Science 5–6

Chapter 4 — Evaluating

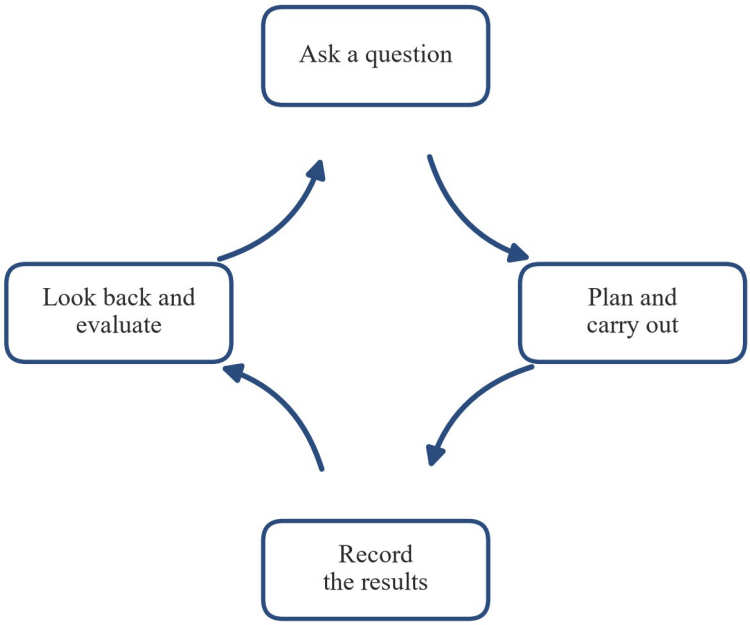
Aboriginal and Torres Strait Islander content has been excluded as accuracy cannot be guaranteed, and it is better to be respectful than cover the entire curriculum inaccurately.

Significant effort has been made to ensure this content is legal. Please send any areas of concern to admin@vwx.au with appropriate document/legal references so errors can be verified and changes be made.

Comparing findings, finding the errors, asking the next question

Words to know method — the steps we follow to carry out a test • **variable** — a thing we could change in a test; in a fair test only one is changed • **trial** — one go at the test; doing repeats means doing extra trials • **error** — a mistake in the method or in a reading that makes the findings harder to trust • **data** — everything we record when we do a test • **evidence** — the readings we choose to support our explanation • **conclusion** — what we decide from the evidence • **absorb** — soak up • **findings** — what the results of a test tell us

A test is not finished when the equipment is packed away. In this section we **evaluate**: we look back at the method and the findings, compare our work with other people's work, hunt down the errors, and choose the evidence that supports our conclusion. Then we do what scientists always do at the end of a test — we ask the next question.



Looking back at the work raises the next question.

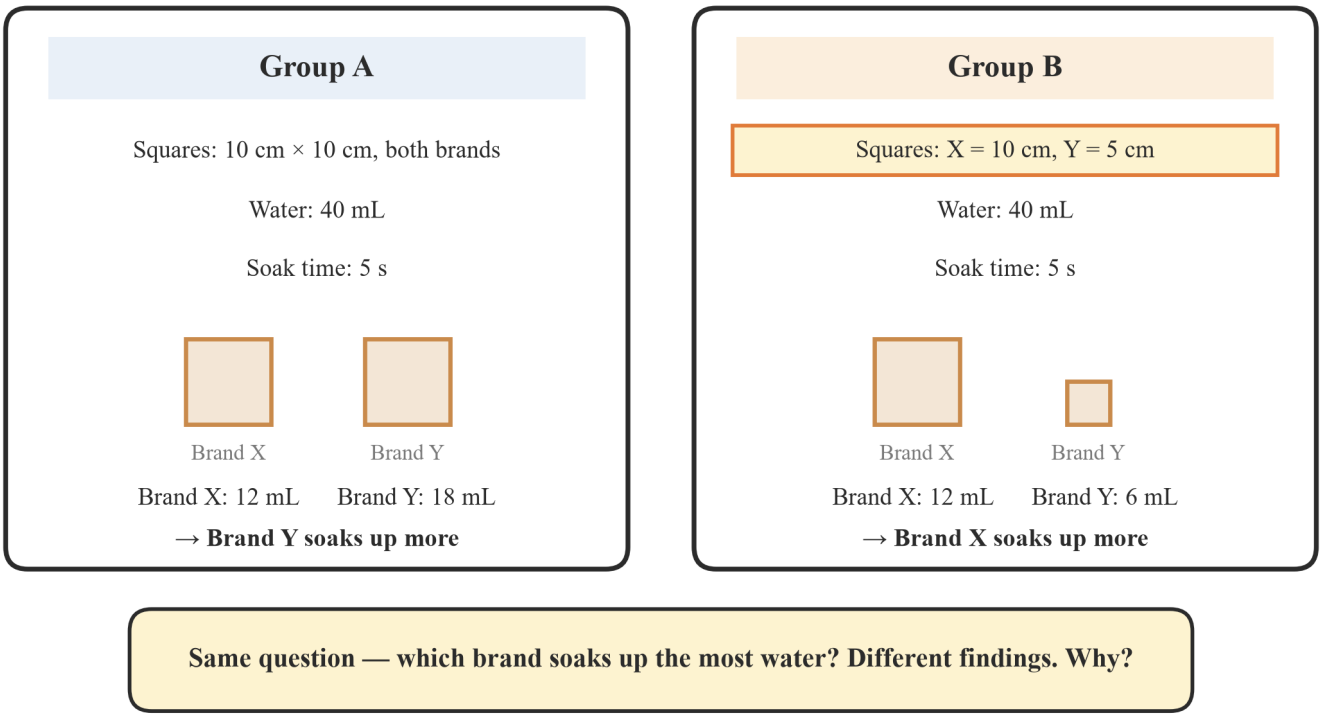
A circle of four boxes joined by arrows: Ask a question, Plan and carry out, Record the results, Look back and evaluate, showing that looking back raises the next question

Theory

Same question, different findings?

Two groups in one class tested the same question — *which brand of paper towel soaks up the most water?* — and got different findings. When that happens, we do not just shrug. We put the two **methods** side by side and look for a difference, because a difference in the method can cause the difference in the findings.

(SCENARIO data)



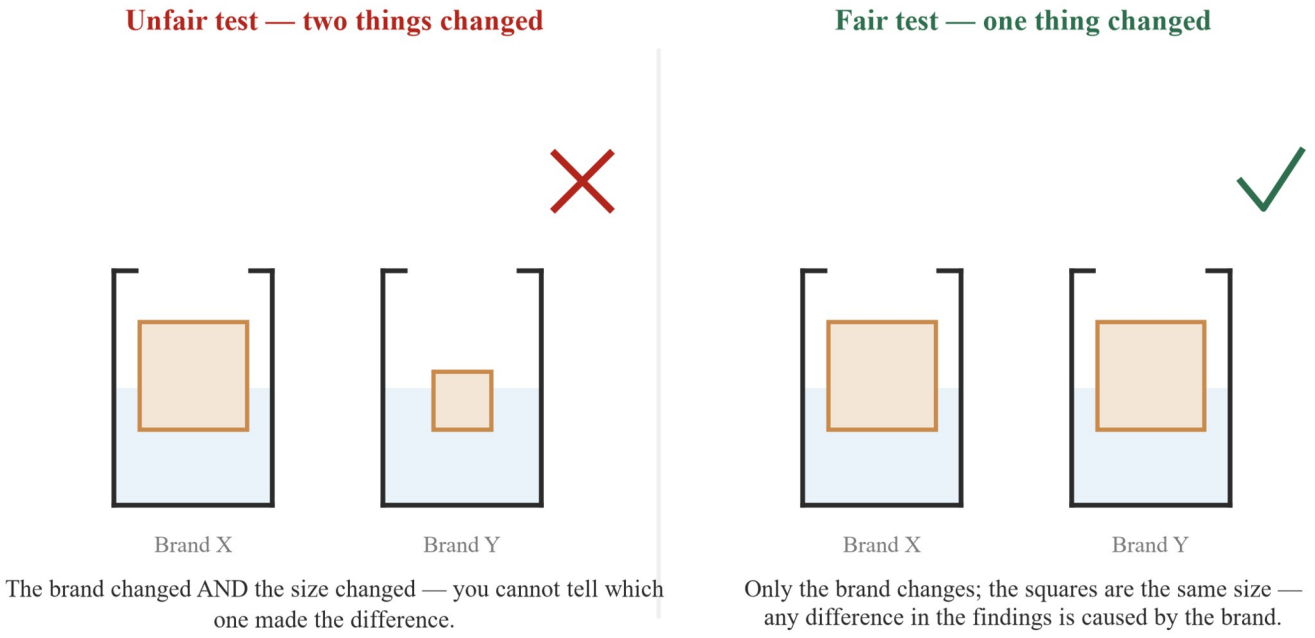
Two method cards for the same question: Group A used the same size square for both brands and found Brand Y soaks up more; Group B used a smaller square for Brand Y and found Brand X soaks up more, with the size difference highlighted

Both groups started 40 mL of water and soaked for 5 s. But Group B used a 5 cm × 5 cm square for Brand Y while Brand X got a 10 cm × 10 cm square. Group B changed two **variables** — the brand *and* the size — so when its findings came out different we cannot tell which variable caused it. Group A's findings are the ones we can trust for the brand question.

The errors Years 5 and 6 actually make

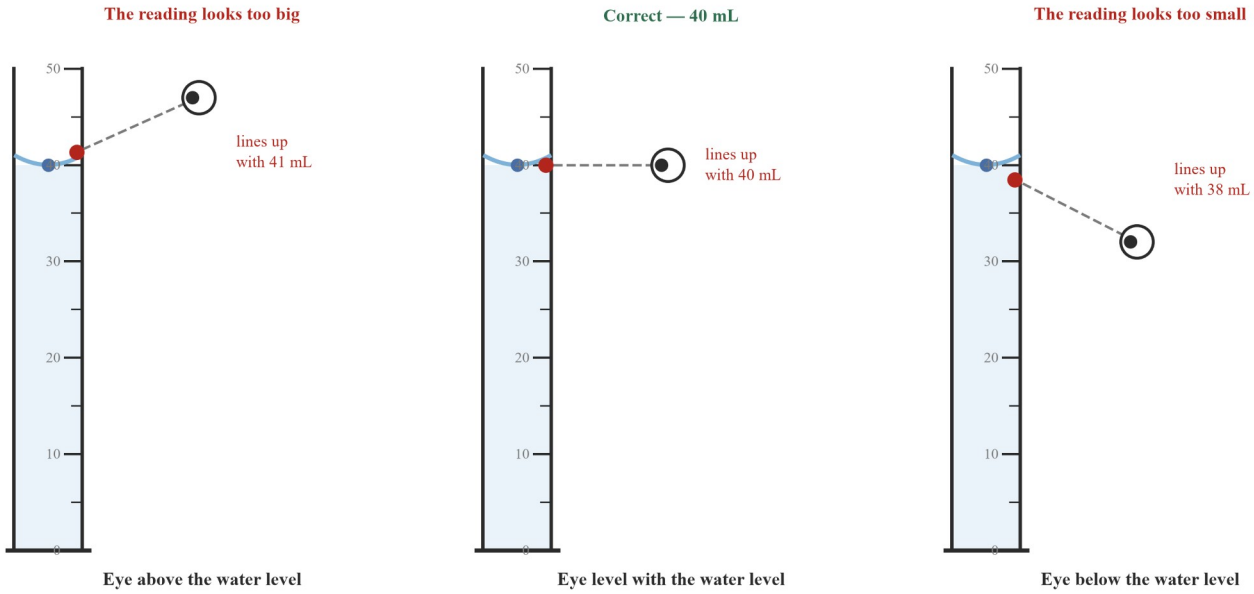
Most of the time, a difference between our findings and someone else's comes from one of four everyday **errors**.

Changing more than one variable. A fair test changes only the variable we are testing. If two things change, the test is unfair and the findings cannot show us anything certain.



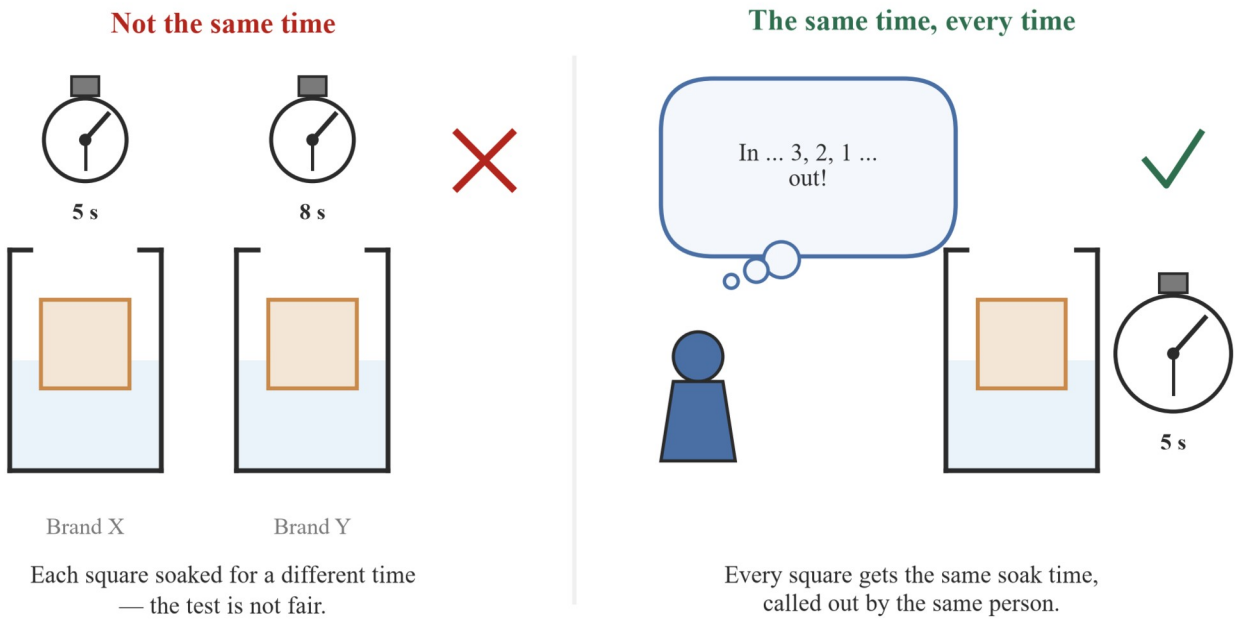
Two beakers with different-size towel squares labelled unfair because the brand and the size both changed, beside two beakers with same-size squares labelled fair because only the brand changed

Misreading a scale. On a measuring cylinder the water surface curves. We read the bottom of the curved surface, and we do it with our eyes level with the water. If our eyes are above the water, the reading looks too big. If our eyes are below, the reading looks too small.



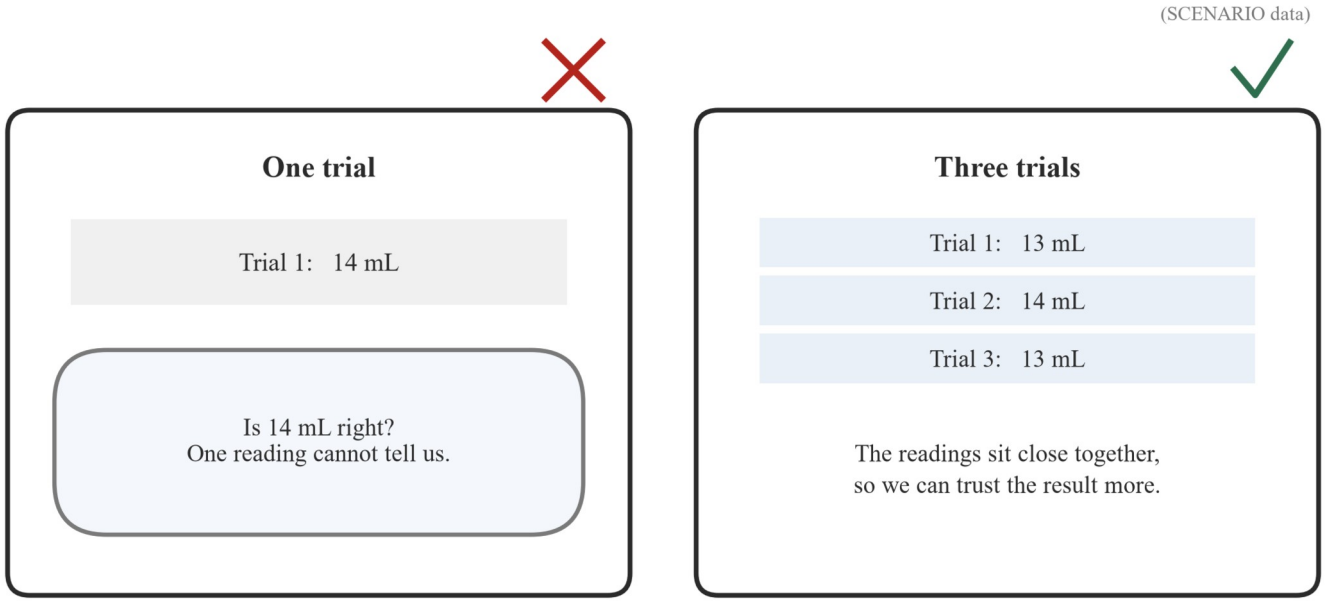
Three measuring cylinders holding 40 mL: an eye above the water level lines up with 41 mL and looks too big, an eye level with the water lines up with 40 mL and is correct, an eye below lines up with 38 mL and looks too small

Not giving every trial the same time. If one towel square soaks for 5 s and another soaks for 8 s, the extra time — not the brand — could cause the difference.



Left: two beakers with stopwatches showing 5 s and 8 s, marked wrong because the squares soaked for different times; right: a person calling out the time so every square gets the same 5 s

Doing too few trials. One reading cannot tell us much. If the one trial went wrong, we would never know.

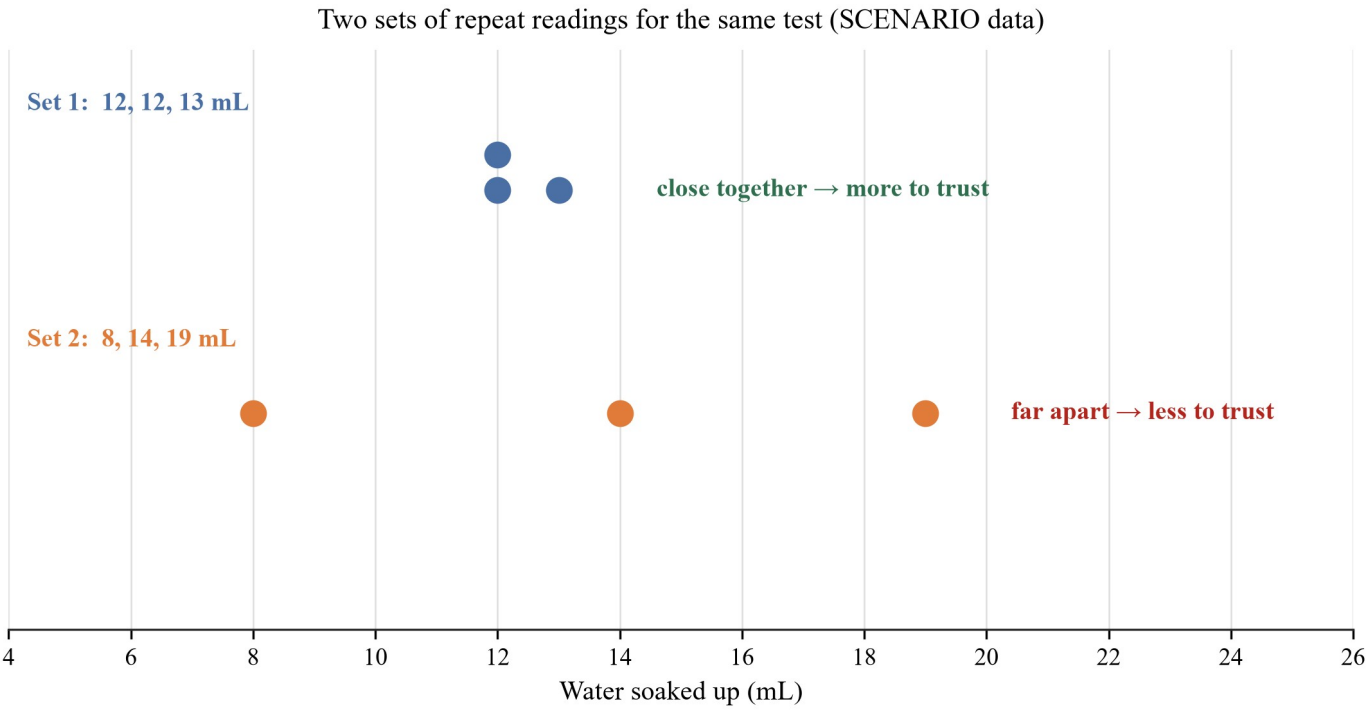


One trial card showing a single reading of 14 mL with a thought bubble saying one reading cannot tell us, beside a three-trials card showing 13, 14 and 13 mL sitting close together

Repeat readings: close together, and believable

Doing three **trials** lets us check our own work. We line the readings up and ask two questions.

Do the readings sit close together? We do not need an average to judge this. We simply compare the readings, see how close they sit, and spot any reading that does not fit. A reading that does not fit gets done again.



Two sets of repeat readings plotted on a number line: 12, 12, 13 mL sitting close together and marked more to trust, and 8, 14, 19 mL spread a long way from each other and marked less to trust

Is the reading believable? Every reading gets a quick sense check. If we poured 40 mL of water, a towel cannot soak up 80 mL — a reading like that cannot be true, so something in the method or the reading went wrong and we do it again.

(SCENARIO data)

Check: is it believable?

We poured 40 mL of water.

Reading: 12 mL soaked up

✓ Believable — 12 mL is less than the 40 mL we poured.

Check: is it believable?

We poured 40 mL of water.

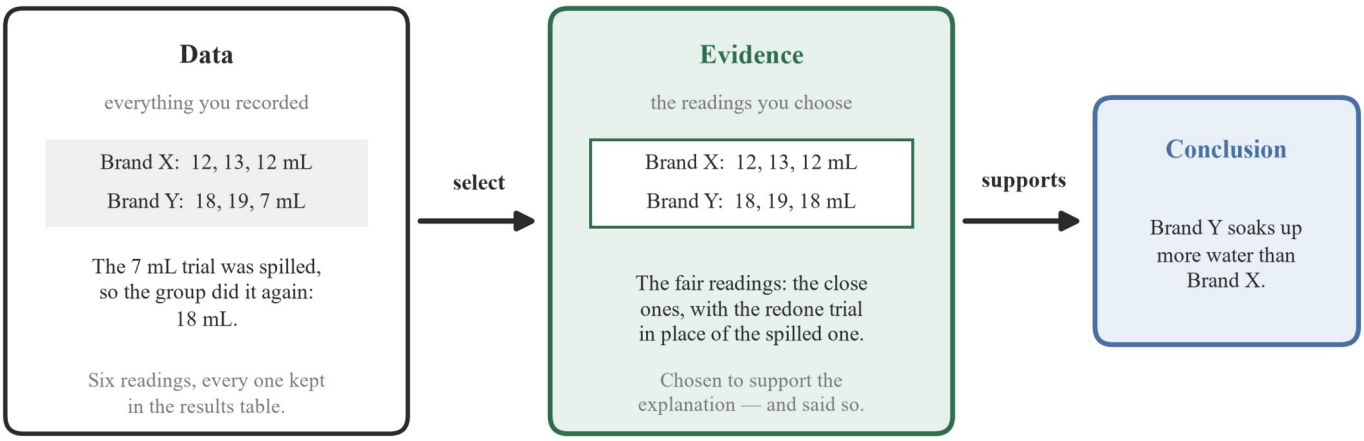
Reading: 80 mL soaked up

✗ Not believable — a towel cannot soak up more water than was poured!

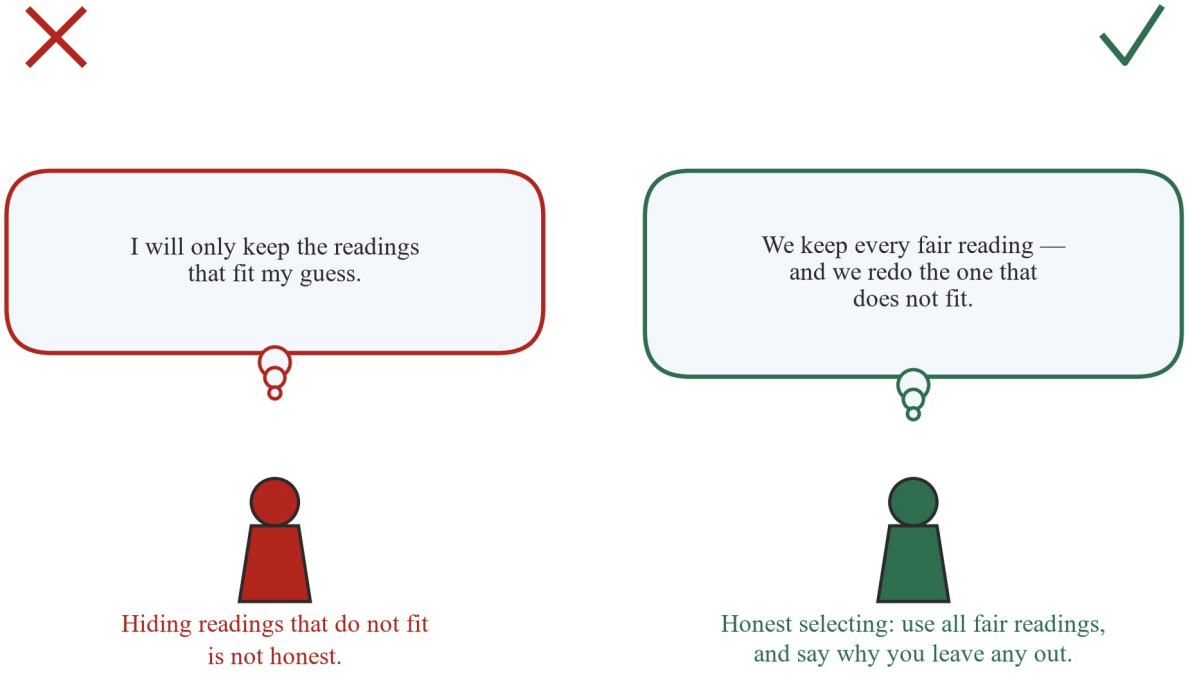
Two cards: a reading of 12 mL soaked up from 40 mL poured is marked believable; a reading of 80 mL soaked up from 40 mL poured is marked not believable

Data and evidence are different things

The **data** is everything we recorded — every reading, including the ones that went wrong. The **evidence** is the readings we *choose* to support our explanation. Choosing is normal and honest: we keep every fair reading, we redo (not hide) any reading that does not fit, and we say why a reading was left out.



Flow from a data box holding all six readings, through an arrow labelled select, to an evidence box holding the fair readings, through an arrow labelled supports, to a conclusion box saying Brand Y soaks up more water than Brand X

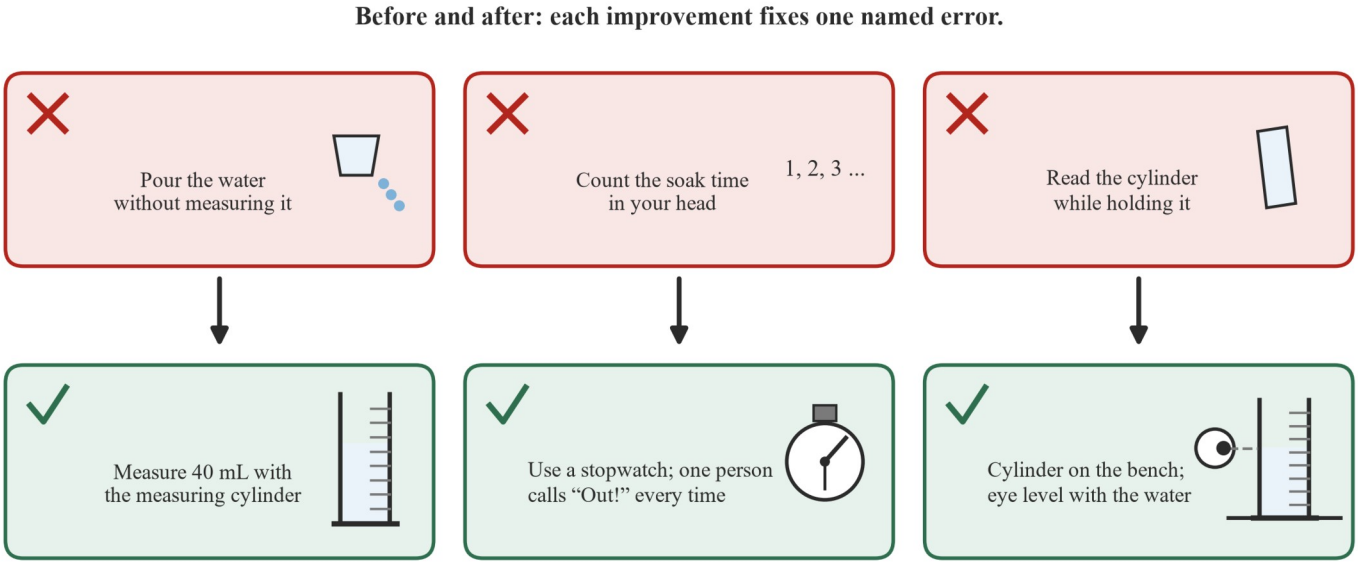


Two students with thought bubbles: one saying I will only keep the readings that fit my guess, marked wrong and not honest; one saying we keep every fair reading and redo the one that does not fit, marked right

Hiding readings that do not fit our guess is not selecting evidence. It is cheating the test.

Making the method better

When we **evaluate** a method, we suggest **improvements** — and a good improvement is specific. “Be more careful” fixes nothing. A good improvement names the error it fixes.



Three before-and-after pairs: pouring without measuring becomes measuring 40 mL with the cylinder; counting the soak time in your head becomes a stopwatch with one person calling out; reading the cylinder while holding it becomes the cylinder on the bench with the eye level with the water

Asking the next question

A finished test is the start of the next one. Looking back at the method and the findings always raises a **further** question — and the best next questions are testable, and change only one variable.

Worked examples

Example 1 — with support

Two groups tested which brand of paper towel soaks up the most water.

	Group A	Group B
Towel squares	10 cm × 10 cm, both brands	Brand X: 10 cm × 10 cm; Brand Y: 5 cm × 5 cm
Starting water	40 mL	40 mL
Soak time	5 s	5 s
Finding	Brand Y soaks up more	Brand X soaks up more

(a) Give one way the two methods were different. **(b)** Which group's finding about the brands can we trust? Explain why. **(c)** Suggest one improvement to Group B's method.

Working	Comment
(a) The size of the towel squares. Group A kept both squares at 10 cm × 10 cm; Group B used a 5 cm × 5 cm square for Brand Y.	*Comparing methods means looking for the difference, step by step.*
(b) Group A. It changed only the variable it was testing — the brand. Group B changed two variables, the brand and the size, so its finding could have been caused by the size.	*Name the error: more than one variable changed.*
(c) Use a 10 cm × 10 cm square for Brand Y, the same size as Brand X.	*A good improvement names the error it fixes.*

Example 2 — with support

One group's three trials for Brand Y gave: 18, 19 and 7 mL.

(a) Which reading does not fit? **(b)** What should the group do about it? **(c)** They did the trial again and got 18 mL. The readings are now 18, 19 and 18 mL. Do the readings now sit close together? What does that mean for how much we trust them? **(d)** A different group wrote down that a square soaked up 80 mL of water, but they had poured only 40 mL. Is that reading believable? Explain.

Working	Comment
(a) 7 mL. It sits far away from 18 and 19.	*Compare the readings; spot the one that does not fit.*
(b) Do that trial again, and check what went wrong. (They found they had squeezed the square over the beaker, so extra water dripped back.)	*A reading that does not fit gets redone — never hidden.*
(c) Yes — 18, 19 and 18 sit close together, so we can trust the result more.	*Close repeats mean more trust. No average needed.*
(d) No. The towel cannot soak up more water than was poured. The reading, or the method, went wrong.	*Ask it every time: is that reading believable?*

Example 3 — by yourself

(SCENARIO) A group asked: *Does the surface of a ramp change how long a marble keeps rolling?*

Their method: roll the marble from the **top** of a carpet ramp, then from **halfway up** a wooden ramp. On the carpet one person starts the timer when the marble is let go; on the wood a different person starts it when the marble reaches the bottom. They did **one trial** on each surface: carpet 3 s, wood 1 s. Their conclusion: the carpet stops the marble sooner.

(a) Find two errors in the method. *(2 marks)* **(b)** Suggest one improvement for each error. *(2 marks)* **(c)** Why is one trial on each surface not enough? *(1 mark)* **(d)** Write one next question for the group. *(1 mark)*

Practice questions

With support

Q1. *(3 marks)* Fill each gap with *error*, *evidence* or *conclusion*. **(a)** The readings we choose to support our explanation are the _____. **(b)** Reading the cylinder from above, so the reading looks too big, is an _____. **(c)** "Brand Y soaks up more water than Brand X" is a _____.

Q2. *(4 marks)* Match each mix-up to the error it is.

Mix-up	Error
1. One square soaks for 5 s, the next for 8 s.	A. changing more than one variable
2. A group does the test once and stops.	B. misreading a scale
3. The eyes are above the water when the cylinder is read.	C. not giving every trial the same time
4. One brand is tested warm, the other cold, *and* the brands differ.	D. doing too few trials

Q3. *(3 marks)* Group C soaked every square for 5 s. Group D soaked Brand X for 5 s and Brand Y for 8 s. Both groups say a different brand wins. **(a)** What is the difference between the two methods? **(b)** Which group's method is fair? **(c)** Why can a different soak time cause different findings?

Q4. *(3 marks)* Two sets of repeat readings for the same test: Set 1: 12, 13, 12 mL. Set 2: 9, 13, 17 mL. **(a)** Which set sits close together? **(b)** Which set do we trust more? **(c)** In Set 2, what should the group do next?

Q5. *(3 marks)* A group's notes say: (i) "The trials gave 12, 13 and 12 mL." (ii) "We choose the readings 12, 13 and 12 mL because they are the fair trials." (iii) "Brand X soaks up about 12 mL each time." Label each note *data*, *evidence* or *conclusion*.

Q6. *(2 marks)* A group finds Brand Y soaks up more water than Brand X. Circle the best next question and say why. A. Which brand is the cheapest? B. Does Brand Y soak up more soapy water too? C. Are paper towels good?

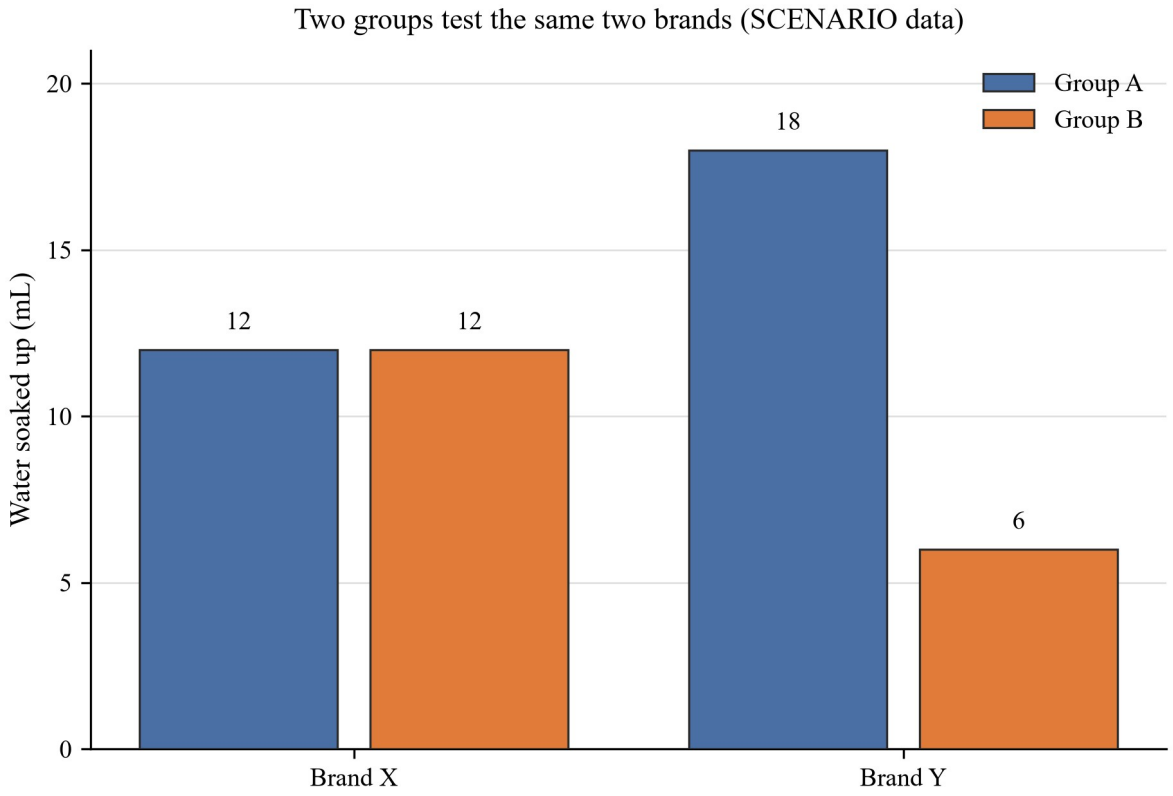
By yourself

Q7. *(4 marks)* Name the error in each mix-up. **(a)** A group reads the cylinder while holding it in one hand. **(b)** Squares of 10 cm × 10 cm for Brand X, 8 cm × 8 cm for Brand Y. **(c)** The timer is started when the marble is let go on one ramp, and when it reaches the bottom on the other. **(d)** A group gets 14 mL in its only trial and writes its conclusion.

Q8. *(3 marks)* A group tests which fertiliser makes bean plants grow taller. Pot 1 gets fertiliser A and sits in the sun. Pot 2 gets fertiliser B and sits in the shade. **(a)** Name the variable they meant to test. **(b)** Name the error in the method. **(c)** Suggest one improvement.

Q9. *(3 marks)* The water in a cylinder sits at 40 mL. Sam reads it with his eyes **above** the water level. **(a)** Will Sam's reading look too big, too small, or correct? **(b)** Which of these is most likely to be Sam's reading: 38 mL, 40 mL or 41 mL? **(c)** Where should Sam's eyes be for a correct reading?

Q10. *(4 marks)* The graph shows two groups' findings for the same two brands of paper towel.



Column graph: Group A found Brand X soaked up 12 mL and Brand Y 18 mL; Group B found Brand X 12 mL and Brand Y 6 mL

(a) Write the finding of each group. **(b)** The groups' methods were checked. Group B had used a 5 cm × 5 cm square for Brand Y. Explain how that error could cause Group B's finding for Brand Y.

Q11. *(3 marks)* A group's method says: 1. Pour some water into the beaker. 2. Count the soak time in your head. 3. Pick the cylinder up and read it. Rewrite each step so the error is fixed.

Q12. *(3 marks)* A class grew four bean seedlings by the window and measured their heights after two weeks: 8, 9, 8 and 3 cm. The teacher's notes say the 3 cm pot was knocked over early in week one and got no water for days. **(a)** Which reading would you leave out of the evidence? **(b)** Why is leaving it out honest, but hiding a reading we simply dislike is not?

Q13. *(3 marks)* A group's notes say a towel square soaked up 60 mL — but they had poured 40 mL into the beaker. **(a)** Is the reading believable? **(b)** Give one thing that could have gone wrong to cause it. **(c)** What should the group do now?

Q14. *(5 marks) (SCENARIO)* A group's results for two brands of paper towel:

Trial	Brand X (mL)	Brand Y (mL)
1	11	17
2	12	18
3	12	9

(a) Which Brand Y reading does not fit? **(b)** What should the group do about it? **(c)** The redone trial gives 18 mL. Do the Brand Y readings now sit close together? **(d)** Which brand soaks up more water? **(e)** Give one reason we can trust this finding.

Q15. *(3 marks)* Another group tests how the surface of a ramp changes how far a toy car rolls. Their ramp wobbles, they push the car by hand each time, and they measure the roll with their hands. Suggest one improvement for each problem, and say which error each one fixes.

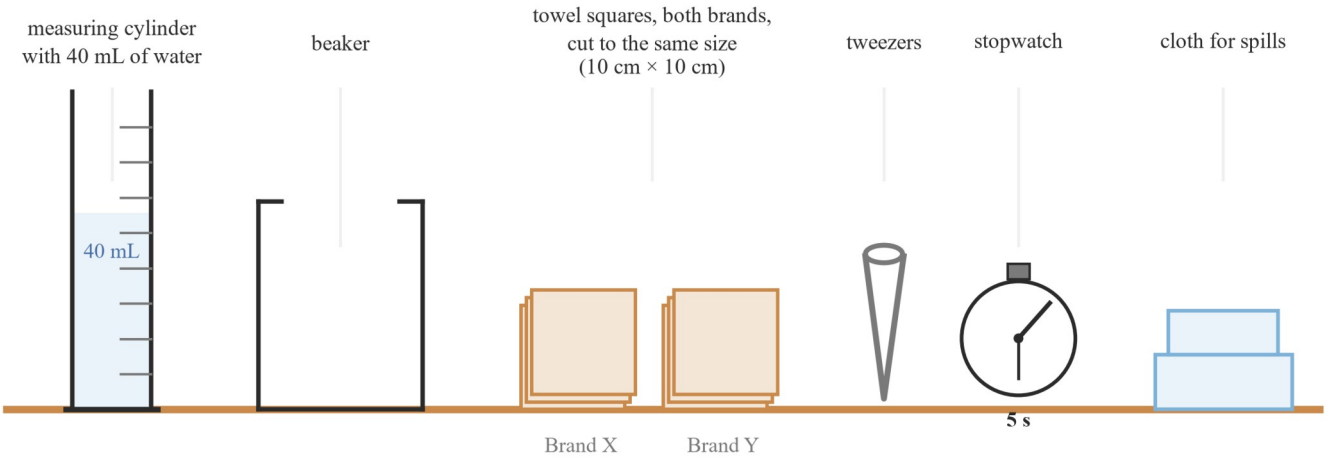
Q16. *(3 marks)* Your class has just found that Brand Y soaks up more water than Brand X. **(a)** Write one next question you could investigate. **(b)** For your question, name the variable you would change. **(c)** Name the variable you would measure.

Practical — Which paper towel soaks up the most water?

> PRACTICAL — this page plans or analyses work done in the classroom, laboratory or field. It is not a substitute for doing it.

Question: Which brand of paper towel soaks up the most water?

Equipment (see the bench picture below): a measuring cylinder, a beaker, towel squares of both brands cut to 10 cm × 10 cm, tweezers, a stopwatch, and a cloth for spills.



Everything on the bench before you start — including the cloth for spills.

A labelled bench: a measuring cylinder holding 40 mL of water, a beaker, two stacks of same-size towel squares labelled Brand X and Brand Y, tweezers, a stopwatch, and a cloth for spills

Safety

Risk	How we stay safe
Spilled water makes the bench and floor slippery.	Keep the cloth on the bench; wipe spills up straight away.
Wet towel squares drip on the way to the cylinder.	Carry squares with the tweezers, over the beaker.

Variables

Variable	Choice
The one we change	the brand of paper towel
The one we measure	the volume of water soaked up (mL)
The ones we keep the same	square size (10 cm × 10 cm); starting water (40 mL); soak time (5 s); drip time (10 s); the same beaker and cylinder

Method

1. Measure 40 mL of water into the beaker with the cylinder. Read it on the bench, eyes level with the water, at the bottom of the curved surface.
2. Lay one square of Brand X on the water. Start the stopwatch.
3. After 5 s, lift the square with the tweezers. Hold it over the beaker and let it drip for 10 s.
4. Pour the leftover water back into the cylinder. Read the volume, the correct way. Record it.
5. Volume soaked up = 40 mL − leftover volume. Record it.
6. Do steps 1–5 two more times for Brand X (three trials), then three times for Brand Y.

Results

Trial	Leftover water (mL)	Soaked up (mL)	Leftover water (mL)	Soaked up (mL)
	Brand X	*Brand X*	*Brand Y*	*Brand Y*
1				
2				
3				

Thinking about the results

- Do your three trials for each brand sit close together? If one reading does not fit, do that trial again.
- Is every reading believable? (Could a square soak up more than 40 mL?)
- Compare your findings with two other groups. If they differ, put your methods side by side and look for the difference.

Claim and defence

Our claim: Brand _____ soaks up the most water. The evidence we choose: _____ Why this evidence supports our claim: _____ One thing we kept the same to hold the test fair: _____ Our next question: _____

Answers

Example 3. (a) The release height changed (top vs halfway) *and* the surface changed — two variables; the timer was started at different moments (uneven timing). (b) Release from the top on both ramps; one person starts the timer the same way each time. (c) One trial could go wrong and we would never know; repeats let us check. (d) Any testable question changing one variable, e.g. *Does the height of the release change how long the marble rolls?*

Q1. (a) evidence (b) error (c) conclusion **Q2.** 1–C, 2–D, 3–B, 4–A **Q3.** (a) Group D gave different soak times (5 s vs 8 s); Group C kept them the same. (b) Group C. (c) A longer soak gives the towel more time to soak, so the extra time — not the brand — could cause the difference. **Q4.** (a) Set 1. (b) Set 1. (c) Look for what went wrong and do the trials again. **Q5.** (i) data (ii) evidence (iii) conclusion **Q6.** B — it is testable and changes one variable (the kind of water). **Q7.** (a) misreading a scale (b) changing more than one variable (c) not giving every trial the same timing (d) doing too few trials **Q8.** (a) the fertiliser (b) two variables changed — fertiliser *and* light (c) put both pots in the same light. **Q9.** (a) too big (b) 41 mL (c) level with the water, at the bottom of the curved surface. **Q10.** (a) Group A: Brand Y soaks up more (18 vs 12 mL). Group B: Brand X soaks up more (12 vs 6 mL). (b) A smaller square has less towel to soak, so Brand Y's soaked volume looks small — the size, not the brand, could cause the finding. **Q11.** 1. Measure 40 mL of water with the cylinder. 2. Time the soak with a stopwatch (one person calls out). 3. Put the cylinder on the bench and read it at eye level. **Q12.** (a) the 3 cm reading (b) the 3 cm pot was not a fair test of the seedlings' growth, and we say why we left it out; hiding a reading we dislike gives no reason and cheats the test. **Q13.** (a) No — 60 mL cannot soak up from 40 mL poured. (b) e.g. the leftover water was read from above, or some water spilled before the reading. (c) Find what went wrong and do the trial again. **Q14.** (a) 9 mL (b) redo that trial (and check for a slip, e.g. a squeeze or a spill) (c) yes — 17, 18, 18 sit close together (d) Brand Y (e) the trials sit close together, and the test changed only the brand. **Q15.** e.g. brace or hold the ramp still (fixes a wobbly, changing ramp); release the car without pushing, from a marked line (fixes changing the push each trial); measure with a ruler or metre stick in cm (fixes measuring with their hands, which are not the same for everyone). **Q16.** Any testable question that changes one variable, e.g. (a) *Does Brand Y soak up more soapy water too?* (b) change: the kind of water (plain vs soapy) (c) measure: the volume soaked up (mL).

Full solutions

Example 3. *(a) 1 mark* — error 1: the release point changed (top of the carpet ramp, halfway on the wood) as well as the surface, so two variables changed. *1 mark* — error 2: the timer was started at different moments (when let go vs when reaching the bottom), so the timing was not the same each trial. *(b) 1 mark* — let the marble go from the top of both ramps, from the same mark. *1 mark* — one person times it, starting the stopwatch the same way on every trial. *(c) 1 mark* — with one trial, a single slip would change the conclusion and nobody would know; repeats let us check the readings against each other. *(d) 1 mark* — e.g. *Does the height we release the marble from change how long it keeps rolling?*

Q1. *1 mark each.* (a) evidence — the readings we choose to support the explanation. (b) error — a mistake that makes the findings harder to trust. (c) conclusion — what we decide from the evidence.

Q2. *1 mark each.* 1–C: different soak times is uneven timing. 2–D: one go is too few trials. 3–B: eyes above the water misreads the scale. 4–A: two things changed at once.

Q3. *(a) 1 mark* — Group D soaked Brand X for 5 s and Brand Y for 8 s; Group C kept the soak time the same. *(b) 1 mark* — Group C (only the brand changes). *(c) 1 mark* — with more soak time the towel has more time to soak, so the time — not the brand — could cause the difference in the findings.

Q4. *(a) 1 mark* — Set 1 (12, 13, 12 sit close together). *(b) 1 mark* — Set 1; close repeats are easier to trust. *(c) 1 mark* — look for what went wrong and do the trials again; the spread (9 to 17) says something in the method kept changing.

Q5. *1 mark each.* (i) data — everything recorded. (ii) evidence — the readings chosen, with a reason. (iii) conclusion — what is decided from the evidence.

Q6. *1 mark* — B. *1 mark* — it can be tested and it changes one variable; A is not a science test of soaking and C cannot be tested as it stands.

Q7. *1 mark each.* (a) misreading a scale (holding the cylinder makes the water tilt). (b) changing more than one variable (brand and size). (c) not giving every trial the same timing. (d) doing too few trials.

Q8. *(a) 1 mark* — the fertiliser (A vs B). *(b) 1 mark* — the light changed with the fertiliser, so two variables changed. *(c) 1 mark* — put both pots in the same place, so only the fertiliser differs.

Q9. *(a) 1 mark* — too big. *(b) 1 mark* — 41 mL. *(c) 1 mark* — level with the water, reading the bottom of the curved surface, cylinder on the bench.

Q10. *(a) 2 marks* — Group A: Brand Y soaks up more (18 mL vs 12 mL); Group B: Brand X soaks up more (12 mL vs 6 mL). *(b) 2 marks* — the 5 cm × 5 cm square has less towel to soak up water, so Brand Y's soaked volume comes out small; the size error, not the brand, could cause Group B's finding.

Q11. *1 mark each.* Step 1 → measure 40 mL with the measuring cylinder (fixes pouring without measuring). Step 2 → use a stopwatch with one person calling the time (fixes counting in your head). Step 3 → cylinder on the bench, eyes level with the water, read the bottom of the curved surface (fixes reading it in the hand).

Q12. *(a) 1 mark* — the 3 cm seedling. *(b) 2 marks* — that pot was knocked over and missed water, so it is not a fair test of how tall these seedlings grow, and we *say* why it is left out (that is honest selecting); hiding a reading just because we dislike it gives no reason and keeps the evidence from matching the data.

Q13. *(a) 1 mark* — not believable: 60 mL cannot soak up from 40 mL poured. *(b) 1 mark* — e.g. the leftover water was read from above (reading looks too big → soaked-up volume looks too big) or water spilled before the leftover was read. *(c) 1 mark* — find the error and do the trial again.

Q14. *(a) 1 mark* — 9 mL (far from 17 and 18). *(b) 1 mark* — do that trial again, checking for slips (a squeeze over the beaker, a spill, a misread scale). *(c) 1 mark* — yes: 17, 18, 18 sit close together. *(d) 1 mark* — Brand Y (about 17–18 mL vs about 11–12 mL). *(e) 1 mark* — the repeats sit close together for both brands, and only the brand was changed, so the difference is caused by the brand.

Q15. *1 mark each.* Hold or brace the ramp so it does not wobble (fixes a ramp that changes between trials). Let the car go from a marked line without pushing (fixes a different push each trial). Measure the roll in cm with a ruler or metre stick (fixes measuring with their hands, which are not the same for everyone).

Q16. *(a) 1 mark* — any testable question, e.g. *Does Brand Y soak up more soapy water too? (b) 1 mark* — the variable changed: the kind of water (plain vs soapy). *(c) 1 mark* — the variable measured: the volume of water soaked up (mL).