the felt case. No allergen-prone loose fibers, sharp-edged sheet material, student cutting or hot glue.

Supervision, participation and cleanup

The completed records are fictional. Follow ordinary school procedures and adult supervision for practical work. Use the supplied data if equipment or conditions are unsuitable.

Follow applicable school-use rules and required protective arrangements. Adult checks stability, covered slots, probe holders, measurement depth and dry working area before release to groups. Water is nonconsumable. No heating, flames, dry ice, heat/cold packs, pressure devices, mains heating or food/medicine applications. Covers remain loose and unsealed.

Demonstrate carrying empty cups and pouring over the tray. Keep bags, books and electrical devices clear. For a spill: stop handling, notify adult, keep others away from wet floor, adult isolates nearby electrical equipment according to school procedure, absorb water and dry the surface. Never ask a student to handle a wet electrical device. A leaking/damaged cup leaves service; retain the run as flagged and replace with a matching stable cup before a fresh run.

Roles: timer/reader, recorder/checker, materials coordinator, evidence explainer. Rotate or pair roles; students can read a large display, dictate records, use accessible tables, or analyze supplied data without pouring/handling. Adult/peer can manipulate under learner direction. Equivalent reasoning does not require manual dexterity or graph drawing.

Cleanup: adult drains water into an appropriate sink under school procedures; no drinking or reuse as drinking water. Rinse/clean probes as instructed by maker, keep electronics dry, dry cups/trays, remove damaged/wet materials, wash hands and leave the floor dry. Store labeled dry assemblies between sessions; recheck stability, sleeve condition and instruments at the next session.

Retain empirical records including flags. Replacement trials get suffix IDs (for example F2a) and fresh samples. Never insert constructed readings into empirical sheets to finish an investigation.

Why these controls and limits matter

These setup choices are for this lesson; they are not established as optimal. Matching geometry/150 mL limits amount and area differences. Covered cups reduce evaporation and keep exposed surfaces comparable; they do not eliminate transfer or evaporation. Do not interpret rises as total energy, calorimetry or instantaneous rates. Cup-reset water approximates common initial wall conditions; wall temperature is not directly verified.

The start gap≤0.5 °C, volume allowance±2 mL, room shift flag>1 °C and reading window±15 s are operational tolerances for this lesson, not statistical uncertainty estimates or certified accuracy. Reject failed starts before timing; retain them separately. Small differences are interpreted cautiously even within tolerances. Prefer reporting each paired rise over a mean that hides procedural flags.

Minimum 60 s plus maker response interval is followed by two common-water reads 30 s apart. A≤0.2 °C change and≤0.5 °C pair agreement screens inconsistent/slow probes but cannot establish calibration. If maker response interval exceeds the practical schedule, lengthen supervised timing or use fallback; do not shorten it. Repeat checks after each phase; record any drift.

Swapping A/B and left/right after draining provides a partial balance. Odd three trials do not cancel every offset. Endpoint-minus-start subtraction reduces a constant offset’s effect but changing offset/response can still bias rises. Comparing concurrent pairs helps with room changes; distinguish runs made on different days. Three fresh pairs are an instructional choice to discuss variation, not a statistical reliability guarantee.

A dry redesign versus original sleeve is a system-level evaluation. If multiple features change, do not credit one alone. Keep assembled dry sleeves between runs; log/rebuild if wet/damaged. Interpret any rebuild as a procedural change. Any condition outside the specified range needs an explicit adapted record or supplied-data route, never a promised ranking.

Troubleshooting without manufacturing success

Negligible warming: check actual room/water gap, elapsed times, probe immersion/response and units. Retain valid small changes. Use supplied data for reasoning, or plan a new explicitly labeled longer-duration series (with fresh pairs and stated schedule); do not append it as the original 10 min result.

Indistinguishable designs: state that this setup did not establish a useful difference. Differences≤0.5 °C deserve particular caution because accepted checks/start gaps are of that scale; this is a teaching caution, not an uncertainty interval. Do not certify larger differences either. Check the full pattern and documented instrument information.

Inconsistent trials: inspect starts, volume, covers, timing, placement, sun/drafts, sleeve moisture and resets. Keep all data. Flag specific procedural concerns and obtain fresh replacements as time allows; mixed usable trials support an inconclusive claim, not selective deletion.

Drift/slow probe: repeat both common-water checks with equal depths, sufficient maker response and stable stirring. Record the disagreement. Replace/check the affected instrument; restart fresh pairs. Do not claim calibration or infer an accuracy specification.

Leaks/spills/wet sleeve: stop, apply spill procedures, remove damaged cup, log wetness/time. Replace with matching dry components, reset and use a new run ID. R2 in the fictional case models retaining this defect, not a clean redesign trial.

Missing/unavailable equipment: use the full constructed case plus questions; no learner must invent readings. One probe alone cannot do the concurrent method; do not silently substitute sequential readings.

Redesign performs worse: check constraints and conditions, report the worse/mixed outcome, seek a plausible next modification and test. Treat a feasible but ineffective design as useful evidence; do not force success.

Supplied-data teaching model

All six completed trials are explicitly fictional constructed records; they are not observations, experiments or computational simulations. Give supplied first-comparison record only for first analysis. Have learners commit to a redesign, question 1 before revealing the supplied redesign/retest record. State that the supplied redesign/retest record describes specified fictional redesigned sleeve and does not test their own proposal.

Model F1: bare cup rise 16.0−13.0=3.0 °C; original sleeve rise 15.3−13.1=2.2 °C; difference 3.0−2.2=0.8 °C. All first pairs give 0.8 °C. Say “less warming in these constructed comparisons,” not “felt cools water” or “no energy enters.” Starts match within 0.5 °C but are not identical; rises are preferable to bare endpoints.

Retest: R1 rises 2.3 versus 1.9 °C; R2 2.4 versus 2.7 °C with wet-sleeve flag; R3 2.3 versus 2.0 °C after logged rebuild. Do not erase R2. Two dry differences 0.4/0.3 °C are small and fewer than the intended three clean trials. Conclude improvement remains unestablished, then specify fresh dry matched retesting. R2 does not prove wet felt is worse: spill/changed conditions were not isolated.

Sparse valid response: “original sleeve warmed less,0.8 each time. Only this setup.” This shows evidence and scope even without polished prose. Ask for unit/run labels if unclear. A drawing/table/oral explanation can carry the same reasoning.

This fallback covers planning, evidence interpretation and claim revision. It does not demonstrate that learners actually constructed, measured or tested a device. Do not report empirical learning outcomes from completing it.

Methods, standards and source scope

The completed numerical records are fictional cases; the blank forms are for your own measurements. Use suitable instruments and ordinary school procedures. No classroom outcomes or product performance are established by the supplied records.

Official source checked 30 September 2026: https://www.nextgenscience.org/pe/ms-ps3-3-energy informs this investigation: engineering a thermal protector, directional transfer and temperature/energy distinction. The investigation offers construction/testing opportunities only when actually carried out; no total transferred-energy calculation or blanket standards coverage is claimed.

Official source: https://www.nextgenscience.org/pe/ms-ets1-3-engineering-design informs comparison and redesign analysis: compare designs using criteria/constraints and recognize useful redesign evidence. Source elements are paraphrased; original tasks/tables/setup text are not copied lessons or diagrams. This investigation alone does not establish full standards coverage or endorsement.

Official source: https://www.nsta.org/nstas-official-positions/safety-and-school-science-instruction informs the practical safety guidance: science activity requires school safety arrangements, training/procedures and attention to hazards.

No device-specific maker accuracy is asserted. Teacher records selected device and checks its official immersion/response/cleaning instructions. Operational matching, stability, timing and repeat criteria are specified lesson choices; they have not been empirically validated. No manufacturer endorsement or accuracy estimate is inferred.

The teaching text, tasks, forms and constructed values are original project materials. Source links identify the ideas used; no organization endorses this resource. Classroom copying and adaptation follow the Mentor Teaching Terms of Use: https://mentorteaching.com/terms-of-use/.

Answers and accepted alternatives

the design brief, question 1

Aim: less temperature rise over 10 minutes than the identified reference under matched conditions. Constraints: at most two 26 × 6 cm sheets, each ≤3 mm; ≤30 cm tape; footprint ≤12 cm diameter; height ≤9 cm; base/top/probe slot remain exposed (any three). Inconclusive: missing/comparably flawed trials, differences small relative to instrument/condition concerns, mixed directions or negligible warming. No fixed achieved temperature is promised.

the design brief, question 2

Accept a testable prediction such as original sleeve will have a smaller 10-minute rise than bare cup, with dry felt possibly slowing transfer through the sides. Other predictions are accepted with reasons and later evidence review. Room surroundings are warmer than the water, so net energy can enter; exposed base, cover, slot and other paths remain. Insulation is not an active cooler or a guarantee of no transfer.

assemble and label, question 1

Accept an accurate labeled description or sketch of bare cup/original sleeve with dry side sleeve only, identical loose covers, equal 150 mL water, central immersed sensing region and tip 1.5 cm above base without contact. Match cup geometry, water amount/start, cover/slot, probe placement, timing, tray/environment; explain how differences could affect comparison.

Answers and accepted alternatives

instruments and trial order, question 1

Do not start using this pair: 0.7 °C exceeds the 0.5 °C consistency limit. Record, check positioning/response, replace/check instrument or use fallback. 0.1 °C is display resolution, not demonstrated accuracy or calibration.

instruments and trial order, question 2

Swap to examine/reduce consistent position and sensor effects. Three timepoints follow the same water/cup run; they are repeated readings, not independent fresh trials. Three fresh paired fills provide repeated runs; they do not establish statistical reliability.

instrument-check record, question 1

Actual equipment facts vary. Identify both exact probes and phase/date/check ID, supported immersion and response conditions, air instrument/location and maker information if available. “Accuracy not available” is acceptable and limits claims. Do not make a device-specific specification up or call this calibration. Record first and second readings of both probes in each common-water condition before/after phase; flag failure to settle, >0.5 °C disagreement, >0.2 °C change or new drift.

Answers and accepted alternatives

First comparison · trial 1, question 1

Empirical values vary. A complete record identifies group/date/time, reference/prototype and cup/sensor positions, volume 150 mL within 2 mL each, dry sleeves/unchanged assembly and identical covers, batch/start/room conditions, response interval, and a linked completed instrument-check record. Use class shared check record if clearly linked by ID. Do not invent missing checks. A failed check is retained and marks the evidence limited/inconclusive.

First comparison · trial 1, question 2

Accept factual logged anomalies or explicitly “none noticed”; classify usable only when matched start, appropriate instruments, full endpoint readings and consistent procedure/room conditions are documented. Flag/missing conditions limit comparison. Retain anomalous rows; add a separately identified fresh replacement run without erasing or changing originals. No universal numeric expected measurements.

First comparison · trial 2, question 1

Empirical values vary. A complete record identifies group/date/time, reference/prototype and cup/sensor positions, volume 150 mL within 2 mL each, dry sleeves/unchanged assembly and identical covers, batch/start/room conditions, response interval, and a linked completed instrument-check record. Use class shared check record if clearly linked by ID. Do not invent missing checks. A failed check is retained and marks the evidence limited/inconclusive.

Answers and accepted alternatives

First comparison · trial 2, question 2

Accept factual logged anomalies or explicitly “none noticed”; classify usable only when matched start, appropriate instruments, full endpoint readings and consistent procedure/room conditions are documented. Flag/missing conditions limit comparison. Retain anomalous rows; add a separately identified fresh replacement run without erasing or changing originals. No universal numeric expected measurements.

First comparison · trial 3, question 1

Empirical values vary. A complete record identifies group/date/time, reference/prototype and cup/sensor positions, volume 150 mL within 2 mL each, dry sleeves/unchanged assembly and identical covers, batch/start/room conditions, response interval, and a linked completed instrument-check record. Use class shared check record if clearly linked by ID. Do not invent missing checks. A failed check is retained and marks the evidence limited/inconclusive.

First comparison · trial 3, question 2

Accept factual logged anomalies or explicitly “none noticed”; classify usable only when matched start, appropriate instruments, full endpoint readings and consistent procedure/room conditions are documented. Flag/missing conditions limit comparison. Retain anomalous rows; add a separately identified fresh replacement run without erasing or changing originals. No universal numeric expected measurements.

Answers and accepted alternatives

Retest · trial 1, question 1

Empirical values vary. A complete record identifies group/date/time, reference/prototype and cup/sensor positions, volume 150 mL within 2 mL each, dry sleeves/unchanged assembly and identical covers, batch/start/room conditions, response interval, and a linked completed instrument-check record. Use class shared check record if clearly linked by ID. Do not invent missing checks. A failed check is retained and marks the evidence limited/inconclusive.

Retest · trial 1, question 2

Accept factual logged anomalies or explicitly “none noticed”; classify usable only when matched start, appropriate instruments, full endpoint readings and consistent procedure/room conditions are documented. Flag/missing conditions limit comparison. Retain anomalous rows; add a separately identified fresh replacement run without erasing or changing originals. No universal numeric expected measurements.

Retest · trial 2, question 1

Empirical values vary. A complete record identifies group/date/time, reference/prototype and cup/sensor positions, volume 150 mL within 2 mL each, dry sleeves/unchanged assembly and identical covers, batch/start/room conditions, response interval, and a linked completed instrument-check record. Use class shared check record if clearly linked by ID. Do not invent missing checks. A failed check is retained and marks the evidence limited/inconclusive.

Answers and accepted alternatives

Retest · trial 2, question 2

Accept factual logged anomalies or explicitly “none noticed”; classify usable only when matched start, appropriate instruments, full endpoint readings and consistent procedure/room conditions are documented. Flag/missing conditions limit comparison. Retain anomalous rows; add a separately identified fresh replacement run without erasing or changing originals. No universal numeric expected measurements.

Retest · trial 3, question 1

Empirical values vary. A complete record identifies group/date/time, reference/prototype and cup/sensor positions, volume 150 mL within 2 mL each, dry sleeves/unchanged assembly and identical covers, batch/start/room conditions, response interval, and a linked completed instrument-check record. Use class shared check record if clearly linked by ID. Do not invent missing checks. A failed check is retained and marks the evidence limited/inconclusive.

Retest · trial 3, question 2

Accept factual logged anomalies or explicitly “none noticed”; classify usable only when matched start, appropriate instruments, full endpoint readings and consistent procedure/room conditions are documented. Flag/missing conditions limit comparison. Retain anomalous rows; add a separately identified fresh replacement run without erasing or changing originals. No universal numeric expected measurements.

Answers and accepted alternatives

compare the first trials, question 1

Empirical answer follows retained records and correct subtraction in °C, with flags. Constructed F1/F2/F3: bare cup rises 3.0, 3.2, 3.1 °C; original sleeve rises 2.2, 2.4, 2.3 °C; paired differences each +0.8 °C. Units and run IDs required. A difference of rises, not just endpoints, is used because starts differ slightly.

compare the first trials, question 2

Constructed answer: original sleeve warmed 0.8 °C less than bare cup in each of three matched constructed pairs; evidence is consistent with less warming for this setup. Limits: fictional values, small number of runs, instrument accuracy not supplied, imperfect sensor/position balance, finite interval/room conditions, other heat paths. Do not claim no transfer, universal best material, statistical significance or cooling. Empirical mixed/small/flagged results warrant inconclusive wording.

compare the first trials, question 3

The comparison evaluates these assembled whole designs in the stated conditions. Side sleeve changes coverage/contact and possibly surface interactions; evidence does not uniquely prove a mechanism or isolate a material property. Accept a matched further comparison of sleeves with the same size/thickness/attachment but different material, or one feature changed with all other conditions maintained; even then bound causal claims to controlled conditions.

Answers and accepted alternatives

commit to a redesign, question 1

Accept any feasible removable dry sleeve with stated quantities/dimensions and tape ≤30 cm, ≤2 supplied 26×6 cm sheets ≤3 mm each, footprint ≤12 cm and height ≤9 cm, exposed base/top/slot, stable cup. Prediction states less/similar/more warming than concurrent original sleeve with a reason. Controls: geometry, water volume/start, room/timing, cover/probe position, surface and systematic sensor/position swaps. No promise of improved performance. Multi-feature redesign can fairly compare whole designs but cannot assign the effect to one feature alone.

commit to a redesign, question 2

Different start temperature, water amount, room/drift and timing could explain an endpoint difference. Compare each fresh concurrent matched pair’s rises; the historical first series is useful context but not a substitute for today’s reference.

reassess the redesign, question 1

Constructed R1: original sleeve +2.3 °C, redesigned sleeve +1.9 °C, difference +0.4 °C, dry; R2: +2.4, +2.7, difference −0.3 °C, wet-sleeve flag; R3: +2.3, +2.0, difference +0.3 °C, rebuilt dry. Retain R2 and explain that it differs procedurally; do not average all three without qualification or remove it because it conflicts. Dry differences are small and there are only two dry retests, one rebuilt: improvement is not established. Empirical answer depends on all records/conditions; accept worse, inconclusive or improved only with bounded evidence.

Answers and accepted alternatives

reassess the redesign, question 2

Constructed redesigned sleeve meets ≤2 sheets, ≤30 cm tape (20), footprint ≤12 cm (9), height ≤9 cm (8), exposed surfaces. Meeting resources does not establish thermal improvement. Revised claim: specified redesigned sleeve gave slightly smaller rises in two dry constructed pairs; flagged wet trial warmed more; evidence insufficient to establish improvement. Next: repeat three fresh matched concurrent original sleeve/redesigned sleeve pairs with fully dry intact sleeves, same assembly/volume/starts/room/timing/probe placement and sensor/position swaps; retain all outcomes. Accept a different declared modification and comparison if feasible and justified.

evidence and explanation, question 1

Warmer surroundings can transfer energy into cooler water. A sleeve may reduce net transfer through some pathways, while uncovered paths and imperfect insulation remain. Positive rise means warming, not cooling. Readings compare temperature changes for equal water/cup conditions, not total energy directly or instantaneous rates. Endpoints cannot uniquely prove a particular transfer mechanism.

evidence and explanation, question 2

Trial variation is changes among fresh runs; display resolution is shown increment (0.1 °C); accuracy is how close a reading is to a defensible reference and needs maker information/checking, not inferred from resolution. Limits may include small trial count, matching tolerance, probe response/offset/drift, wet/rebuilt sleeve, ambient/position differences, short duration, approximate cup reset or constructed-data provenance. A consistency check is not calibration; there is no statistical reliability/guaranteed ranking claim.

Answers and accepted alternatives

evidence and explanation, question 3

Accept correct tabulation or graph with time in min, temperature °C, reference/prototype labels, all three readings and run ID/provenance; annotate anomaly if relevant. Constructed F1 shows both warm; bare cup rise 3.0 vs original sleeve 2.2 °C. Other record patterns accepted. Do not demand a graph or claim an instantaneous rate from endpoints.

a fresh later check, question 1

Y rises 3.0 °C; Z rises 2.0 °C. Start gap 2.0 °C exceeds 0.5 °C and gives different thermal driving conditions; neither rise comparison nor final reading establishes better protection. The readings at 10/15/20 follow one sample, not three fresh trials. Keep record flagged; start fresh concurrent 150 mL pairs, both 12–14 °C within 0.5 °C, room 20–24 °C at least 6 °C warmer, same geometry/covers/probe placement and timing; obtain three fresh matched trials with swaps and all readings.

Fictional teaching response — sound for the named task

Named task: Supplied data part 1, Compare the first trials, questions 1–3

Fictional response — sound for the named task: “F1 rises are 3.0°C bare and 2.2°C original sleeve; F2 3.2 and 2.4°C; F3 3.1 and 2.3°C. The bare-minus-original rise difference is 0.8°C in every pair, and the supplied first-comparison record flags no anomalies. The complete original assembly shows less reported warming in these constructed matched comparisons. It does not isolate felt alone, establish instrument accuracy or prove physical performance. For a hypothetical next test, I would compare a dry two-sheet sleeve with the concurrent original one-sheet sleeve, using two supplied 26 × 6 cm sheets and 20 cm total tape for the proposed sleeve, keeping its base/top/slot clear and checking the stated size constraints. Use three fresh matched 150 mL pairs starting near 13°C in 22°C room air for 10 minutes, with matching cups, covers, probe checks/placement and the supplied between-trial position/probe schedule. Retain starts, 5/10-minute readings and flags. This tests that added-layer assembly proposal, not felt alone, and does not approve apparatus or supply its outcome.”

Visible evidence: All-pair labeled rise calculations, flags, a bounded assembly comparison and a limit/test direction for the named analysis questions.

Not established: Physical apparatus skills, the learner’s full redesign proposal, or that a later supplied redesign result tests the learner’s own unbuilt idea.

Clarify: Ask why each sample’s own start is subtracted and why a whole assembly differs from a material-only causal claim.

Next teaching/evidence: Complete and keep the Commit to a redesign proposal in part 1 before giving supplied-data part 2. That later record describes the specified fictional design. Reserve Later reasoning assessment for another occasion.

This response is invented for teacher interpretation. Accept defensible alternatives. Record access, prior exposure and the actual response; do not assign an ability label or claim mastery from this example.

Supplied data, part 1 — compare and keep the proposal

Supplied data, part 2 — new evidence after the proposal

Reserve: Later reasoning assessment

Fictional teaching response — partially developed

Named task: Supplied data part 1, Compare the first trials, questions 1–3

Fictional response — partially developed: “The original sleeve warmed less in all three trials.”

Visible evidence: A comparison consistent with the supplied rises.

Not established: The six rise calculations, difference/flag record or limits on the claim.

Clarify: Ask for each 10-minute-minus-start calculation and the bare-minus-original rise difference, then whether these are constructed or actual observations.

Next teaching/evidence: Complete the supplied Compare the first trials table and explain its limit. Keep the redesign proposal before part 2; if subtraction itself is unclear, use the practice-only A value is not a change file from Read Temperature Data Before Explaining It.

This response is invented for teacher interpretation. Accept defensible alternatives. Record access, prior exposure and the actual response; do not assign an ability label or claim mastery from this example.

A value is not a change

Reserve until after this practice: A value is not a change — fresh check

Reserve: Later reasoning assessment

Read Temperature Data Before Explaining It — matching teacher directions and answers

Fictional teaching response — plausible error

Named task: Supplied data part 1, Compare the first trials, questions 1–3

Fictional response — plausible error: “F1’s difference is 0.7°C because 16.0 − 15.3 = 0.7. That proves felt alone causes a 0.7°C improvement.”

Visible evidence: Accurate subtraction of the final endpoints.

Not established: The requested difference in rises: starts differ by 0.1°C, so rises are 3.0 and 2.2°C and their difference is 0.8°C. The comparison is of complete assemblies, and the fictional record does not prove a material-only cause.

Clarify: Have the learner calculate and label both rises before comparing them, then list what differs between assemblies.

Next teaching/evidence: Use the practice-only A value is not a change file, then its separate reserved fresh check. Return to all three supplied F pairs and a bounded assembly claim before retaining the redesign proposal and revealing part 2.

This response is invented for teacher interpretation. Accept defensible alternatives. Record access, prior exposure and the actual response; do not assign an ability label or claim mastery from this example.

A value is not a change

Reserve until after this practice: A value is not a change — fresh check

Reserve: Later reasoning assessment

Read Temperature Data Before Explaining It — matching teacher directions and answers