

Science 9–10

Chapter 3 — Planning and conducting

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Attribution: Classbasics Pty Ltd, vwx.au

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.

Contents

Section	Page
Planning valid, reproducible investigations: methodology, controls, and sources of error and bias in sampling and observation	2
Safe and ethical investigation: risk assessment, animal and human ethics, and protocols for cultural sites and artefacts on Country and Place	19
Equipment, calibration, sample size and precision — generating and recording data with digital tools	42
Test A	58
Test B	62
Test A — Section A — Multiple choice	67
Test A — Section B — Short/structured response	69
Test B — Section A — Multiple choice	71
Test B — Section B — Short/structured response	73
Mark record — Test A	76
Mark record — Test B	77

Planning valid, reproducible investigations: methodology, controls, and sources of error and bias in sampling and observation

Chapter 3 — Planning and conducting (Science, Levels 9 and 10)

Subtopic 3A · VC2 Science — Planning and conducting

CD code: VC2S10I02 (split delivery †; this subtopic is the accountable owner) **Content description**

(verbatim): “valid, reproducible investigations to answer questions and test hypotheses can be planned and conducted, including identifying and controlling for possible sources of error and bias in sampling or in making observations; safe, ethical investigations include undertaking risk assessments and following protocols when accessing cultural sites and artefacts on Country and Place” **This subtopic’s limb (TOC §4 clause split):** valid, reproducible design — methodology, controls, and sources of error and bias in sampling and observation. The partner 3B delivers the other limb: safety, ethics, risk assessment, and protocols on Country and Place. Every plan made here hands over to 3B before fieldwork begins.

Elaboration ids drawn on: VC2S10I02.E03 · VC2S10I02.E05 · VC2S10I02.E06 (ids VC2S10I02.E01, VC2S10I02.E02 and VC2S10I02.E07 are drawn on by Subtopic 3B; VC2S10I02.E04 is withheld under the TOC’s D10 ICIP decision and is not taught here) **Achievement standard (verbatim):** “They plan a range of valid, reproducible and safe scientific investigations and explain how they have addressed any ethical and cultural considerations when generating or using primary and secondary data.” — the planning limb is served here; the safe, ethical and cultural limb is served with Subtopic 3B.

Theory

The plan is where the quality is decided. By the time results come in, most of an investigation’s quality has already been fixed — by the plan. A perfect graph of badly planned data is still bad evidence. This subtopic is about the plan: choosing a method that suits the question, building the comparison that makes the test fair, and keeping error and bias out of the two places they most easily sneak in — *sampling* (who or what you look at) and *observation* (how you record what you see). The safety and ethics side of planning — risk assessments, and protocols when accessing cultural sites and artefacts on Country and Place — is Subtopic 3B’s limb of the same content description, and it sits alongside everything taught here.

Three words this subtopic works to. The curriculum uses *valid*, *reproducible* and *bias* without fixing a single wording, so this book works to these definitions, and every question here can be answered from them:

- **valid** — an investigation is **valid** when it actually tests what it set out to test: the comparison is fair (only the factor being tested differs between the things compared), and the sampling and the observations genuinely answer the question asked.
- **reproducible** — an investigation is **reproducible** when someone else follows your written method and gets results that agree with yours.
- **bias** — **bias** is a push, built into the method, that shifts results one way. A mistake can push a result either way; bias pushes it the *same* way every time, so repeating a biased method does not fix it.

(The Victorian Curriculum uses these terms without a single fixed wording, so this subtopic works to the definitions above.)

Choosing a methodology (elaboration VC2S10I02.E03). The first planning decision is *how* the question will be answered — the **methodology**. Elaboration VC2S10I02.E03 names a range of scientific investigation methodologies: fieldwork, locating and using information sources, surveys, modelling and simulations, case studies, participant observations and interviews. The method is not chosen by what sounds impressive; it is chosen by what the question needs.

Choosing a methodology — match the method to the question

Fieldwork measuring or observing in a real place <i>"How many species of plant grow along the creek?"</i>	Survey asking many people the same questions <i>"How do students travel to school?"</i>	Secondary data using data others have collected, from information sources you can check <i>"What does the fossil record show about past life?"</i>
Modelling and simulation testing an idea with a representation you can change and rerun <i>"What happens to the population if the waterhole dries each summer?"</i>	Case study looking at one case in depth <i>"How did this one wetland recover after fencing?"</i>	Participant observation and interviews watching people and asking them questions <i>"How do visitors actually use the walking track?"</i>

A grid of six investigation methodologies — fieldwork, survey, secondary data, modelling and simulation, case study, and participant observation and interviews — each with a short description and an example question it suits

Each methodology suits a shape of question:

- **Fieldwork** — measuring or observing in a real place — suits questions about what is actually happening out in the world, like counting traffic or surveying plants along a creek.
- **Surveys** — asking many people the same questions — suit questions about a whole group when you can only ask some of them.
- **Secondary data** — using data others have collected, from information sources you can check — suits questions you cannot test yourself, like what the fossil record shows about past life.
- **Modelling and simulations** — testing an idea with a representation you can change and rerun — suit questions that are too slow, too big or too complex to test directly.
- **Case studies** — looking at one case in depth — suit questions about how one particular place or event behaved.
- **Participant observation and interviews** — watching people and asking them questions — suit questions about how people actually behave and why.

A question can earn more than one method, and a good plan often pairs two of them. What matters for validity is that the method *can* answer the question — a survey cannot measure what people do when nobody is asking them, and a single case study cannot speak for every case.

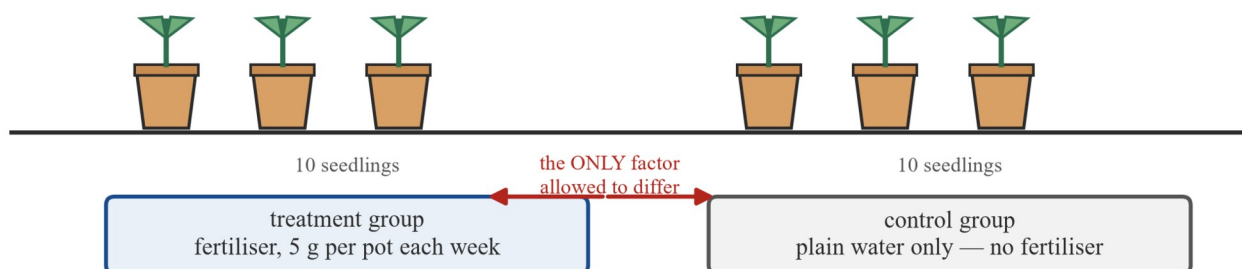
Variables and the fair test (skills from Subtopic 2A). Most experiments compare two or more setups. The factor you change on purpose is the **independent variable**; the thing you measure is the **dependent variable**; and everything you keep the same is a **controlled variable**. A test is fair — and the comparison valid — only when one factor differs and everything else that could matter is controlled. If two factors differ, and the result changes, you cannot say which factor caused it.

A fair test: one factor differs — everything else is kept the same

controlled variables — kept the same for every pot, both groups:



dependent variable: growth in height, measured after 3 weeks

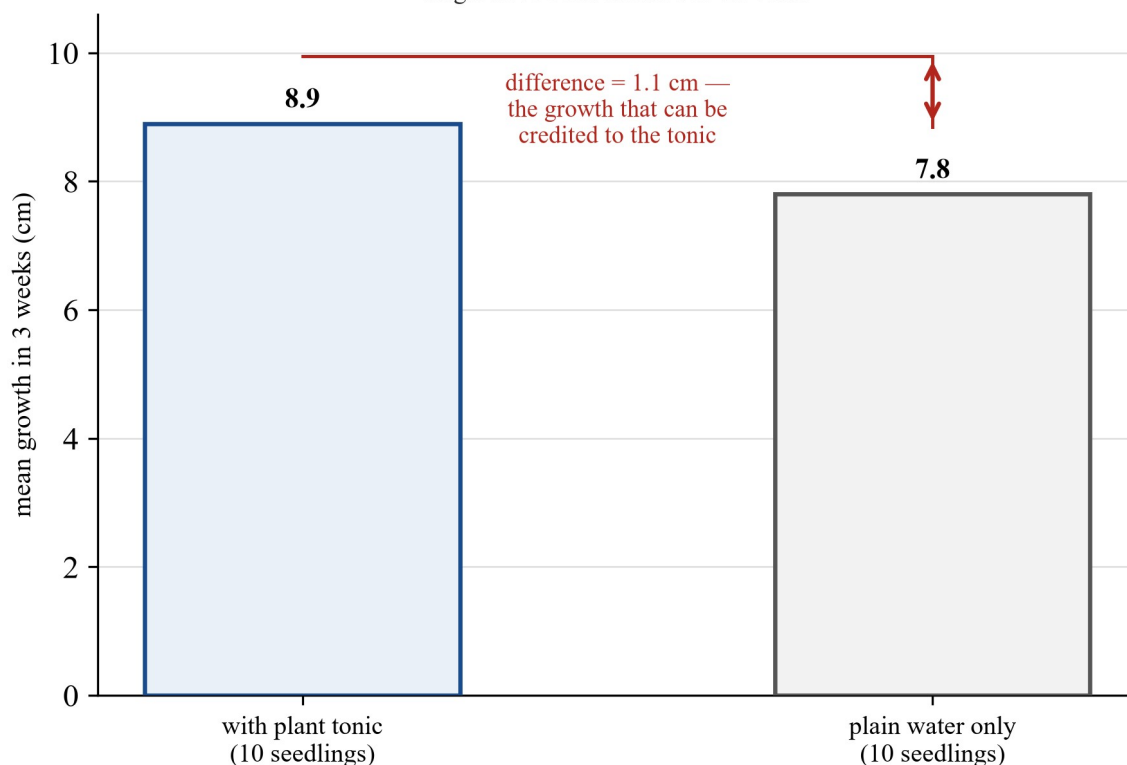


A fair test with two groups of potted seedlings on a bench: a treatment group given fertiliser and a control group given plain water, with callout boxes listing the controlled variables kept the same for every pot, and the dependent variable labelled as growth in height after 3 weeks

In the figure, the only factor allowed to differ is the fertiliser. Every pot gets the same size pot, the same soil, the same volume of water, the same light, the same bean variety and the same starting height, and each group starts with 10 seedlings. If any one of those controlled variables were left uncontrolled — say, one group sat on a sunnier bench — the test would no longer be valid, because a difference in growth could be blamed on the light just as easily as on the fertiliser.

Control groups: what would have happened anyway? Living things grow, heal and change with no help at all, so an experiment on a treatment needs something to compare it with — a **control group**, treated the same way in every respect except for the factor being tested. The control group shows what would have happened *anyway*.

Simulated data for practice: without the control group, all 8.9 cm might have been credited to the tonic



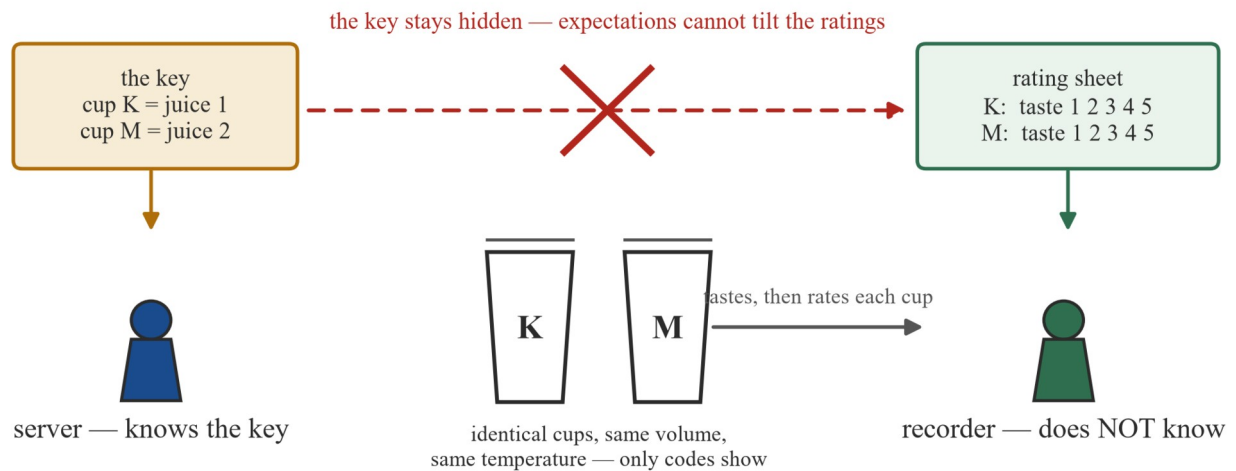
Bar chart of simulated data for practice: seedlings given a plant tonic grew a mean of 8.9 cm in 3 weeks while the control group given plain water grew a mean of 7.8 cm, so the difference that can be credited to the tonic is 1.1 cm

The simulated data in the figure makes the point exactly. The tonic group grew a mean of 8.9 cm. Without a control group, all 8.9 cm might have been credited to the tonic. The control group — plain water only, everything else identical — grew 7.8 cm anyway. So the growth that can actually be credited to the tonic is the difference: $8.9 - 7.8 = 1.1$ cm. The control group turns “it grew” into “it grew *more than it would have anyway*”.

Placebos and blind tests. When the thing being tested is taken by people, expectations do part of the work: people who *believe* they are getting a treatment often report feeling better. A fair comparison therefore gives the control group a **placebo** — a look-alike treatment with no active ingredient, so both groups believe they might be getting the real thing. (In this book, *placebo* means a pretend treatment, identical in looks to the real one but with no active ingredient, given to the control group so that expectations are the same in both groups.)

Expectations can tilt the measuring side too. If the person recording the results knows which participant got the real treatment, they may — without any dishonesty — hear or see what they expect. A **blind test** hides the key: the cups, tablets or samples carry only codes, and the person recording the results does not know which code is which.

A blind test: the person recording the results does not know which is which

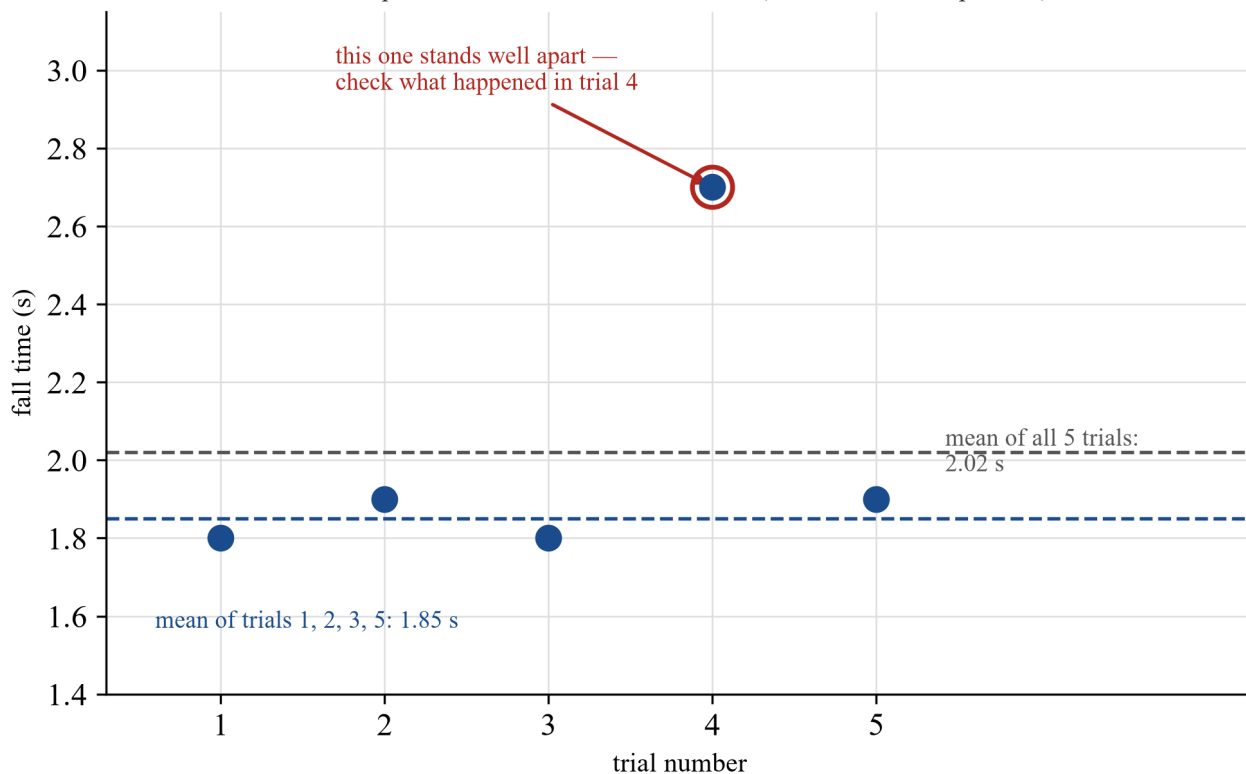


A blind taste test: a server who knows the key hands out identical cups labelled only K and M, while the recorder rates each cup on a sheet without knowing which juice is which, so expectations cannot tilt the ratings

In the figure, the server knows that cup K is juice 1 and cup M is juice 2, but the recorder does not — the key stays hidden until every rating is written down. The cups are identical, the volumes the same and the temperatures the same, so the only thing left to differ is the taste itself. The same idea protects the plant-tonic test: the person who measures the seedlings should not know which pots got the tonic.

Reproducibility, part one: repeat trials. No single measurement is trusted. Any one reading can be pushed by a small mistake — a late thumb on the stopwatch, a misread scale. So a good plan **repeats the same measurement several times** under the same conditions, and the results are compared before anything is concluded.

Five repeat trials of the same measurement (simulated data for practice)

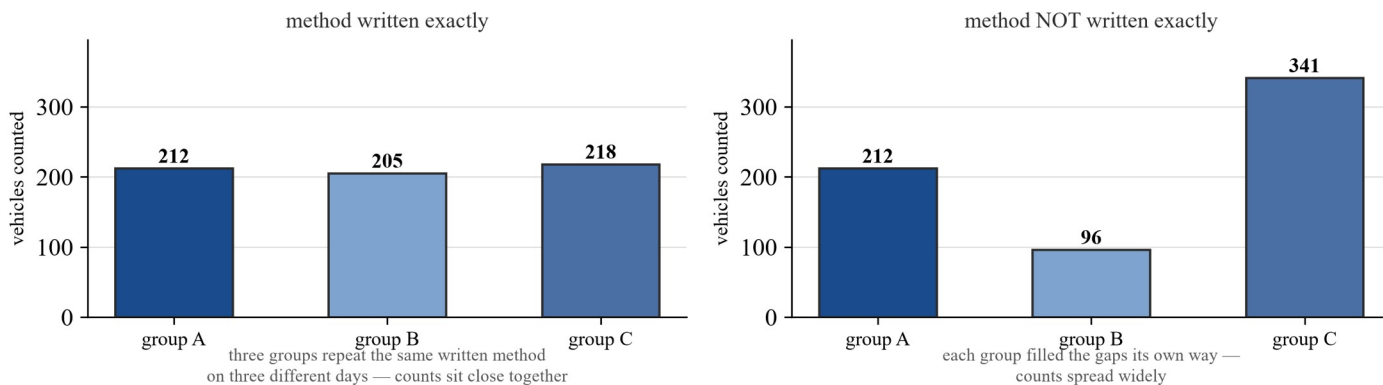


Scatter plot of five repeat trials of the same fall-time measurement, simulated data for practice: trials 1, 2, 3 and 5 sit close together between 1.8 and 1.9 seconds while trial 4 stands apart at 2.7 s; the mean of all 5 trials is 2.02 s and the mean of trials 1, 2, 3 and 5 is 1.85 s

The five trials in the figure are five drops of the same paper cone, timed by the same observer: 1.8, 1.9, 1.8, 2.7 and 1.9 s. Four of the five sit close together; trial 4 stands well apart. That apartness is information, not a nuisance — it says *check what happened in trial 4*. Perhaps the cone hit the desk on the way down, or the stopwatch was started late. The mean of all five trials is 2.02 s; the mean of trials 1, 2, 3 and 5 is 1.85 s. You never silently delete a result that looks wrong — you check what happened, and you only set it aside if you find a reason. Repeat trials like these catch one-off mistakes; they cannot catch bias, because a biased method pushes every trial the same way.

Reproducibility, part two: a method someone else can follow. Reproducibility is tested by a *different* person repeating your investigation from what you wrote down. That only works if the method is written in enough detail that no gap is left for the other person to fill in their own way.

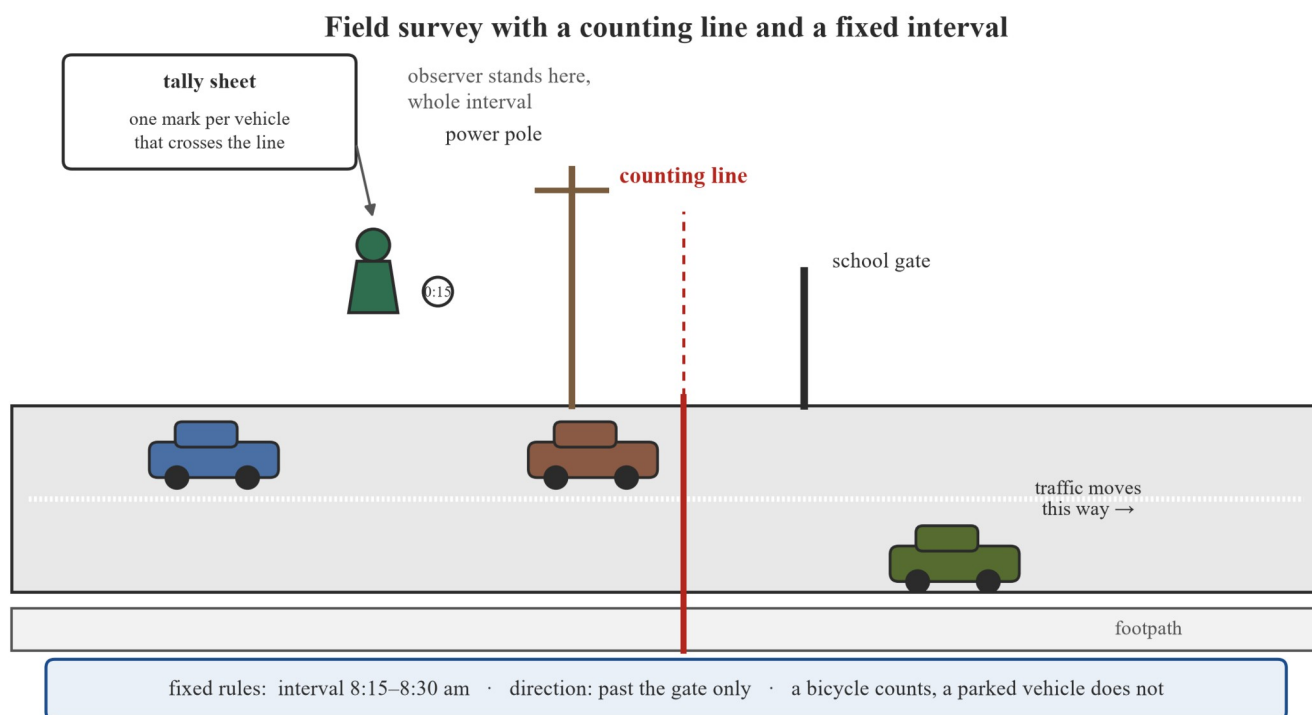
Reproducibility check on the traffic survey — simulated data for practice



Two bar charts of simulated traffic counts for practice: three groups repeating the same exactly written method counted 212, 205 and 218 vehicles, sitting close together, while three groups following a method that was not written exactly counted 212, 96 and 341, spread widely

The simulated counts in the figure show what a gap costs. Three groups followed the *same exactly written* method on three different days and counted 212, 205 and 218 vehicles — close together, because every decision had been made for them in advance. Three other groups followed a vague method (“count the traffic outside school in the morning”) and counted 212, 96 and 341. Nothing went “wrong” for them — each group simply filled the gaps its own way: a different interval, a different direction, a different rule for bicycles. The spread of 245 between the highest and lowest count is not the traffic changing; it is the method changing. Reproducibility is therefore a property of the *writing* as much as the science: if a stranger could not follow your method without asking you a single question, it is not finished.

A worked planning case: the traffic survey (elaboration VC2S10I02.E06). Elaboration VC2S10I02.E06 asks you to determine the reproducibility of a field investigation using survey techniques that answer a question such as “How much traffic passes the school during a designated period of time?” The figure shows a plan built to be reproducible.



A labelled street scene for the traffic survey: an observer with a tally sheet and a 15-minute timer stands in one spot for the whole interval, a red counting line runs between the power pole and the school gate, and a strip of fixed rules reads interval 8:15 to 8:30 am, direction past the gate only, a bicycle counts and a parked vehicle does not

Every choice that could otherwise be filled in differently is fixed by a written rule:

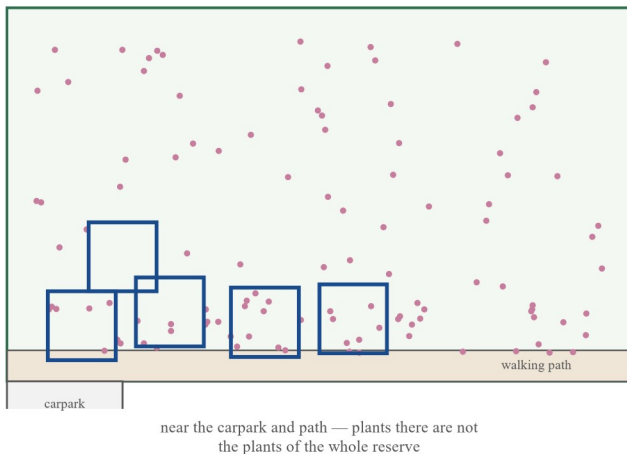
- **the interval** — 8:15–8:30 am, not “the morning”;
- **the counting line** — the imaginary line between the gate post and the power pole; a vehicle counts when its front crosses that line;
- **the direction** — traffic travelling past the gate only, not the other side of the road;
- **what counts as a vehicle** — a bicycle counts; a parked vehicle does not;
- **the observer** — stands in the one spot, marked on the plan, for the whole interval.

With the rules written, a second group can repeat the survey next week and compare counts meaningfully. Without them, “we counted the traffic” is the vague method of the previous figure, and the counts will disagree for reasons that have nothing to do with traffic. The same discipline — fixed interval, fixed line, fixed rules for what counts — carries over to any field count, from vehicles to birds at a feeder to plants in a reserve. (Standing safely, and getting any needed permissions, is the 3B side of this plan.)

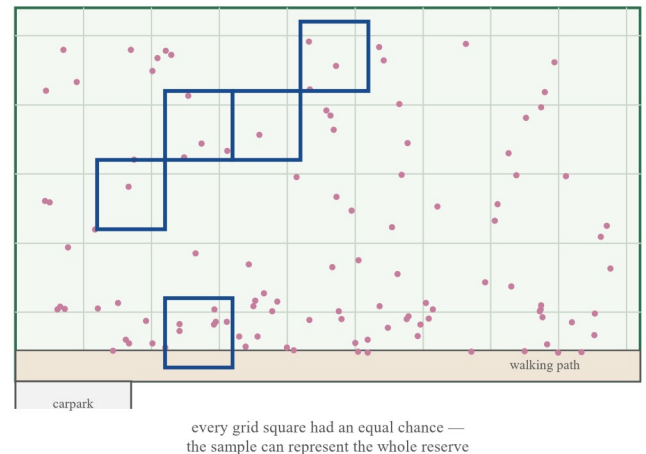
Sampling: does what you look at stand for the whole? Often you cannot look at everything — every plant in a reserve, every person in a school — so you look at a **sample**, a part chosen to stand for the **population**, the whole group the question is about. (In this book, *population* means the whole group a question is about, and *sample* means

the part of it that is actually looked at.) A sample does its job only if it is **representative** — if it resembles the population rather than some special corner of it.

(a) quadrats placed where it was easy



(b) quadrats placed at random grid squares

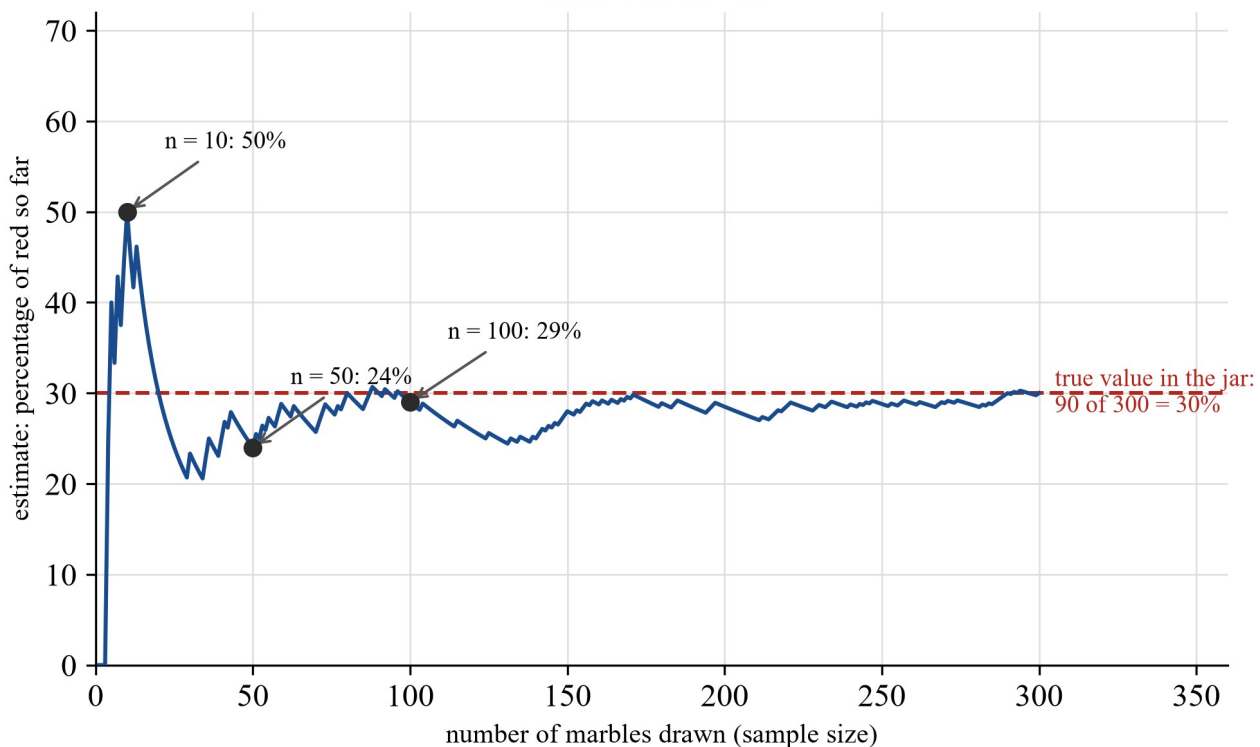


Two maps of the same plant reserve with flowers dotted across it: in panel a, five quadrats placed where it was easy, clustered near the carpark and walking path; in panel b, five quadrats placed at randomly chosen grid squares spread across the whole reserve

The figure shows the same reserve surveyed two ways, using a **quadrat** — a square frame of known size, placed on the ground, inside which everything of interest is counted. (In this book, *quadrat* means a square frame of known size used to sample plants or slow-moving animals in the area it covers.) In panel (a) the quadrats were placed where it was easy — beside the carpark and the walking path. That is a **convenience sample**, chosen for ease, and it is biased: plants that like the disturbed ground near a path are counted too often, and plants of the open reserve almost never. In panel (b) the reserve is divided into a grid and the squares are chosen so that every square has an equal chance — here the random squares chosen were (2,3), (3,1), (3,4), (4,4) and (5,5). Some path edge and some open reserve both get their fair share of quadrats, so the sample can represent the whole reserve. Convenience sampling is bias in the *sampling* half of the content description; random choice of sampling points is the standard fix.

Sample size: bigger settles, but bigger does not un-bias. A second fact about sampling is that small samples jump around. The simulated marble jar in the next figure holds 300 marbles, 90 of them red — a true value of 30%.

Drawing marbles without looking (simulated): bigger samples settle closer to the true value



Line graph of simulated marble draws for practice: the running estimate of the percentage of red marbles swings wildly at first — 50% after 10 draws, 24% after 50 — then settles closer to the true value of 30% as the sample grows, reading 29% at 100 draws and 30% at 300

Drawing without looking, the estimate after 10 marbles is 50% red — wildly off the truth. After 50 draws it is 24%; after 100, 29%; and by the time all 300 are drawn it sits at the true 30%. The estimate settles closer to the true value as the sample grows, which is why a good plan says *how many*, not just *where*: ten quadrats, ten survey replies or ten trials give a shakier answer than a well-chosen larger sample. But notice what size does **not** do: if the marbles were drawn only from the top of the jar, and red marbles had settled to the bottom, even 300 draws would not give the truth. **A bigger sample settles a fair method; it never fixes a biased one.** Bias is fixed by *how* the sample is chosen, size fixes only how much the estimate jumps around.

Error and bias in making observations. The second place the content description names is *making observations*. Observers are not cameras. We see what we expect to see, we tire, and we use vague words that mean different things to different people. The free notes on the left of the next figure show all three failures; the structured sheet on the right shows the fixes.

Two ways to record behaviour — only one of them can be repeated or checked

free notes

8:02 lots of birds at the new feeder

8:05 they seem more aggressive here

8:09 one chased another; I think

8:12 a peck? hard to tell

8:15 had to leave — enough?

“aggressive” is never defined — a second observer would record something different

structured tally sheet

Bird behaviour tally — feeder study

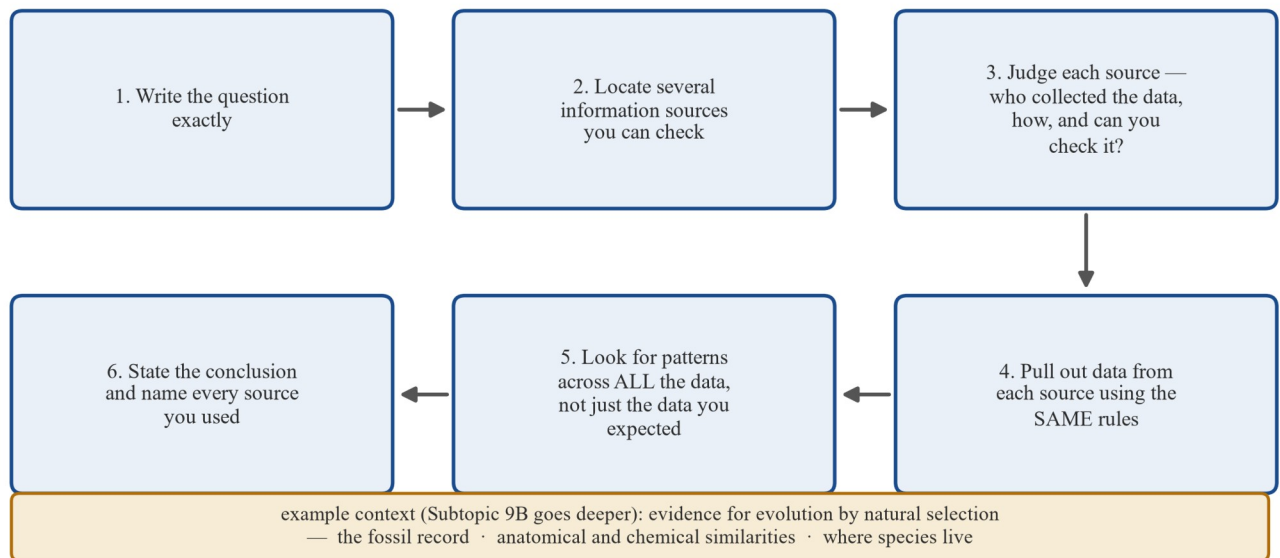
behaviour	what counts (operational definition)
peck	beak touches or strikes another bird
chase	flying after another bird
call	call directed at another bird
interval	tally in 30-second blocks
time	8:00–8:10 am, every visit
distance	observe from 5 m, behind the fence
any observer, any day, records the same behaviours the same way	

Two recording sheets side by side: on the left, free notes full of vague words like *aggressive* and *hard to tell* that a second observer could not repeat; on the right, a structured tally sheet that gives each bird behaviour an operational definition saying exactly what counts

The free notes say “they seem more aggressive here” — but *aggressive* is never defined, so a second observer would record something different, and even the same observer would decide differently on a different day. The structured sheet replaces the vague word with an **operational definition** — a definition written as a rule anyone can follow: a *peck* counts when a beak touches or strikes another bird; a *chase* counts when one bird flies after another. (In this book, *operational definition* means a definition of a behaviour or measurement written as a rule that any observer can follow, so that two observers record the same thing the same way.) The sheet also fixes the interval (30-second blocks), the time (8:00–8:10 am, every visit) and the distance (5 m, behind the fence). With those rules, any observer on any day records the same behaviours the same way — the observation side of reproducibility. Where expectations are the worry, the observer side borrows the blind test’s trick: if the recorder does not know which group is which, expectations have nothing to push on.

Secondary data: using evidence you did not collect yourself (elaboration VC2S10I02 .E05). Not every investigation generates its own data. **Primary data** is data you collect yourself; **secondary data** is data someone else collected, which you locate from information sources and re-use for your question. (In this book, *secondary data* means data collected by someone else, used from a source you can check.) Elaboration VC2S10I02 .E05 gives the classic case: using secondary data to analyse evidence for the theory of evolution by natural selection — the fossil record, chemical and anatomical similarities between living things, and the geographical distribution of species. No school can dig up the fossil record, so the investigation is planned around *locating, judging and using* other people’s data carefully.

An investigation using secondary data — data someone else collected

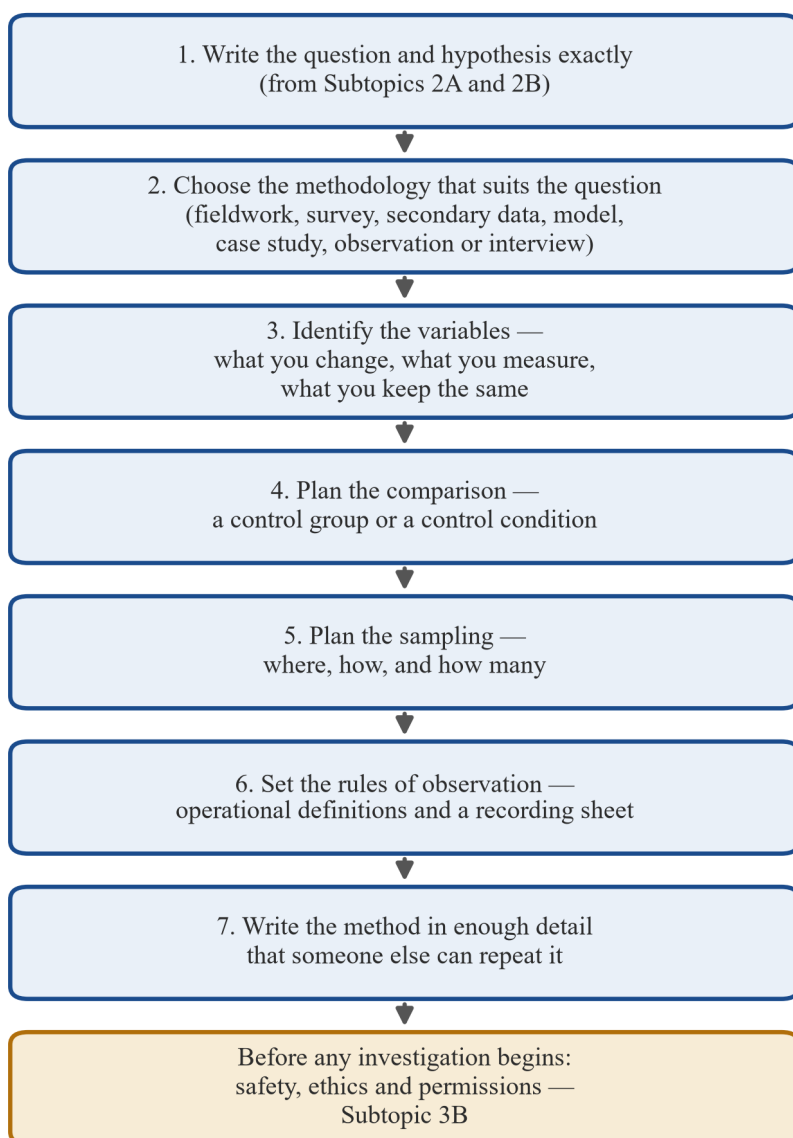


A six-step flow chart for an investigation using secondary data: write the question exactly; locate several information sources you can check; judge each source by who collected the data, how, and whether you can check it; pull out data from each source using the same rules; look for patterns across all the data, not just the data you expected; state the conclusion and name every source used

The method in the figure is the fair-test idea translated to data you did not collect: pull data from every source using the *same* rules (the control on extraction), and look for patterns across *all* the data, not just the data you expected (the control on your own expectations). The checks that matter are the checks you cannot do on your own measuring — who collected the data, how, and whether the source can be checked against others. (Subtopic 9B goes deeper into the evolution evidence itself; here the point is the method.) A secondary-data investigation is reproducible in its own way: someone else should be able to follow your steps to the same sources and pull out the same numbers using your rules.

The whole plan in one sequence. Every decision above sits in one planning order, shown in the final figure: from the exact question and hypothesis (skills from Subtopics 2A and 2B), through methodology, variables, comparison, sampling and observation rules, to a method written in enough detail that someone else can repeat it.

Planning a valid, reproducible investigation



A vertical flow chart of the planning sequence: write the question and hypothesis exactly; choose the methodology; identify the variables; plan the comparison with a control group or condition; plan the sampling of where, how and how many; set the rules of observation with operational definitions and a recording sheet; write the method in enough detail to repeat — with an amber note that safety, ethics and permissions from Subtopic 3B come before any investigation begins

The amber box is the handover: before any investigation begins, the safety, ethics and permissions side is checked — risk assessments and protocols, including on Country and Place, are Subtopic 3B's limb of this same content description, and no plan is finished without it. Likewise the plan hands measurements over to Subtopic 3C for the right equipment and units, and the finished results hand over to Subtopic 5A, where validity and reproducibility are *evaluated* after the data is in. This subtopic's job is the plan itself — made valid, made reproducible, and written down.

Activities

Activity 1 — Reproducibility of the traffic survey (elaboration VC2S10I02 . E06). In groups of three, plan and run the school-gate traffic survey of the figure, twice. First run: each group writes its *own* method in five lines or fewer and counts for 15 minutes exactly as it wrote it. Second run (a different day): all groups follow *one* shared method with the fixed rules from the figure — interval, counting line, direction, what counts, where the observer stands. Tabulate the three counts from each run (tabulating is a skill from Subtopic 4A) and compute the spread

(highest minus lowest) of each run. Which run is reproducible, and which planning decisions made the difference? (Any real survey outside the gate waits on 3B's safety and permissions checks; until then, run the activity on simulated data for practice — for example, a teacher-called stream of “vehicle” cards.)

Activity 2 — The sampling game: convenience versus random. Using a printed grid map of a “reserve” (a sheet of squared paper with dots scattered as plants, like the sampling figure), place five quadrats (a square cut to one grid square) where it is easiest to reach from the bottom edge, and count the plants inside each. Then choose five squares by random numbers — draw column and row numbers from a bag — place the quadrats there, and count again. Compare the two mean counts with the class's combined data, which stands for the true population. Which method's means sit closer to the true value across the class, and why did the convenience sample misjudge the reserve?

Activity 3 — A blind taste test (control, placebo and blind in one plan). With the teacher's check that no participant has an allergy or dietary restriction, run a blind taste test of two similar juices. A server, out of sight, pours identical volumes into identical cups coded K and M and keeps the key; a recorder who does not know the key rates each cup from 1 to 5 for sweetness. The whole class's ratings are the treatment data. Then discuss: (i) what a placebo version of this test would look like if one “juice” were plain coloured water claimed to be a sports drink; (ii) which variables the plan controls (cup, volume, temperature, codes); (iii) what the blind hides from whom, and which bias it removes. Record the class plan as a numbered method that another class could repeat.

Worked examples

Example 1 — scaffolded

A group plans to test whether a plant tonic increases seedling growth. They propose: “put ten seedlings on the window sill, give them the tonic, and measure them after 3 weeks.” (a) Name the independent and dependent variables the plan needs. (b) Add the comparison the plan is missing, and say what it shows. (c) List three controlled variables, and say what goes wrong if one of them is left uncontrolled.

Working	Comment
Independent variable: whether the seedling gets the tonic or not; dependent variable: growth in height over 3 weeks	The changed factor and the measured thing — name both exactly.
A control group: ten more seedlings, treated the same in every way except plain water instead of tonic	A treatment alone proves nothing, because seedlings grow anyway.
The control group shows the growth that would have happened anyway; only growth above it can be credited to the tonic	This is the whole reason for the control.
Controlled variables (any three): same pot size, same soil, same volume of water, same light, same bean variety, same starting height	Each one is a factor that could also change growth.
If, say, light were uncontrolled (tonic group on the sill, control in the shade), a taller tonic group could be blamed on the light — the test would not be valid	Name the factor, then the wrong conclusion it allows.

Example 2 — scaffolded

Two groups run the school-gate traffic survey on the same morning, one week apart. Group A counts 212 vehicles; group B counts 84. Group A's method is the exactly written method of the figure. Group B's method says: “we counted the cars outside school after first bell.” (a) Give two ways group B's method differs from group A's fixed rules. (b) Explain why the two counts cannot be compared as evidence about traffic. (c) Rewrite group B's method so that a third group could reproduce group A's count.

Working	Comment
Difference 1: group B counted cars only — the fixed rules count bicycles too; difference 2: “after first bell” is not the fixed 8:15–8:30 am interval, and no counting	Read each gap against the five fixed rules.

line or direction is set

The counts disagree because the *methods* disagree, not because the traffic changed — the investigation is not reproducible

Reproducibility fails when gaps are filled in differently.

Rewritten: stand at the marked spot from 8:15 to 8:30 am; count every vehicle whose front crosses the line between the gate post and the power pole, travelling past the gate only; a bicycle counts, a parked vehicle does not; one tally mark per vehicle

A finished method leaves a stranger no question to ask.

Example 3 — unscaffolded

A student wants to know what plants grow in the bushland reserve behind the school. Their plan: “walk in from the carpark, and put the quadrat down wherever the plants look interesting, five times.” Explain why this sample would misjudge the reserve, name the kind of sampling it is, and rewrite the placement rule so the sample can represent the whole reserve.

The plan is a **convenience sample** — quadrats placed where it is easy or where plants catch the eye. It misjudges the reserve because the plants near the carpark and track are not the plants of the whole reserve: disturbed, trampled ground selects for a few tough species, and choosing “where plants look interesting” counts showy patches too often and sparse ground not at all. The bias is built into *where* the sample is taken, so repeating the same walk would repeat the same bias. The fix is to let chance choose: divide the reserve into a grid, number the squares, and draw five pairs of numbers at random; the quadrat goes exactly on the squares drawn — even the bare-looking ones. Every square then has an equal chance, so the sample can represent the whole reserve.

Example 4 — unscaffolded

A class studies behaviour at a new bird feeder. One observer’s sheet reads: “lots of birds, a bit aggressive today, one bullied another, stayed a while.” A second observer, at the feeder at the same time next week, records almost nothing. Explain, using the ideas of this subtopic, why the two records disagree, and redesign the observation side of the plan so that any observer on any day records the same behaviours the same way.

The records disagree because nothing in the first sheet is *operational*: “lots”, “aggressive” and “a while” are vague words that each observer fills in differently, and “aggressive” invites expectation to do the counting — an observer who expects bullying at a new feeder will see it. The redesign fixes interval, distance and definitions: observe 8:00–8:10 am on every visit, from 5 m behind the fence; use a structured tally sheet in which each behaviour has an operational definition (a *peck* counts when a beak touches or strikes another bird; a *chase* counts when one bird flies after another; a *call* counts when directed at another bird), tallied in 30-second blocks. With rules like these, the second observer’s sheet becomes comparable to the first — and if the feeder itself is ever tested against a no-feeder control, the observer should not know which week is which, so expectation has nothing to push on.

Practice questions

Question 1 (scaffolded) In your own words, give the meaning this book works to for: (a) valid; (b) reproducible; (c) bias. For (c), also say how bias differs from a one-off mistake.

Question 2 (scaffolded) Match each question to the methodology that suits it best: (a) How many species of plant grow along the creek? (b) How do students in our school travel to school? (c) What does the fossil record show about past life? (d) What happens to a town’s water use if a year of drought repeats? (e) How did this one wetland recover after fencing? (f) How do visitors actually use the walking track?

Question 3 (scaffolded) In the seedling fair test of the figure: (a) name the independent variable; (b) name the dependent variable; (c) list two controlled variables.

Question 4 (scaffolded) In the plant-ionic figure: (a) what does the control group show? (b) what single number is the growth that can be credited to the ionic, and how is it found?

Question 5 (scaffolded) In the repeat-trials figure: (a) which trial stands well apart? (b) state the mean of all five trials and the mean without the apart trial; (c) what should be done about the apart trial before any conclusion?

Question 6 (scaffolded) In the sampling figure: (a) why do the quadrats near the carpark and path give a biased picture of the reserve? (b) what does choosing squares so every grid square has an equal chance fix?

Question 7 A survey asks “how do students travel to school?” but is run only at the bicycle shed, between 8:25 and 8:30 am. Name the sampling problem, say who is missed, and say whether asking *more* students at the bike shed would fix it.

Question 8 (a) Explain why the recorder in the blind-test figure must not know which cup is which. (b) In a medicine test, what does the control group receive instead of the medicine, and why?

Question 9 The marble-jar figure shows estimates of 50% at 10 draws, 24% at 50, 29% at 100 and 30% at 300, against a true value of 30%. (a) Describe what happens to the estimate as the sample grows. (b) If red marbles had settled to the bottom of the jar and draws were taken from the top only, would drawing all 300 marbles give the true value? Explain what this says about sample size and bias.

Question 10 The free-notes observer wrote “they seem more aggressive here”. Rewrite *aggressive* for a bird-feeder study as an operational definition a second observer could follow.

Question 11 A group claims “the creek is cleaner upstream” using secondary data from two websites. (a) Give two checks the group should make on each source before trusting it. (b) Why is this a secondary-data investigation rather than fieldwork, and when would fieldwork be the better methodology?

Question 12 Two groups repeat the traffic survey from the *same* exactly written method and count 212 and 205. A third group, using a method it wrote itself, counts 341. Does the 341 show the method is not reproducible? Explain.

Question 13 A student plans to test a new sports drink: “ten teammates drink it for a month, and I time their 1 km runs before and after.” Give two reasons any improvement could not be credited to the drink, and add one planning feature that fixes each.

Question 14 Classify each planning fix as one that mainly tackles **bias** or one that mainly tackles **one-off mistakes and spread**: (a) repeating a measurement five times and using the mean; (b) choosing quadrat squares by random numbers instead of near the path; (c) a recorder who does not know which cup holds which juice; (d) a bigger sample from a fair method; (e) fixed rules for what counts as a vehicle.

Question 15 Explain why “a bigger sample fixes a biased method” is false, using the marble-jar idea of red marbles settled to the bottom.

Question 16 A case study follows one wetland for a year after fencing. (a) What does the case-study methodology do well? (b) What can it *not* claim about wetlands in general, and which other methodology could begin to test that?

Answers

Question 1 (a) valid — the investigation tests what it set out to test: a fair comparison and sampling/observations that answer the question; (b) reproducible — someone else following the written method gets results that agree; (c) bias — a push built into the method that shifts results one way; a mistake pushes either way, bias the same way every time.

Question 2 (a) fieldwork; (b) survey; (c) secondary data; (d) modelling and simulation; (e) case study; (f) participant observation and interviews. **Question 3** (a) fertiliser or not; (b) growth in height after 3 weeks; (c) any two of pot size, soil, volume of water, light, bean variety, starting height. **Question 4** (a) the growth that would have happened anyway; (b) 1.1 cm — 8.9 minus 7.8 cm. **Question 5** (a) trial 4 (2.7 s); (b) 2.02 s and 1.85 s; (c) check what happened in trial 4; set it aside only with a reason. **Question 6** (a) plants near the carpark and path are not the plants of the whole reserve — ease, not chance, chose them; (b) it lets every square an equal chance, so the sample can represent the whole reserve. **Question 7** see solution — convenience sample; walkers, cars and buses and early arrivals missed; more of the same biased sample does not fix it. **Question 8** see solution. **Question 9** (a) it settles closer to the true value; (b) see solution. **Question 10** any rule a second observer could follow — see solution. **Question 11** see solution. **Question 12** No — see solution. **Question 13** see solution. **Question 14** (a) spread; (b) bias; (c) bias; (d) spread; (e) bias. **Question 15** see solution. **Question 16** see solution.

Fully-worked solutions

Question 1 (a) An investigation is **valid** when it actually tests what it set out to test — the comparison is fair (only the factor being tested differs) and the sampling and the observations genuinely answer the question asked. (b) An investigation is **reproducible** when someone else follows the written method and gets results that agree with yours. (c) **Bias** is a push, built into the method, that shifts results one way. A one-off mistake can push a result either way and is smoothed by repeats; bias pushes the same way every time, so repeating the biased method repeats the bias.

Question 2 (a) **Fieldwork** — measuring or observing in a real place. (b) **Survey** — asking many people the same questions. (c) **Secondary data** — using data others collected, from sources you can check. (d) **Modelling and simulation** — a representation you can change and rerun. (e) **Case study** — one case in depth. (f) **Participant observation and interviews** — watching people and asking them questions.

Question 3 (a) The **independent variable** is whether a pot gets fertiliser or not. (b) The **dependent variable** is growth in height, measured after 3 weeks. (c) Any two of: same pot size, same soil, same volume of water, same light, same bean variety, same starting height.

Question 4 (a) The control group (plain water, everything else identical) shows the growth that would have happened *anyway* — seedlings grow with no tonic at all. (b) **1.1 cm**, found by subtraction: 8.9 cm (tonic) minus 7.8 cm (control). Only that difference can be credited to the tonic.

Question 5 (a) **Trial 4**, at 2.7 s, stands well apart from the other four (1.8–1.9 s). (b) Mean of all five trials: **2.02 s**; mean of trials 1, 2, 3 and 5: **1.85 s**. (c) Check what happened in trial 4 — perhaps the cone struck the desk or the watch started late — and set the result aside only if a reason is found. Never silently delete it.

Question 6 (a) The quadrats sit on disturbed ground near the carpark and walking path, chosen for ease; plants that like that ground are counted too often and plants of the open reserve almost never, so the sample does not represent the reserve. (b) Random choice gives every grid square an equal chance, so path edge and open ground both get their fair share — the sample can represent the whole reserve.

Question 7 Running the survey only at the bicycle shed is a **convenience sample** with a built-in push: it over-counts riders and misses everyone who walks, is driven, catches a bus, or arrives before 8:25. Asking more students *at the bike shed* does **not** fix it — a bigger sample settles a fair method but never un-biases an unfair one; the fix is to sample the whole school, for example by drawing names or classes at random.

Question 8 (a) If the recorder knew which cup held which juice, expectation could tilt the ratings — without any dishonesty, they might rate the expected favourite higher. Hiding the key means expectations have nothing to push on. (b) The control group receives a **placebo** — a look-alike tablet or drink with no active ingredient — so both groups believe they might have the real thing, and any extra improvement in the medicine group cannot be explained by expectation.

Question 9 (a) The estimate swings wildly for small samples (50% at 10 draws, 24% at 50) and settles closer to the true 30% as the sample grows (29% at 100, 30% at 300): bigger samples settle closer to the true value. (b) **No**. With red settled to the bottom and draws from the top, every draw — even all 300 — comes from the wrong part of the jar, so the estimate settles confidently on the *wrong* value. Size fixes only how much a fair estimate jumps around; only *how* the sample is chosen can fix bias.

Question 10 Any definition written as a follow-able rule. For example: *aggressive contact* counts when one bird's beak touches or strikes another bird, or when one bird flies after another across more than one perch. The test is: could a second observer, reading only the rule, record the same events the same way?

Question 11 (a) Two of: who collected the data, and how? Do the two sources agree where they overlap? Can the source be checked — is it a body with a method stated, not just a claim? Is the data recent and from the same creek sites? (b) It is secondary because the group used data someone else collected. Fieldwork becomes the better methodology when the question is about *now*, at *this* creek, and the existing data is old, from other sites, or unchecked — provided the 3B safety and permissions side is cleared first.

Question 12 **No**. Reproducibility compares groups following the *same* written method. The two groups on the shared method (212 and 205) agree closely — that pair *is* the reproducibility check, and it passes. The third group changed the method, so its 341 shows only that a different method gives a different count — exactly the vague-method panel of the reproducibility figure. The fix is for the third group to follow the shared method, not to abandon it.

Question 13 Reason 1: there is **no control group** — the teammates might run faster a month later anyway (training, growth, practice on the run). Fix: a second, similar group that runs the same month on a look-alike drink with no active ingredient (a placebo). Reason 2: the test is **not blind** — the teammates expect the drink to help, and the timer is the believer. Fix: code the drinks so neither the runners nor the timer knows who has which until all times are recorded. (Sampling too: teammates are a convenience sample, not a random one.)

Question 14 (a) **Spread/mistakes** — repeats and a mean smooth one-off errors. (b) **Bias** — random placement removes the push toward easy ground. (c) **Bias** — hiding the key removes expectation's push on the ratings. (d) **Spread** — a bigger fair sample settles the estimate; it cannot move a bias. (e) **Bias** (and reproducibility) — fixed rules stop each observer counting a different traffic.

Question 15 If red marbles settle to the bottom and draws come from the top, each draw is taken from the wrong part of the jar. A bigger sample then repeats the same wrong draw more times: the estimate settles, but on the wrong value — confidently wrong, in fact. Bias lives in *how* the sample is chosen, so only changing the choosing (drawing from a stirred jar, or from random depths) can fix it; size only shrinks the jumping-around of a fair method.

Question 16 (a) A case study records one case **in depth** over time — the full, real story of this wetland's recovery, with detail a broad survey would miss. (b) With only one wetland, there is no comparison and no random choice of cases, so it cannot claim that wetlands *in general* recover after fencing — this wetland may be special. A **survey** of many fenced wetlands (or fieldwork comparing several fenced and unfenced wetlands) could begin to test the general claim.

Where this leads: the planning skills taught here — validity, reproducibility, controls, random sampling, operational definitions and blind tests — are the working core of the VCE Science “key science skills” (planning and conducting investigations), reference-only for this book; VCE adds formal treatment of random and systematic error and of uncertainty, which this book leaves to evaluation in Subtopic 5A and to measurement in Subtopic 3C. The safe, ethical and cultural limb of VC2S10I02 — risk assessments and protocols on Country and Place — is delivered by Subtopic 3B, equipment and units by Subtopic 3C, and the evolution evidence used here as a secondary-data example is treated in full in Subtopic 9B.

Safe and ethical investigation: risk assessment, animal and human ethics, and protocols for cultural sites and artefacts on Country and Place

Chapter 3 — Planning and conducting (Science, Levels 9 and 10)

Subtopic 3B · VC2 Science — Planning and conducting

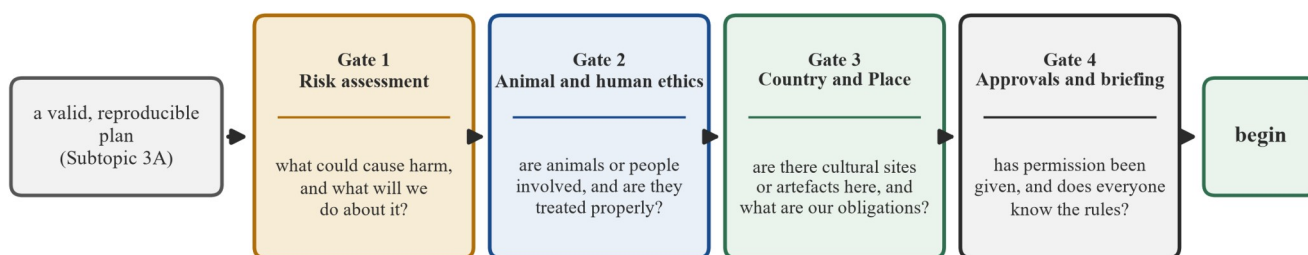
CD code: VC2S10I02 (split delivery †; Subtopic 3A is the accountable owner) **Content description**

(verbatim): “valid, reproducible investigations to answer questions and test hypotheses can be planned and conducted, including identifying and controlling for possible sources of error and bias in sampling or in making observations; safe, ethical investigations include undertaking risk assessments and following protocols when accessing cultural sites and artefacts on Country and Place” **This subtopic’s limb (TOC §4 clause split):** safe, ethical investigations — risk assessments, animal and human ethics, and protocols when accessing cultural sites and artefacts on Country and Place. The partner 3A delivers the other limb: valid, reproducible design. Every plan made there passes through this subtopic’s four gates before the work begins. **Elaboration ids drawn on:** VC2S10I02.E01 · VC2S10I02.E02 · VC2S10I02.E07 (ids VC2S10I02.E03, VC2S10I02.E05 and VC2S10I02.E06 are drawn on by Subtopic 3A; VC2S10I02.E04 is withheld under the TOC’s D10 ICIP decision and is not taught here) **Achievement standard (verbatim):** “They plan a range of valid, reproducible and safe scientific investigations and explain how they have addressed any ethical and cultural considerations when generating or using primary and secondary data.” — the safe, ethical and cultural limb is served here; the planning limb is served with Subtopic 3A.

Theory

A plan is not finished when it works — it is finished when it is safe and right. Subtopic 3A builds a plan that can answer a question validly and reproducibly. This subtopic asks the two questions that must be answered before that plan is allowed to run: *could it hurt anyone?* and *is it fair to everyone it touches — people, animals, and the places and cultural material it might meet?* A plan that gets the science right but injures a student, pressures a participant, or disturbs a cultural site is a failed plan. The content description’s second limb makes that official: safe, ethical investigations include undertaking risk assessments and following protocols when accessing cultural sites and artefacts on Country and Place.

Before an investigation begins, the plan passes four gates



a plan that fails any gate is changed before it runs — not after something goes wrong

this subtopic teaches all four gates; Gate 4 is also where SIT-2 assesses your risk assessment and ethics statement

A flow of boxes: a valid, reproducible plan from Subtopic 3A passes through four gates before the investigation may begin — Gate 1 risk assessment, Gate 2 animal and human ethics, Gate 3 Country and Place, Gate 4 approvals and briefing — and a plan that fails any gate is changed before it runs, not after something goes wrong

The four gates in the figure are the shape of this subtopic. Gate 1 asks what could cause harm and what will be done about it. Gate 2 asks whether animals or people are involved and whether they are treated properly. Gate 3 asks whether cultural sites or artefacts are present and what our responsibilities are. Gate 4 checks that permission has been given and that everyone knows the rules before anyone moves. Nothing here is a formality: a plan that fails a gate is changed before it runs.

The words this subtopic works to. The curriculum uses *hazard*, *risk*, *protocol* and *Country and Place* without fixing a single wording, so this book works to these definitions, and every question here can be answered from them:

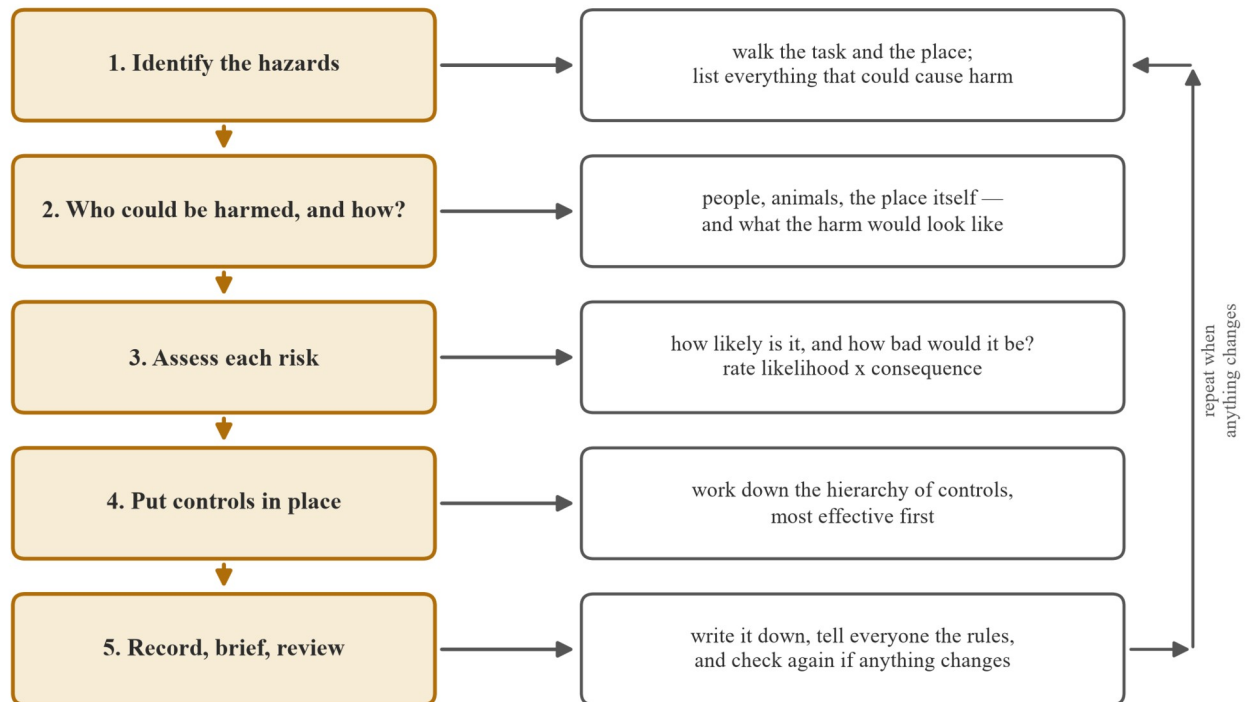
- **hazard** — anything that can cause harm: a chemical, a flame, a slippery floor, a steep bank, a living thing.
- **risk** — the chance that a hazard actually causes harm, together with how bad that harm would be. A hazard can be large while the risk is small, once controls are in place.
- **risk assessment** — the written process of finding the hazards, rating the risks and deciding the controls, done *before* the work starts.
- **ethical consideration** — a question about what is right and fair in the treatment of the people and animals an investigation involves or affects.
- **protocol** — an agreed rule for how to act in a situation. On *Country and Place*, protocols are the correct ways to behave there — followed because they are right, and because in many cases they are also responsibilities the law puts on us.
- **Country and Place** — the curriculum's term for the lands, waters and places that Aboriginal and Torres Strait Islander Peoples belong to and carry responsibilities for, including cultural sites and artefacts. This book uses the term the way the curriculum does, and teaches the checks and protocols that protect those places.

(The Victorian Curriculum uses these terms without a single fixed wording, so this subtopic works to the definitions above.)

Hazard and risk are not the same word. A **hazard** is a source of harm; a **risk** rates what that hazard is likely to do. A beaker of dilute acid is a hazard. Whether it is a big risk depends on how likely a splash is and how bad the splash would be — which is exactly what controls change. That distinction is why the same chemical can be handled safely in a school lab under one set of rules and not under another: the hazard stayed, the risk was moved.

The risk assessment in five steps (elaboration VC2S10I02.E07). Elaboration VC2S10I02.E07 names the skill: identifying the potential hazards of the chemicals or biological materials and processes an investigation uses, and how they should be addressed. The figure turns that into a five-step routine that is always done before the work starts — and repeated whenever anything changes.

A risk assessment in five steps — done BEFORE the work starts



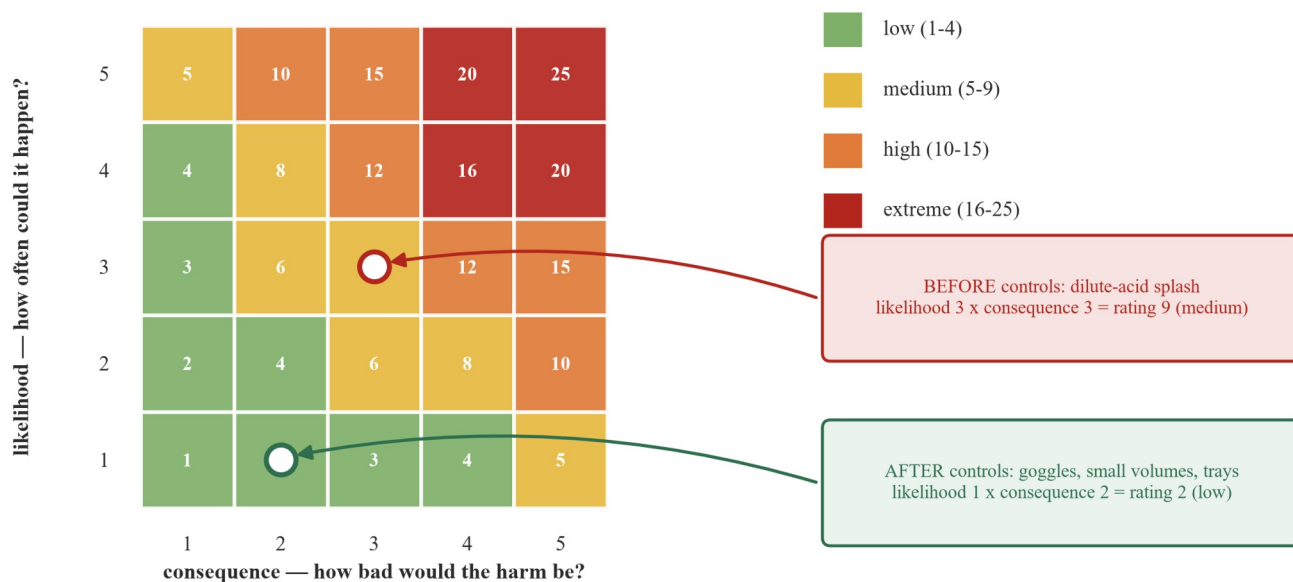
The five steps of a risk assessment in order: identify the hazards; work out who could be harmed and how; assess each risk by rating likelihood times consequence; put controls in place, most effective first; then record it, brief everyone, and review — with a loop arrow back to the start, to repeat when anything changes

Step 3 rates each risk as **likelihood** × **consequence**, and step 4 chooses controls from the most effective down. Step 5 is what makes it an assessment rather than a wish: it is written down, everyone is briefed on the rules, and it is checked again if the task, the place or the weather changes.

The risk grid: likelihood × consequence. Ratings are usually made on a grid. Likelihood is scored 1 (unlikely) to 5 (very likely), consequence 1 (minor) to 5 (very severe), and the risk rating is the two multiplied. The rating then sits in a zone: 1–4 low, 5–9 medium, 10–15 high, 16–25 extreme. The point of the grid is not the arithmetic; it is that controls must move a task *down and left* — less likely, and less bad if it happens — before the work begins.

risk rating = likelihood x consequence

controls move a task DOWN and LEFT on this grid — that is the whole point of doing one

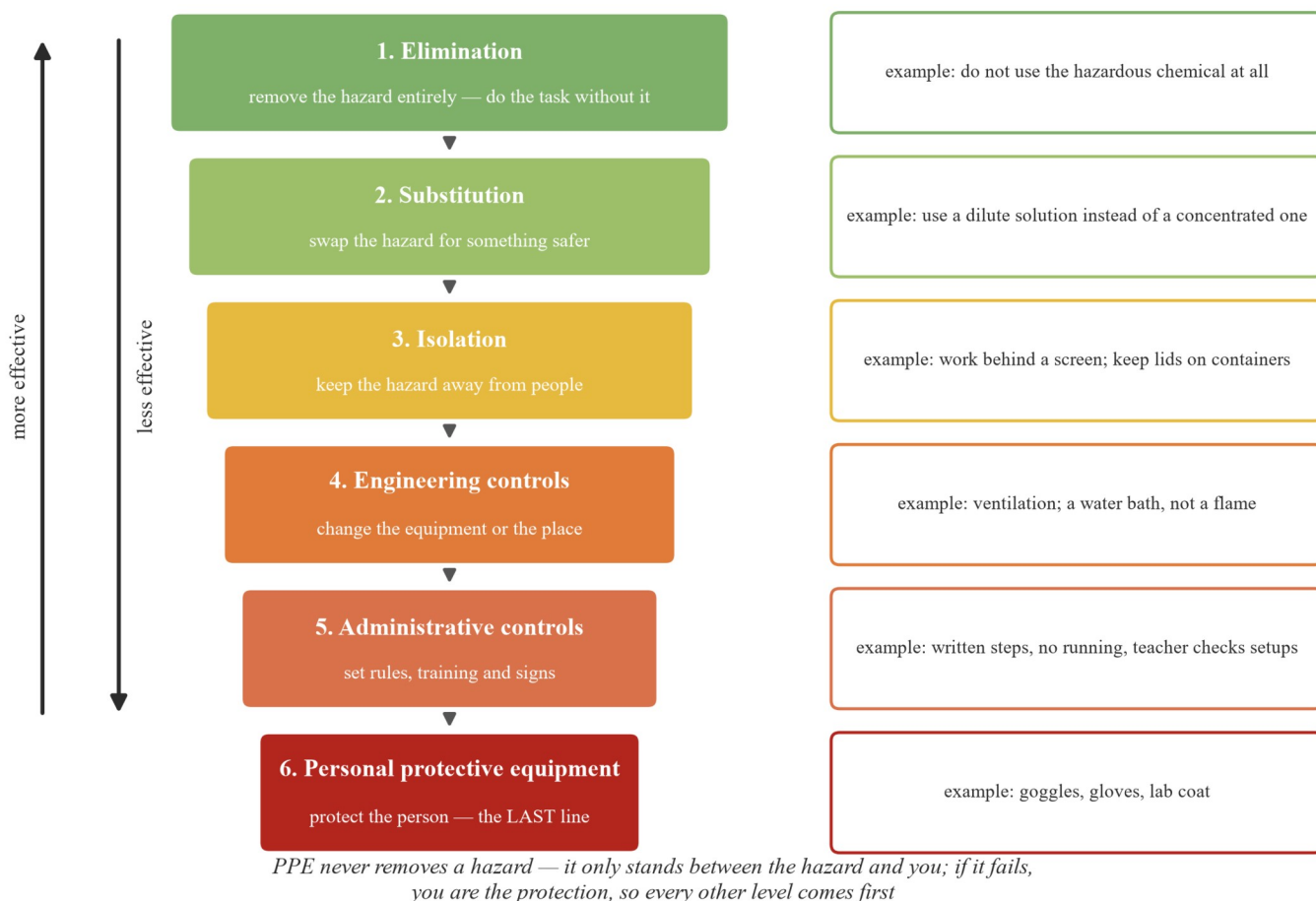


A five-by-five risk grid with likelihood on one axis and consequence on the other, each cell coloured by its zone — low, medium, high or extreme — with a worked re-rating: a dilute-acid splash rated likelihood 3 times consequence 3 equals 9, medium, before controls, and likelihood 1 times consequence 2 equals 2, low, after controls such as goggles, small volumes and trays

In the figure, a splash of dilute acid onto skin or eyes is rated likelihood $3 \times$ consequence $3 = 9$ (medium) before controls. With the controls in place — dilute acid only, small volumes, splash-proof goggles for everyone, bottles standing in trays — the same task is likelihood $1 \times$ consequence $2 = 2$ (low). The hazard never left the room; the risk was moved from medium to low. That move is the whole purpose of the exercise.

The hierarchy of controls — fix the hazard before you wrap the person. When a risk needs controls, the order of thinking matters. The **hierarchy of controls** is that order: an arrangement from the controls that deal with the hazard itself down to the controls that only stand between the hazard and the person. (In this book, *hierarchy* means an order from most to least effective.)

The hierarchy of controls — fix the hazard itself before you wrap the person



The hierarchy of controls as a funnel from most to least effective: elimination, substitution, isolation, engineering controls, administrative controls, and personal protective equipment last — each with an example, such as using a dilute solution instead of a concentrated one, and a water bath rather than a flame

The rule the figure teaches is that personal protective equipment (PPE) is the *last* line, not the first. Goggles do not stop a splash happening; they only stand between the splash and an eye. Every level above them either removes the hazard (do the task without it, or swap it for something safer) or keeps it away from people (screens, lids, ventilation, written rules). A plan that reaches for goggles before it has asked “can we remove or swap this hazard?” has done the hierarchy backwards.

Know the symbols, read the sheet. Chemical containers carry hazard symbols that say, at a glance, what the chemical can do. The simplified drawings in the figure are the nine you are most likely to meet; on any real container the label is checked first, and the label points to the full story — the **safety data sheet (SDS)**. (In this book, *safety data sheet* means the sheet that comes with a chemical, listing its hazards and how to handle, store and clean it up safely.) Step 1 of the risk assessment leans on both: the symbol says what to look for, the SDS says what the controls should be.

Hazard symbols you may meet on a chemical container

simplified classroom drawings — on a real container, always read the label and the safety data sheet (SDS)

 flammable warns: catches fire easily you: keep away from flames; store safely	 toxic warns: poison — can kill or cause serious illness you: never taste; avoid breathing vapour	 corrosive warns: burns skin, eyes and some materials you: goggles and gloves; carry carefully
 oxidising warns: feeds other fires — makes burning fiercer you: keep away from flammable things	 irritant / harmful warns: can irritate or harm skin, eyes or breathing you: avoid contact; ventilate the room	 compressed gas warns: gas under pressure — can burst if heated you: store upright; keep cool
 explosive warns: can explode if heated, hit or mixed wrongly you: never heat or grind; teacher handles it	 environmental hazard warns: harms living things in water and soil you: never pour down the sink or outside	 serious health hazard warns: can cause long-term illness after exposure you: avoid all exposure; special handling rules

Nine simplified hazard symbols in red diamonds — flammable, toxic, corrosive, oxidising, irritant or harmful, compressed gas, explosive, environmental hazard, and serious health hazard — each with what it warns about and what you should do, such as keeping flammables away from flames and never pouring environmental hazards down the sink

A completed risk assessment: the marble-chips investigation. The figure shows what a finished assessment looks like for a common lab task — how the strength of an acid changes the speed of its reaction with marble chips (calcium carbonate). The chemistry itself belongs to Chapter 11; here the investigation is only the *task to be made safe*.

A completed risk assessment — investigation: how does acid concentration change the rate of reaction of marble chips (calcium carbonate)? (the reaction itself is Chapter 11)

hazard	who could be harmed	rating BEFORE controls	controls	rating AFTER controls
dilute acid splashes onto skin or eyes	students handling the acid	likelihood 3 x consequence 3 rating 9	dilute acid only, small volumes, goggles for everyone, bottles in trays	likelihood 1 x consequence 2 rating 2
carbon dioxide gas builds up if a flask is stoppered	the group at that setup	likelihood 2 x consequence 4 rating 8	never stopper the flask; let gas escape into a gas syringe or through loose tubing	likelihood 1 x consequence 3 rating 3
glass breaks while fitting tubing or syringes	hands of the student fitting it	likelihood 2 x consequence 2 rating 4	check glassware first; push tubing slowly with a cloth; teacher checks each setup	likelihood 1 x consequence 2 rating 2
spilled liquid makes the bench or floor slippery	anyone walking past the bench	likelihood 2 x consequence 2 rating 4	wipe spills at once; keep bags off the floor; walk, never run, in the lab	likelihood 1 x consequence 1 rating 1

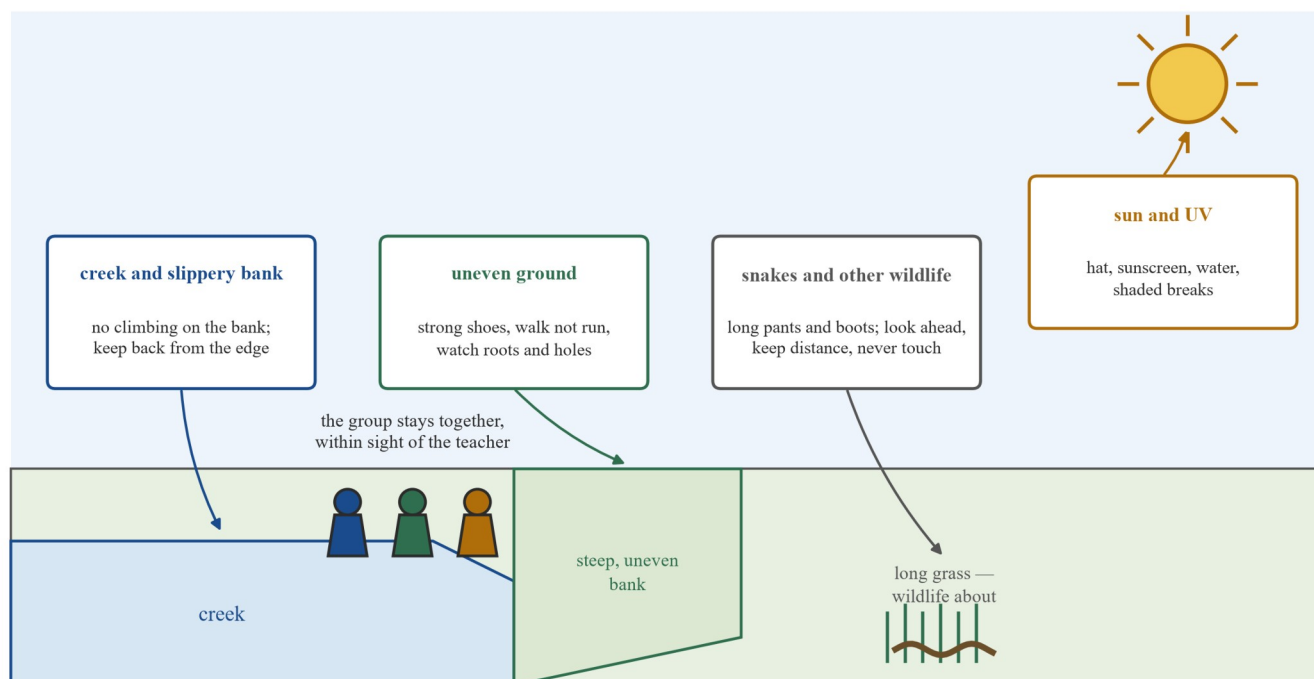
every control moves its rating down — the column on the right is what the teacher checks before any group starts; a rating still in the higher zones means the plan is not yet safe to run

A completed risk assessment table for the marble-chips investigation with four hazards — dilute acid splashes, gas building up in a stoppered flask, broken glass while fitting tubing, and spilled liquid making the floor slippery — each with who could be harmed, a before-controls rating, the controls chosen, and a lower after-controls rating

Read the table the way a teacher checks it. Every hazard names *who* could be harmed and *how*. Every rating is shown before and after controls, and every after-rating is lower — dilute acid splash 9 to 2, stoppered-flask gas 8 to 3, broken glass 4 to 2, slippery floor 4 to 1. And the controls follow the hierarchy’s order of thinking: swap to dilute acid and small volumes before relying on goggles; rule “never stopper the flask” before rule “walk, never run”. A rating still sitting in the high or extreme zone after controls means the plan does not run — it goes back and gets better controls.

Fieldwork has its own hazards. Outside the lab there is no bench, no trays and no teacher at every elbow — the place itself supplies the hazards, and the risk assessment has to be done for *that* place, on the day, not copied from last year’s excursion.

Fieldwork has its own hazards — name each one, control each one



A labelled fieldwork scene beside a creek: four hazards, each with its controls — the creek and slippery bank (no climbing, keep back from the edge), uneven ground (strong shoes, walk not run), snakes and other wildlife in long grass (long pants and boots, keep distance, never touch), and sun and UV (hat, sunscreen, water, shaded breaks) — while the group stays together within sight of the teacher

The figure's scene is the same five-step assessment applied to a creek: the hazards are named (water and slippery banks, uneven ground, wildlife, sun), the people who could be harmed are the group and anyone walking alone, and each hazard gets a control — including the administrative ones, like staying within sight of the teacher. Notice the pattern from the lab holds outdoors: the controls that keep people *away* from the hazard (keep back from the bank, keep distance from wildlife) come before the controls that wrap the person (hat, boots, sunscreen).

Animals in investigations: the discussion happens first (elaboration VC2S10I02 .E01). Elaboration

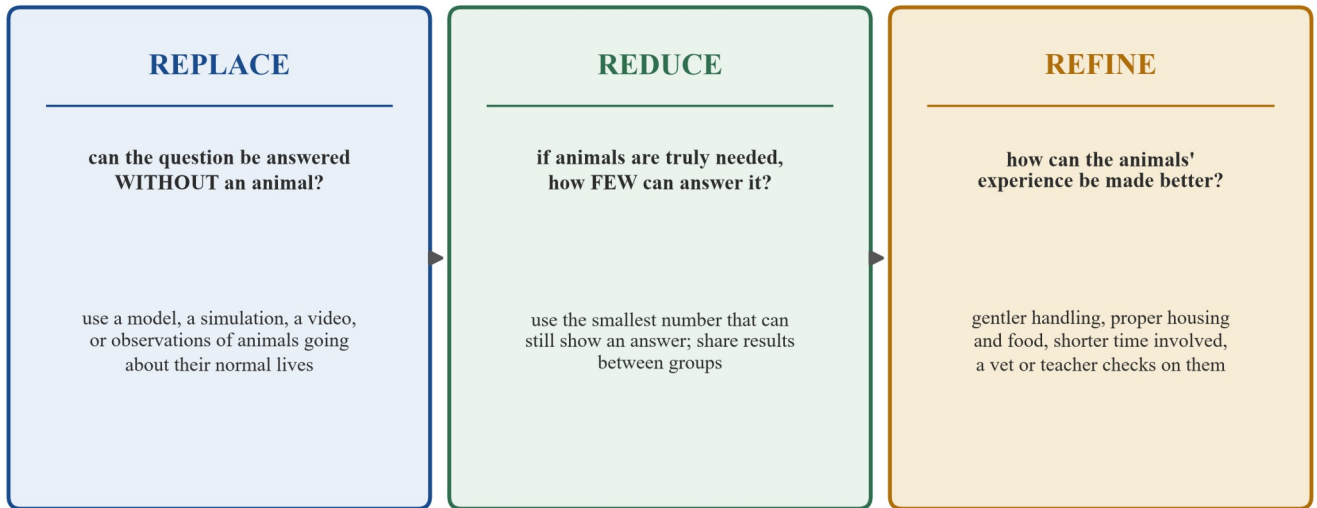
VC2S10I02 .E01 asks for a discussion of the ethical and social issues involved in the care and use of animals for scientific purposes, *before* starting an investigation that involves animals. The discussion is real, and both sides of it deserve their strongest form.

One view is that careful, regulated use of animals has been hard to replace: much of what medicine knows about vaccines, anaesthetics, nutrition and disease came from work with animals, and some questions — how a whole living body grows, heals or reproduces — cannot yet be answered fully by a model or a screen. On this view, the honest position is not “never”, but “only when needed, only with approval, and only with the animals’ welfare held as a duty”.

The other view is that animals cannot agree to take part, can suffer, and do not get to weigh the benefit against the cost the way a human participant can — so the burden of proof sits entirely on the people using them. On this view, every avoidable use is wrong, and the steady improvement of models, simulations and non-invasive observation makes “unavoidable” rarer every year.

The curriculum does not ask you to pretend one of these views is silly. It asks you to weigh both fairly *before* a plan involving animals is written — and then to apply the rules that society has already settled: animals are used only when the question truly needs them, in the smallest numbers, with the gentlest methods, under proper approval.

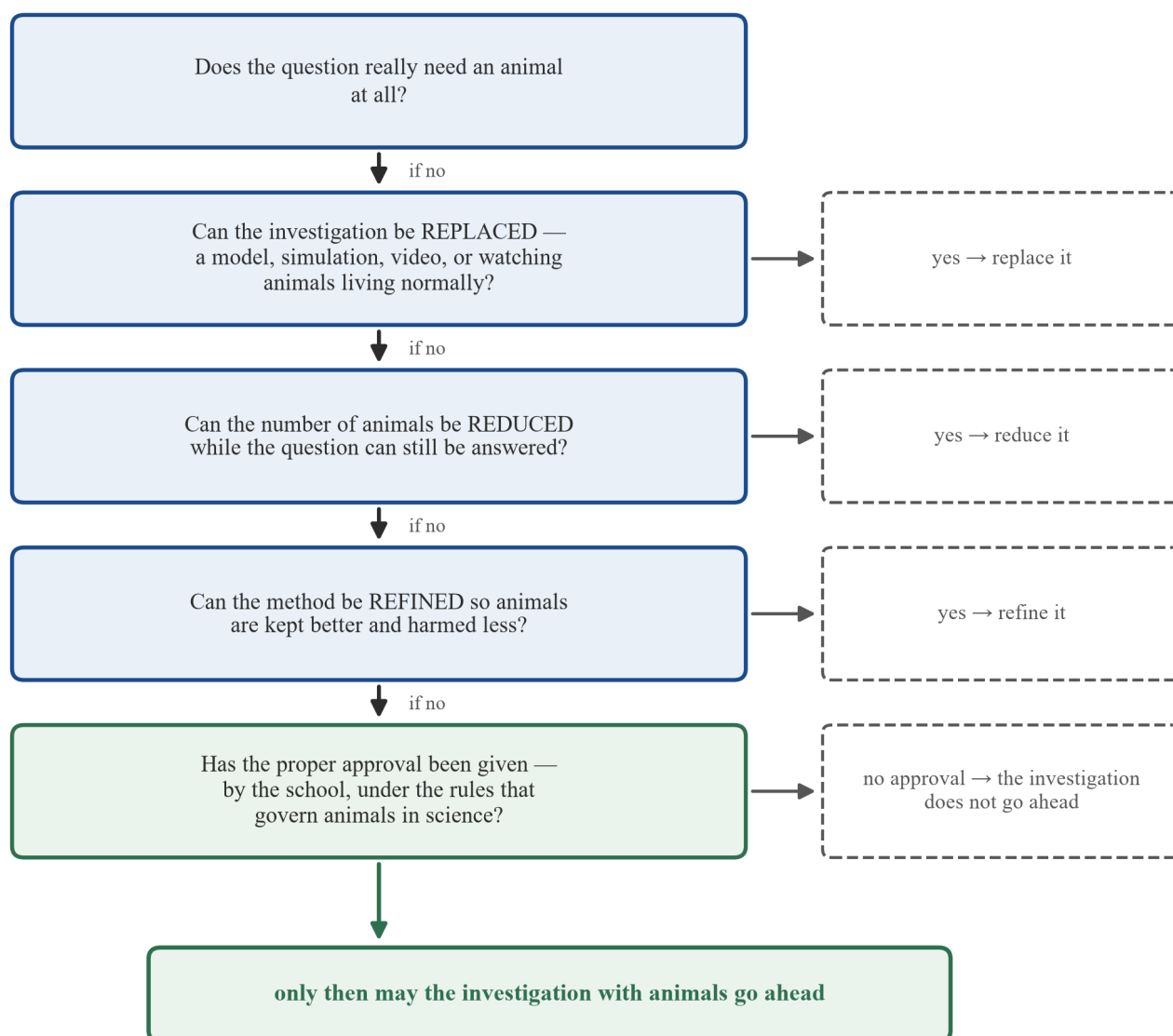
The three Rs of animal ethics — Replace, Reduce, Refine



The three Rs of animal ethics as three panels: Replace — can the question be answered without an animal, using a model, simulation, video or observations of animals living normally; Reduce — if animals are truly needed, use the smallest number that can still show an answer and share results between groups; Refine — make the animals' experience better with gentler handling, proper housing and food, shorter time, and checks on their welfare

The three Rs in the figure are the working form of that settled position. **Replace** asks whether a model, a simulation, a video, or watching animals going about their normal lives can answer the question at all. **Reduce** asks, if animals are truly needed, how few can still give an answer. **Refine** asks how the animals' experience can be made better — gentler handling, proper housing and food, the shortest possible time, and someone checking on their welfare. They are applied in that order, every time.

The questions answered before any animal is used



The questions answered before any animal is used, in order: does the question really need an animal at all; can it be replaced; can the number be reduced; can the method be refined so animals are kept better and harmed less; and has the proper approval been given — yes answers lead to replacing, reducing or refining, and without approval the investigation does not go ahead; only then may the investigation with animals go ahead

The decision chain in the figure ends at approval, and that ending is not a formality. In Australia, the care and use of animals for scientific purposes is governed by the *Australian Code for the Care and Use of Animals for Scientific Purposes*, and school investigations sit under rules made from it — which is why a school investigation involving animals is approved by the school's processes before it begins, never after. The Code is named here rather than quoted: the detail belongs to your teacher and to the Code itself. What belongs to you is the habit — replace, reduce, refine, approval, in that order, every time.

- State of Victoria, *Aboriginal Heritage Act 2006* (Vic), legislation.vic.gov.au/in-force/acts/aboriginal-heritage-act-2006/06-016 — retrieved 2026-08-21.
- State of Victoria, *Aboriginal Heritage Regulations 2018* (Vic), legislation.vic.gov.au/in-force/statutory-rules/aboriginal-heritage-regulations-2018/018 — retrieved 2026-08-21.
- Australian Government (NHMRC and partner councils), *Australian Code for the Care and Use of Animals for Scientific Purposes*, 8th ed., 2013, nhmrc.gov.au — retrieval attempted 2026-08-21 (host unreachable from this environment; the instrument is named only, per the TOC's D10 approach).

- NHMRC, *National Statement on Ethical Conduct in Human Research*, 2018 (updated 2023), nhmrc.gov.au — retrieval attempted 2026-08-21 (host unreachable from this environment; the instrument is named only).
- AIATSIS, *AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research*, 2020, aiatsis.gov.au — retrieval attempted 2026-08-21 (host blocks automated retrieval; the instrument is named only).

People in investigations: five principles. When people take part — answering a survey, sitting a reaction-time test, being observed — the ethics are about consent, fairness and privacy. Five principles cover school investigations, and they apply before the first question is asked, not after.

Five principles when people take part in your investigation

1. Informed consent	2. Voluntary participation	3. Privacy	4. No harm	5. Ethics approval
participants are told, in plain language, what will happen — before they agree <i>your survey: explain it on the consent form, and let people ask</i>	nobody is forced to take part, and anyone can stop at any time without giving a reason <i>your survey: “you may skip any question, or stop, at any time”</i>	collect no names or details that could identify a person; results are anonymous <i>your survey: no names, no classes, no photos without permission</i>	keep risk and discomfort low — including from the questions themselves <i>your survey: nothing private, embarrassing or upsetting</i>	formal research is reviewed by an ethics committee before it starts <i>your classwork: your teacher checks the plan first</i>

Five principles when people take part in an investigation, each with what it means and what it looks like in a class survey: informed consent — participants are told in plain language what will happen before they agree; voluntary participation — nobody is forced, and anyone can stop at any time without giving a reason; privacy — no names or details that could identify a person, results anonymous; no harm — risk and discomfort kept low, including from the questions themselves; ethics approval — formal research is reviewed by an ethics committee before it starts, and in classwork the teacher checks the plan first

The principles are not abstract. **Informed consent** means the participant is told, in plain language, what will happen — before agreeing, not after. **Voluntary participation** means nobody is pressed into taking part, and anyone can stop at any time without giving a reason. **Privacy** means no names or details that could identify a person are collected, so the results cannot be traced back to anyone. **No harm** includes harm a question can cause: a survey has no right to ask things that are private, embarrassing or upsetting. **Ethics approval** means that in formal research an ethics committee reviews the plan before it starts; in classwork the same gate is your teacher checking the plan first. The national framework researchers work under is the NHMRC *National Statement on Ethical Conduct in Human Research* — named here, not quoted; the five principles above are the school-facing shape of it.

The consent form is where the principles become visible. A participant information and consent form is simply the five principles written down where a participant can read them and agree to them — or not.

A model participant information and consent form

Participant information and consent form	
What this investigation is about: We are studying how sleep affects reaction time in Year 10 students, for a school science assessment.	in plain language, BEFORE agreeing — informed consent
What you would be asked to do: Press a button as fast as you can when a light turns green. This takes about five minutes, once.	exactly what will happen and for how long
Any risks or discomfort: None expected. You will sit comfortably the whole time.	honest about risks, even when they are small
What happens to your results: Your results are recorded with a code only. No name, class or photo is collected; no one can identify you.	privacy built in — nothing that could identify
Your rights: Taking part is voluntary. You may skip any part or stop at any time, for any reason, without giving a reason.	voluntary, and the right to withdraw
Who to ask: Ask your science teacher before, during or after.	
[] I have read the information above, I understand it, and I agree to take part.	
_____ signature of participant (and of a parent/carer if under 18)	_____ date

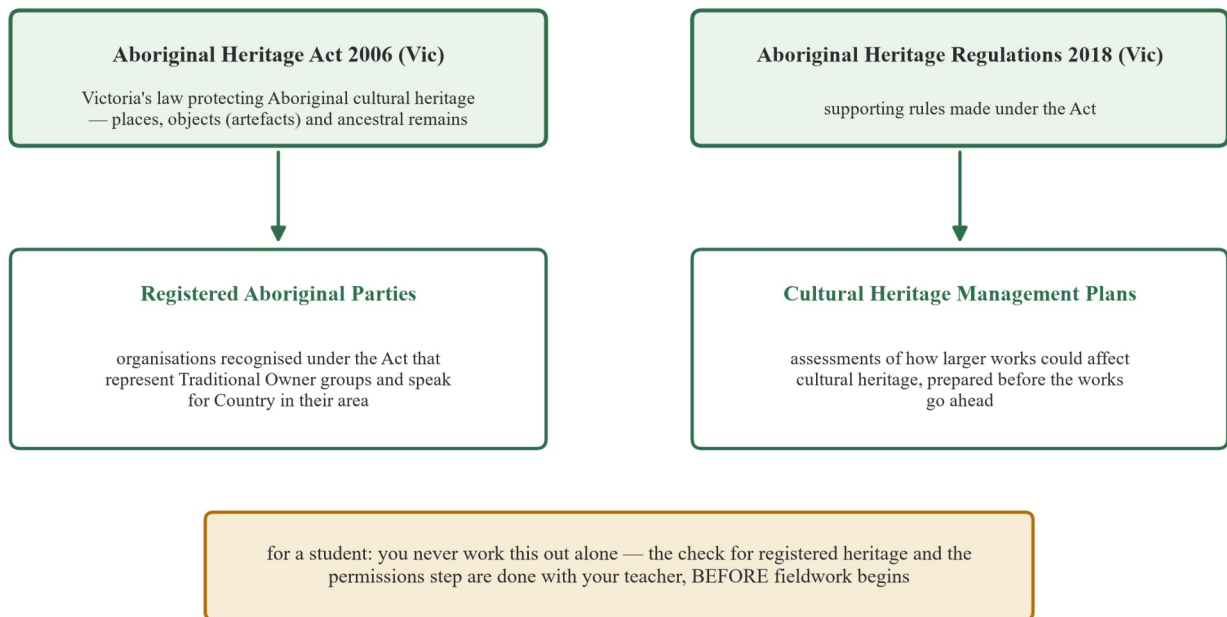
A model participant information and consent form for a class investigation of sleep and reaction time, with annotated parts: what the investigation is about, explained in plain language before agreeing; exactly what the participant would be asked to do and for how long; an honest statement of any risks, even small ones; what happens to the results — recorded with a code only, nothing that could identify anyone; the participant's rights — voluntary, and the right to skip any part or stop at any time; who to ask; and a signature line for the participant and a parent or carer if under 18

Read the model form against the principles. It says what will happen and for how long; it is honest about risks even when they are tiny; it builds privacy in by collecting codes instead of names; and it states the right to skip or stop at any time. A form that misses one of these is not consent — signing a form you were not properly informed by agrees to nothing. For participants under 18, a parent or carer signs as well, and the participant's own agreement still matters: a signed form does not override a student who says, on the day, "I don't want to do this."

Country and Place: the checks before fieldwork. Some of the places science classes visit — creeks, coasts, reserves, rock outcrops — are also places of cultural significance to Aboriginal and Torres Strait Islander Peoples, and some objects lying in the landscape are artefacts protected by law. This subtopic teaches the *checks and the behaviour*, not cultural knowledge: nothing here asks anyone to share or explain cultural knowledge, and no student is expected to. What is expected is the habit of checking before you go, and following every condition once you are there.

Victoria's framework for protecting Aboriginal cultural heritage

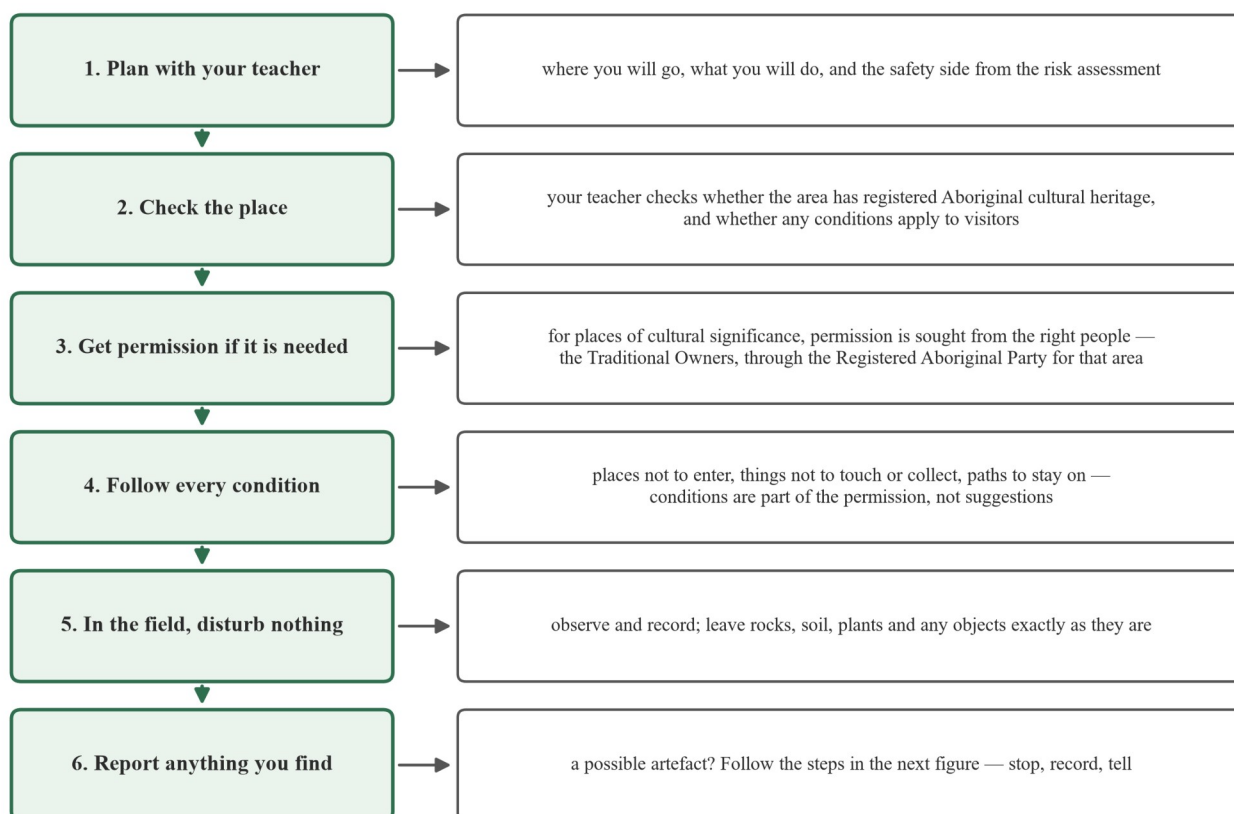
the instruments are named only — the law itself says what they do; your teacher and the relevant authorities give the detail for any real place



Victoria's framework for protecting Aboriginal cultural heritage, with the instruments named only: the Aboriginal Heritage Act 2006 (Vic) and the Aboriginal Heritage Regulations 2018 (Vic) above two pillars — Registered Aboriginal Parties, organisations recognised under the Act that represent Traditional Owner groups and speak for Country in their area, and Cultural Heritage Management Plans, assessments of how larger works could affect cultural heritage — with a note that a student never works this out alone: the check and the permissions step are done with the teacher before fieldwork begins

In Victoria, the *Aboriginal Heritage Act 2006* protects Aboriginal cultural heritage — places, objects and ancestral remains — with supporting rules in the *Aboriginal Heritage Regulations 2018*. Under the Act, **Registered Aboriginal Parties** represent Traditional Owner groups and speak for Country in their area, and **Cultural Heritage Management Plans** assess how larger works could affect heritage before those works go ahead. The instruments are named here, not unpacked — the law itself says what they do, and the detail for any real place comes from your teacher and the relevant authorities. The student-facing rule is the amber one in the figure: you never work this out alone. The check for registered heritage and the permissions step are done with your teacher, before fieldwork begins.

The permissions sequence before fieldwork on Country

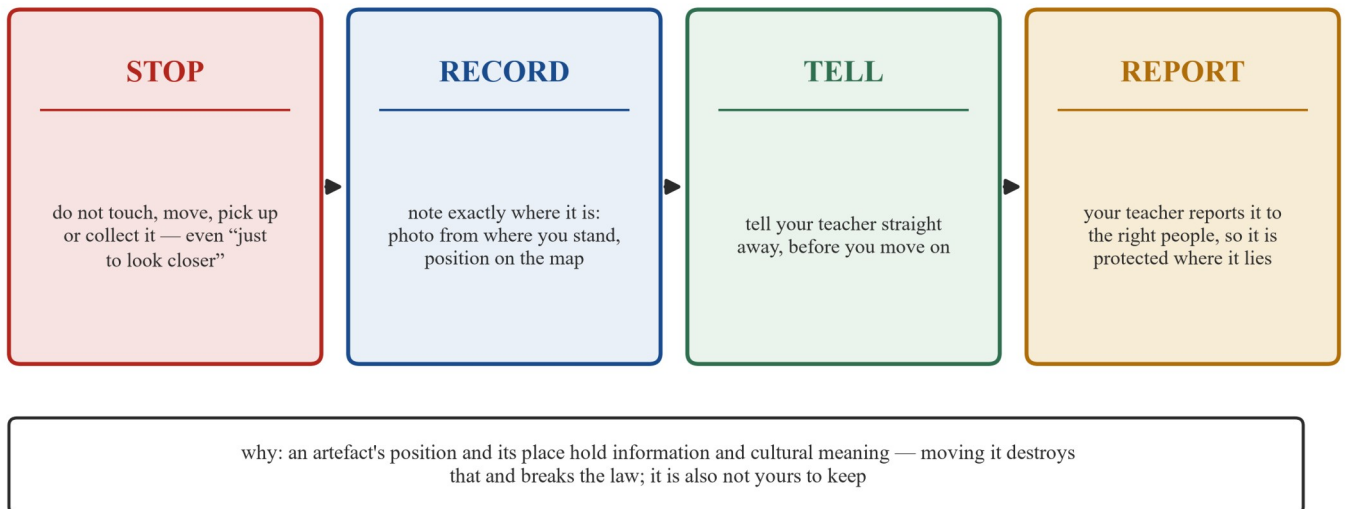


The permissions sequence before fieldwork on Country, in six steps: plan with your teacher; your teacher checks whether the area has registered Aboriginal cultural heritage and whether any conditions apply to visitors; if permission is needed, it is sought from the right people — the Traditional Owners, through the Registered Aboriginal Party for that area; follow every condition, because conditions are part of the permission, not suggestions; in the field, disturb nothing — observe and record, and leave rocks, soil, plants and any objects exactly as they are; report anything you find by stopping, recording and telling your teacher

The sequence in the figure is Gate 3 and Gate 4 made concrete. Conditions deserve a hard look, because they are the part most easily treated as suggestions: a place not to enter, a thing not to touch or collect, a path to stay on. A condition is part of the permission — stepping over it is not a small breach of etiquette but the withdrawal of the permission itself. And “disturb nothing” is the field rule that protects both the science and the place: a sample taken home is a sample removed from every record that gives it meaning.

If you find a possible artefact: stop, record, tell, report. Artefacts are not souvenirs, not even interesting stones; an object lying in the landscape may be Aboriginal cultural heritage, and its position is part of what it is.

You find a possible artefact in the field — what you do



What to do if you find a possible artefact in the field, in four steps: STOP — do not touch, move, pick up or collect it, even just to look closer; RECORD — note exactly where it is, with a photo from where you stand and its position on the map; TELL — tell your teacher straight away, before you move on; REPORT — your teacher reports it to the right people, so it is protected where it lies — because an artefact's position and its place hold information and cultural meaning, and moving it destroys that and breaks the law

The four steps are in order for a reason. Stopping protects the object and its position. Recording — a photo from where you stand, the position marked on a map — keeps the information that picking it up would destroy. Telling your teacher straight away moves the decision to the person who can act on it. Reporting, by the teacher, to the right people means the object is protected where it lies, under the law that covers it. Taking it “just to look closer”, or to show the class, removes an object that was never yours to keep — and in Victoria, harming or removing Aboriginal cultural heritage without authority is against the law.

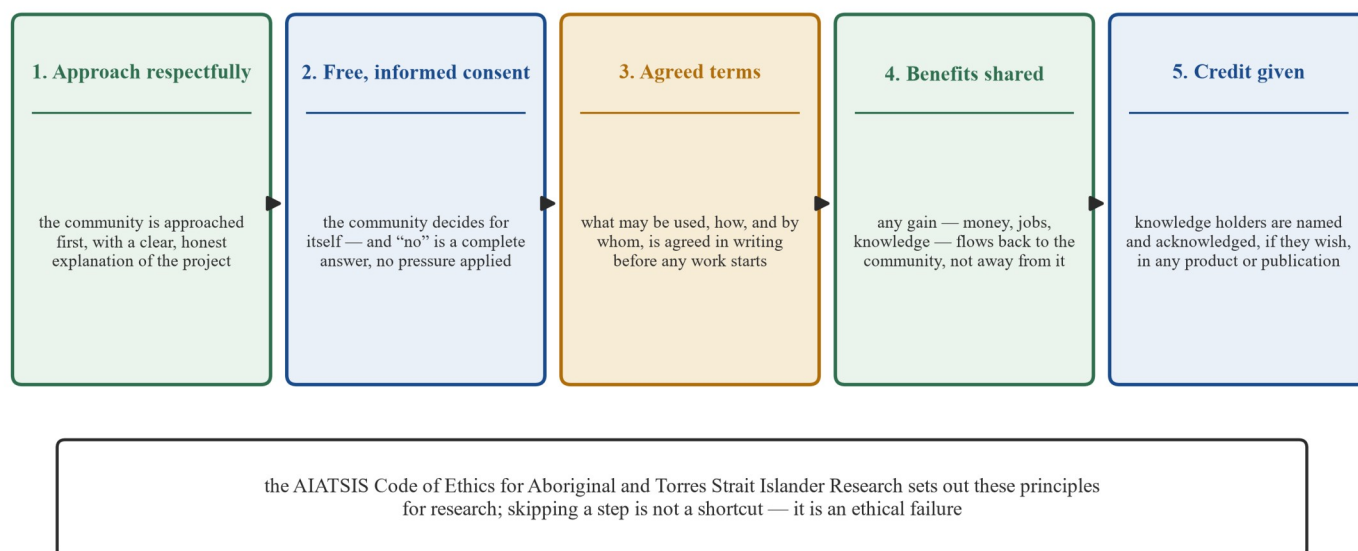
Working with Aboriginal and Torres Strait Islander knowledge: consent, agreement, benefit-sharing, credit (elaboration VC2S10I02 . E02). The last part of this subtopic covers knowledge, not just places and objects. Elaboration VC2S10I02 . E02 reads:

“recognising and implementing ethical requirements when collaborating with Aboriginal and/or Torres Strait Islander people to develop commercial products that utilise traditional ecological knowledges, for example the use of emu oil as an anti-inflammatory product, kangaroo apple as an analgesic and banksia flowers in antiseptic formulations”

The elaboration's examples are quoted here as the curriculum's own; this book does not teach the knowledge behind them, which belongs to its holders. What the elaboration asks you to recognise is the *requirements*: when knowledge like this is used — and especially when money is made from it — there are obligations, and they run in a fixed order.

Working with Aboriginal and Torres Strait Islander knowledge — the obligations

Indigenous Cultural and Intellectual Property (ICIP): the rights of Aboriginal and Torres Strait Islander Peoples to their knowledge and cultural expressions — others need permission to use them



Working with Aboriginal and Torres Strait Islander knowledge — the obligations, in five steps: approach respectfully, with a clear, honest explanation of the project to the community first; free, informed consent, where the community decides for itself and “no” is a complete answer with no pressure applied; agreed terms, in writing, before any work starts, covering what may be used, how, and by whom; benefits shared, so any gain — money, jobs, knowledge — flows back to the community, not away from it; and credit given, with knowledge holders named and acknowledged, if they wish, in any product or publication — the AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research sets out these principles, and skipping a step is not a shortcut but an ethical failure

The rights behind that order have a name: **Indigenous Cultural and Intellectual Property (ICIP)** — the rights of Aboriginal and Torres Strait Islander Peoples to their knowledge and cultural expressions, which others need permission to use. The *AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research* sets out the principles the figure shows, for research; the same shape governs commercial use. Three points carry into any classroom discussion of it. First, consent is *free and informed*: the community is approached first, told honestly what is proposed, and “no” is a complete answer — a project that pressures, or that starts before agreement, has already failed. Second, agreement and benefit-sharing are not gestures: what may be used, how and by whom is agreed in writing before work starts, and any gain flows back to the community, not away from it. Third, credit is given where it is due, if the knowledge holders wish — named and acknowledged in any product or publication. Skipping a step is not a shortcut; it is an ethical failure.

The four gates together: writing the safety and ethics statement. Everything in this subtopic collects into one short written statement that sits beside the 3A plan, and that the achievement standard assesses — “explain how they have addressed any ethical and cultural considerations when generating or using primary and secondary data”. A statement that passes the four gates says:

- **Gate 1 — risk:** the hazards were identified (chemicals, equipment, place), each was rated likelihood × consequence, and the controls chosen from the hierarchy are listed; the after-controls ratings are all in the low zone.
- **Gate 2 — animals and people:** whether animals or people are involved; if animals, the three Rs and the approval; if people, informed consent, voluntary participation, privacy, no harm, and the teacher’s check of the plan.
- **Gate 3 — Country and Place:** whether the place was checked for cultural sites or artefacts, what permissions and conditions apply, and the stop–record–tell–report rule if anything is found; and if Aboriginal or Torres Strait Islander knowledge is used, the consent, agreement, benefit-sharing and credit steps.

- **Gate 4 — approvals and briefing:** who approved the plan, and that everyone involved has been briefed on the rules before the work begins.

Written like that, the statement is not paperwork. It is the plan's answer to the two questions this subtopic began with — could it hurt anyone, and is it fair to everyone it touches.

Activities

Activity 1 — Risk-assess the marble-chips investigation (elaboration VC2S10I02.E07). In pairs, use the five-step routine to build your own risk assessment for the marble-chips investigation *before* looking at the completed table in the figure. Step 1: from the task description (dilute acid, glassware, tubing, gas produced, liquids on the bench), list four hazards. Step 2: for each, name who could be harmed and how. Step 3: rate each likelihood $\times$ consequence on the grid and note its zone. Step 4: choose controls, working down the hierarchy — and write the after-controls rating each control earns. Then compare your table with the completed one: for any hazard you missed, which step would have caught it? For any rating you gave that differs, argue your case from the grid, not from feelings.

Activity 2 — The creek plan, end to end. Take the creek survey plan from Subtopic 3A (same question, same sampling rules) and pass it through the four gates. Gate 1: complete a field risk assessment for the scene in the figure, adding any hazard your own local creek would contribute. Gates 3 and 4: write the permissions sequence for your chosen site as six numbered steps from the figure, including who checks, who is asked if permission is needed, and the conditions the group would follow. Gate 2 will usually read “no animals; no human participants” — but if your plan surveys people (for example, asking visitors to the creek to answer two questions), the consent form rules apply: draft the form. The output is one page: the 3A method on the left, the four-gate checks on the right.

Activity 3 — The ethics panel. Split into panels of three. Each panel receives three proposals: (A) collect 40 tadpoles from the school creek and keep them in classroom tanks at two temperatures until they become frogs; (B) survey Year 10 reaction time, collecting names and classes “so the results can be checked”, with participation announced as compulsory for the class; (C) a group's market-day pitch that uses, as its selling point, Aboriginal knowledge of a local plant found on a website, without asking anyone. Rule on each proposal using the decision figures of this subtopic: for A apply replace, reduce and refine, and the approval gate; for B apply the five human-participant principles and redraft the two worst lines of the proposal; for C apply the five steps of the knowledge figure and name which steps were skipped. Panels return written rulings that give each proposal's strongest point before its failures — a ruling that cannot state a proposal's case fairly is not finished.

Worked examples

Example 1 — scaffolded

A group proposes to warm a small volume of a flammable liquid in a beaker over a burner. Their draft risk assessment rates the fire hazard as likelihood $3 \times$ consequence $5 = 15$ (high) and proposes one control: “wear goggles”. (a) Explain why goggles alone leave the rating high. (b) Apply the hierarchy of controls to propose two controls above PPE, and re-rate the risk after all controls, using likelihood 1 and consequence 3. (c) State the zone the task must sit in before it runs, and what happens if it cannot get there.

Working	Comment
Goggles protect eyes only; the fire hazard — a flammable liquid over an open flame — is untouched, so likelihood stays near 3 and consequence near 5	PPE is the last line of the hierarchy; it never removes or isolates the hazard.
Higher controls: substitution/engineering — heat with a water bath or electric hot plate, not a flame; elimination/reduction — use the smallest volume that works, or a less flammable liquid if the task allows	Work down the hierarchy: fix the hazard before you wrap the person.
Re-rating: likelihood $1 \times$ consequence $3 = 3$	No flame and a small volume make a fire unlikely, and what remains (hot liquid) would burn, not ignite the room.

Working	Comment
3 sits in the low zone (1–4); a task may run only when its after-controls rating is in the low zone	If the rating cannot be brought to low, the plan does not run — it is changed until it can.

Example 2 — scaffolded

A group’s tadpole proposal: “collect 40 tadpoles from the creek, keep them in two classroom tanks at different temperatures, and count how many develop into frogs. We will keep them until the end of term.” (a) Apply **Replace**: give one way the question could be partly answered without collecting animals, and say what the animal work still adds. (b) Apply **Reduce** and **Refine**: two changes each, drawn from the figure. (c) Name the gate the proposal has not yet reached, and what happens without it.

Working	Comment
Replace: observe and count tadpoles at the creek over several weeks, or use an existing dataset or video of development; keeping animals in tanks is only needed if the question requires conditions (like controlled temperature) the creek cannot give	Replace is asked first, every time — but “partly” is honest: some questions genuinely need the animal, which is what the next Rs govern.
Reduce: fewer tadpoles per tank, with the class sharing and combining results across groups; Refine: return the tadpoles to the creek as soon as each one completes metamorphosis rather than keeping them to the end of term; proper food, shade and clean water, with a named person checking them daily	Reduce asks how few can still answer; Refine asks how the animals’ experience can be better and shorter.
The approval gate: the investigation must be approved under the school’s processes, made from the Australian Code; without approval the investigation does not go ahead	The chain ends at approval — “only then may the investigation with animals go ahead.”

Example 3 — unscaffolded

A student’s survey plan reads: “I will survey my teammates about sleep and reaction time. They have to answer every question because it’s for a grade. I’ll collect names so I can check they’re telling the truth, and I’ll put the class results on the noticeboard with names so it’s interesting.” Audit the plan against the five human-participant principles: name each principle it breaks, what the breach does to the participant, and the fix.

The plan breaks four of the five. **Voluntary participation** — “they have to answer” coerces; the fix is that taking part is a choice, with “you may skip any question, or stop, at any time” written on the form. **Informed consent** — teammates are told nothing before agreeing about what the survey is for or what happens to answers; the fix is a plain-language statement read *before* anyone agrees. **Privacy** — names are collected without need and results are published with names; the fix is codes instead of names and group results only, so no one can be identified. **No harm** — publishing named results about sleep invites embarrassment and teasing; the fix is reporting totals and means only. The fifth principle, ethics approval, is the one the plan observes in form — the teacher checks the plan — but the check exists to catch exactly these failures, so the plan as written would (and should) be sent back. A survey that cannot be run without pressure, names and public results is not a survey with flaws; it is a plan that has not yet earned its participants.

Example 4 — unscaffolded

On a creek excursion, a student picks up a smooth stone “just to look closer”, decides it might be an artefact, pockets it, and the group moves on. Later they say: “it’s safe in my bag — I’ll bring it back next week and put it where I found it.” Explain what is wrong with what happened, and what should have happened, using the steps of this subtopic.

Three things went wrong, in order. First, the object was moved at all: the stop rule exists because an artefact’s position is part of its information and its cultural meaning — once lifted, that is lost even if the stone goes back, because “where I found it” is then a guess, not a fact. Second, the removal may be against the law: in Victoria, Aboriginal cultural heritage is protected, and taking an object that might be heritage is not made lawful by good intentions or a plan to return it. Third, the decision was made by the student alone, when every step of the sequence hands the

decision up: stop, record, tell. What should have happened is the figure's order — stop at the moment it looked interesting, record it with a photo from where they stood and a mark on the map, and tell the teacher straight away, leaving the object where it lay. And now, with the stone in the bag, the same rule applies one day late: tell the teacher immediately, so the find can be reported to the right people and handled properly. The weekly-delay “put it back” plan repeats the original error — a student deciding, alone, what happens to a possible artefact — and reporting fixes what can still be fixed.

Practice questions

Question 1 (scaffolded) In your own words, give the meaning this book works to for: (a) hazard; (b) risk; (c) risk assessment. For (b), say how a large hazard can be a small risk.

Question 2 (scaffolded) On the risk grid: (a) rate a hazard scored likelihood 2, consequence 4, and name its zone; (b) a control halves neither factor but lowers likelihood to 1 and consequence to 2 — what is the new rating and zone?

Question 3 (scaffolded) Match each control to its level of the hierarchy: (a) doing the task with water instead of a flammable solvent; (b) a fume extractor or ventilation over the bench; (c) a written rule that lids stay on containers; (d) splash-proof goggles; (e) a screen between the task and the class; (f) not doing the hazardous task at all.

Question 4 (scaffolded) From the symbols figure: (a) which symbol belongs on a container that catches fire easily? (b) which warns that a chemical burns skin and eyes? (c) a container shows the environmental-hazard symbol — give the rule for disposing of it.

Question 5 (scaffolded) Match each change to a tadpole investigation to its R: (a) using a video and an existing dataset instead of collecting animals; (b) ten tadpoles per group with results shared across the class; (c) returning each tadpole to the creek as soon as it completes metamorphosis; (d) a named person checking housing and food daily.

Question 6 (scaffolded) Match each line of the model consent form to the principle it serves: (a) “this takes about five minutes, once”; (b) “you may skip any part or stop at any time”; (c) “your results are recorded with a code only”; (d) “none expected — you will sit comfortably”; (e) the teacher checking the plan before it runs.

Question 7 A lab task is rated likelihood 4 × consequence 3 before controls. (a) What is the rating and zone? (b) Propose one control from each of three *different* levels of the hierarchy (not PPE), and say which factor — likelihood or consequence — each mainly lowers. (c) Re-rate to likelihood 1 × consequence 2 and state what the move down and left on the grid shows.

Question 8 Explain, using the hierarchy of controls, why “everyone wears goggles” is a weaker plan than “use a dilute solution in small volumes, in trays, and goggles as well”.

Question 9 For a fieldwork day at a sunny, grassy creek with a steep bank: name three hazards from the scene, and for each give one control that keeps people away from the hazard and one that protects the person.

Question 10 “Animals should never be used in any investigation.” “Animals may be used whenever we learn something from it.” Give the strongest case for each view, then state the settled middle position this subtopic works to, and the three Rs it applies.

Question 11 A group wants to investigate whether tadpoles in the school creek develop faster in warmer water. They propose 60 tadpoles, kept until the end of term, in tanks on the window sill. Apply Reduce and Refine to rewrite the proposal in four changes.

Question 12 A survey plan says: “participants will be told the results after the investigation finishes, and must complete all questions.” Rewrite both clauses so the plan serves informed consent and voluntary participation.

Question 13 A group plans to publish its sleep-survey results as a named list “so the data is more interesting”. (a) Name the principle breached and the harm it allows. (b) Rewrite the publishing rule so the data stays useful but nobody can be identified.

Question 14 Put these fieldwork steps in the correct order and name the gate they belong to: (a) follow every condition, including paths to stay on; (b) your teacher checks whether the area has registered Aboriginal cultural

heritage; (c) report anything found by stopping, recording and telling; (d) plan the visit with your teacher; (e) seek permission through the Registered Aboriginal Party for that area if needed; (f) in the field, disturb nothing.

Question 15 A classmate finds a possible artefact and proposes: “let’s take it to school, photograph it properly under the good light, and return it on the next excursion.” Refuse the proposal using two distinct reasons from this subtopic, and give the correct four steps.

Question 16 A company wants to develop an antiseptic product using the banksia example from elaboration VC2S10I02.E02. (a) List, in order, the five steps of the knowledge figure the company must follow. (b) Explain what “no is a complete answer” means for step 2, and which step makes the agreement real before work starts.

Answers

Question 1 (a) a hazard is anything that can cause harm; (b) risk is the chance the hazard causes harm, together with how bad that harm would be; (c) the written process of finding hazards, rating risks and deciding controls before the work starts. Large hazard, small risk: controls make the harm unlikely and less bad. **Question 2** (a) 8, medium; (b) 2, low. **Question 3** (a) substitution; (b) engineering controls; (c) administrative controls; (d) PPE; (e) isolation; (f) elimination. **Question 4** (a) flammable; (b) corrosive; (c) never pour it down the sink or outside. **Question 5** (a) Replace; (b) Reduce; (c) Refine; (d) Refine. **Question 6** (a) informed consent; (b) voluntary participation; (c) privacy; (d) no harm; (e) ethics approval. **Question 7** see solution. **Question 8** see solution. **Question 9** see solution. **Question 10** see solution. **Question 11** see solution. **Question 12** see solution. **Question 13** see solution. **Question 14** d, b, e, a, f, c — Gates 3 and 4 (the permissions sequence). **Question 15** see solution. **Question 16** (a) approach respectfully; free, informed consent; agreed terms in writing; benefits shared; credit given; (b) see solution.

Fully-worked solutions

Question 1 (a) A **hazard** is anything that can cause harm — a chemical, a flame, a slippery floor, a steep bank, a living thing. (b) A **risk** is the chance that the hazard actually causes harm, together with how bad that harm would be. A large hazard can be a small risk once controls make the harm unlikely and less severe — dilute acid in small volumes in trays is still acid, but the splash is unlikely and small. (c) A **risk assessment** is the written process of finding the hazards, rating the risks and deciding the controls, done before the work starts.

Question 2 (a) $2 \times 4 = 8$, which sits in the **medium** zone (5–9). (b) $1 \times 2 = 2$, the **low** zone (1–4).

Question 3 (a) **Substitution** — the hazard is swapped for something safer. (b) **Engineering controls** — the equipment or place is changed. (c) **Administrative controls** — a rule. (d) **Personal protective equipment** — the person is wrapped; the last line. (e) **Isolation** — the hazard is kept away from people. (f) **Elimination** — the hazard is removed entirely.

Question 4 (a) The **flammable** symbol. (b) **Corrosive**. (c) Never pour it down the sink or outside — it harms living things in water and soil; disposal follows the SDS and school rules.

Question 5 (a) **Replace** — the question is answered, or partly answered, without animals. (b) **Reduce** — the smallest number that can still show an answer, with shared results. (c) **Refine** — shorter time involved. (d) **Refine** — welfare checked and housing kept proper.

Question 6 (a) **Informed consent** — exactly what will happen and for how long, before agreeing. (b) **Voluntary participation** — the right to skip or stop at any time. (c) **Privacy** — nothing that could identify a person. (d) **No harm** — honest about risks, even tiny ones. (e) **Ethics approval** — the plan is checked before it runs.

Question 7 (a) $4 \times 3 = 12$, **high** (10–15). (b) Any three different levels, for example: substitution — use a safer chemical (mainly lowers consequence); isolation — a screen between the task and the class (lowers likelihood of anyone nearby being harmed); administrative — a written rule that only the teacher pours the liquid (lowers likelihood). (c) $1 \times 2 = 2$, **low**. The move down and left shows the controls have made the harm less likely *and* less bad — which is what a risk assessment is for.

Question 8 Goggles are PPE, the last line of the hierarchy: they stand between a splash and an eye, but the splash can still happen, and everything goggles miss (skin, the bench, the floor) stays exposed. “Dilute solution, small volumes, trays” works higher in the hierarchy — substitution and isolation — so the hazard itself is weaker and kept away from people: the splash is less likely and less bad, not just intercepted. Goggles *as well* then protect what remains. The hierarchy’s rule is exactly this: fix the hazard before you wrap the person.

Question 9 Any three hazards with one away-control and one person-control each. Example — creek and slippery bank: away-control, keep back from the edge and no climbing; person-control, strong shoes with grip. Sun and UV: away-control, shaded breaks and working in shade where possible; person-control, hat, sunscreen and water. Wildlife in long grass: away-control, keep distance and stay on the mown or cleared path, never touch; person-control, long pants and boots. The group-stays-together rule is administrative and sits above all three.

Question 10 Strongest case for “never”: animals cannot agree to take part, can suffer, and cannot weigh benefit against cost — so the burden of proof sits wholly on the user, and models, simulations and observation make avoidable uses rarer every year. Strongest case for “whenever we learn”: some questions about living bodies — growth, healing, reproduction — cannot yet be answered by a model, and much of medicine’s knowledge of vaccines, anaesthetics and disease came from animal work, so banning all use would cost lives too. The settled middle: animals are used only when the question truly needs them, in the smallest numbers, with the gentlest methods, under proper approval — the three Rs, applied in order: **Replace** where possible, **Reduce** the numbers to the fewest that can answer, **Refine** the method so animals are kept better and harmed less.

Question 11 Four changes, two per R. **Reduce:** (1) cut 60 to the fewest that can still show an answer, with groups sharing and combining results; (2) use tadpoles already present in the creek by observing marked individuals where they live, if that can answer part of the question. **Refine:** (3) return each tadpole to the creek as soon as it completes metamorphosis, not at the end of term; (4) move the tanks off the hot window sill to a shaded, stable spot with proper food, clean water and a named person checking daily. (Underlying all four: the approval gate, before any collection.)

Question 12 First clause: “participants will be told, *before* agreeing, what the survey is about, what they will be asked to do, and what happens to their answers” — consent is informed and comes before agreeing, not after the results exist. Second clause: “taking part is voluntary; you may skip any question, or stop, at any time, without giving a reason” — no question is compulsory, and a grade never depends on taking part.

Question 13 (a) The principle is **privacy**, and the harm is real: a named list about sleep lets readers single out individuals, inviting embarrassment and teasing — the survey itself becomes the harm the no-harm principle forbids. (b) “Results are published as group totals and means only; no names or details that could identify a person are collected or shown.” The data stays useful — totals and means answer the question — while no result can be traced to anyone.

Question 14 d, b, e, a, f, c. Plan with your teacher (d); your teacher checks the area for registered Aboriginal cultural heritage (b); permission is sought through the Registered Aboriginal Party for that area if needed (e); follow every condition (a); in the field, disturb nothing (f); report anything found by stopping, recording and telling (c). The sequence belongs to **Gates 3 and 4** — Country and Place, then approvals and briefing — and note that (c) is also the standing field rule whenever something is found.

Question 15 Two distinct reasons. First, the information reason: an artefact’s position and its place hold information and cultural meaning — removing it destroys that, and a later “return” restores nothing, because the exact position and its record are gone; the photo under school light is worth less than the photo from where it stood. Second, the law and authority reason: in Victoria, Aboriginal cultural heritage is protected by law, and a student taking a possible artefact is not made lawful by good intentions — the decision is not the student’s to make. The correct steps: **STOP** (do not touch, move or collect it), **RECORD** (photo from where you stand, position on the map), **TELL** (your teacher, straight away), **REPORT** (the teacher reports it to the right people, so it is protected where it lies).

Question 16 (a) In order: **approach respectfully** — the community first, with a clear, honest explanation; **free, informed consent** — the community decides for itself; **agreed terms** — what may be used, how and by whom, in writing, before work starts; **benefits shared** — any gain flows back to the community; **credit given** — knowledge holders named and acknowledged, if they wish. (b) “No is a complete answer” means the community may refuse without giving reasons, and no pressure may be applied to change it — a project that continues after a refusal, or that asks again until it gets a yes, has failed step 2 whatever it produces. The step that makes the agreement real before work starts is step 3, **agreed terms in writing**: consent without terms leaves what may be used, how and by whom undecided, and work begun before terms are agreed is work begun without agreement.

Where this leads: the safety and ethics habits taught here — risk assessments, the hierarchy of controls, the three Rs, informed consent, and the checks and protocols on Country and Place — are part of the VCE Science “key science skills” (planning and conducting investigations safely and ethically), reference-only for this book, and they are what SIT-2 assesses when a plan is submitted with its risk assessment and ethics statement. The valid, reproducible half of the plan is Subtopic 3A, equipment and units Subtopic 3C, and the evaluation of the finished investigation Subtopic 5A; the chemistry of the marble-chips investigation used here as a risk-assessment case is Chapter 11. The knowledge

behind elaboration VC2S10I02.E02's examples remains with its holders; this book teaches the obligations around it, not the knowledge itself.

Equipment, calibration, sample size and precision — generating and recording data with digital tools

Chapter 3 — Planning and conducting (Science, Levels 9 and 10)

Subtopic 3C · VC2 Science Inquiry — Planning and conducting

CD code: VC2S10I03 (delivered whole by this subtopic) **Content description (verbatim):** “equipment can be selected and used to generate and record data sets that show precision, including consideration of sample size and using digital tools as appropriate” **Elaboration ids drawn on:** VC2S10I03.E01 · VC2S10I03.E02 · VC2S10I03.E03 · VC2S10I03.E04 · VC2S10I03.E05 (all five elaborations of this CD; none is withheld under the TOC’s D10 ICIP decision) **Achievement standard (extract served):** “They select and use equipment to generate and record data, ensuring the use of suitable sample sizes and assessing the precision of multiple measurement readings.” **Maths interlocks (D11):** VC2M9M04 — absolute, relative and percentage error, used here as absolute, relative and percentage uncertainty; VC2M9ST01/VC2M9ST02 — sampling method and sample size in survey reports, the same ideas behind the sample plots in this subtopic.

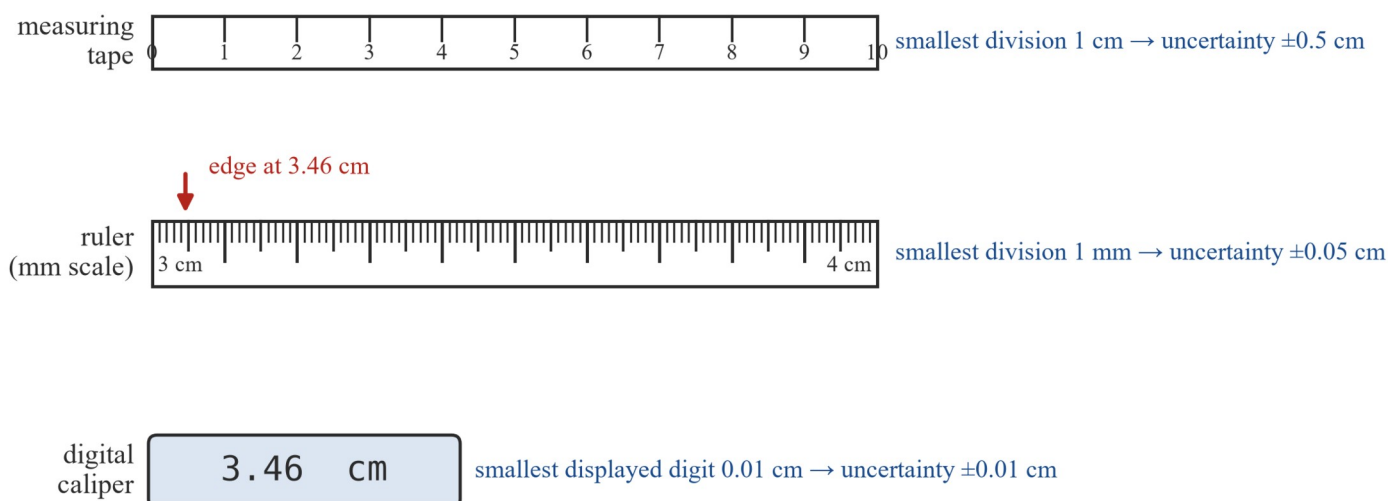
Theory

Every measurement carries an uncertainty. When you measure something, the last digit is never perfect. The scale on the instrument, the way you read it, and the thing being measured all leave a little doubt about the exact value. The **uncertainty** is the range the true value probably lies in, and we write it with the $\pm$ symbol: a length recorded as 3.46 ± 0.05 cm means the true length is probably somewhere between 3.41 cm and 3.51 cm. This subtopic is about choosing equipment that gives a sensible uncertainty, checking that equipment before use, reading it properly, and taking enough readings to trust the result.

All data sets in this subtopic are simulated data for practice — invented numbers that behave like real measurements, clearly labelled as such. No real-world data is quoted.

Choosing equipment: the smallest division sets the uncertainty. The first thing to look at on any scale is its **smallest division** — the gap between two neighbouring marks. On an analogue instrument (one with a scale you read with your eyes), the rule of thumb is that the uncertainty in a single reading is **half the smallest division**, because you can reliably see whether the pointer or edge is closer to one mark or the other. On a digital instrument, the uncertainty is $\pm$ **the smallest displayed digit**: a display that shows 23.6°C can only be read to the nearest 0.1°C .

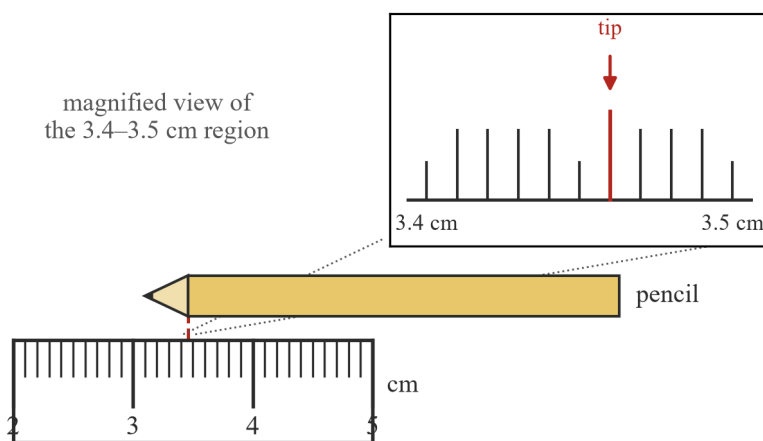
Three instruments measuring the same object — three different uncertainties



Three instruments measuring the same object: a measuring tape with 1 cm divisions giving an uncertainty of ± 0.5 cm, a millimetre ruler giving ± 0.05 cm, and a digital caliper reading 3.46 cm giving ± 0.01 cm.

The figure above shows three instruments measuring the same object. The tape's smallest division is 1 cm, so a reading carries ± 0.5 cm. The ruler's smallest division is 1 mm (0.1 cm), so a reading carries ± 0.05 cm. The caliper displays two decimal places, so its reading carries ± 0.01 cm. Same object, three different uncertainties: the instrument choice decides how precise your data can be, so pick the instrument to match the job. Measuring a room? The tape is fine. Measuring a small metal cube for a density calculation? The caliper.

Estimation: reading between the marks. Good practice with an analogue scale is to estimate **one digit beyond the smallest division**. If the smallest division is 1 mm, you estimate to the nearest 0.1 mm. That estimated digit is the uncertain one — two careful people might record 3.46 cm and 3.47 cm for the same edge, and both would be right, because the last digit is an estimate.



The tip lies between 3.4 cm and 3.5 cm. Estimate one digit beyond the smallest division: length = 3.46 cm.

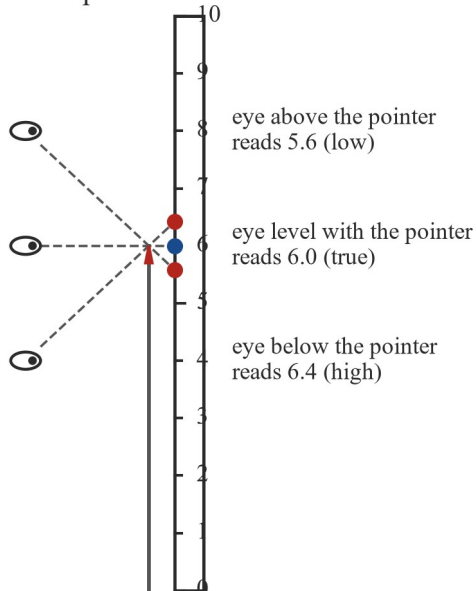
Smallest division 1 mm $\rightarrow$ uncertainty in the reading ± 0.05 cm (half a division).

A pencil tip measured against a millimetre ruler, with a magnified view showing the tip between 3.4 cm and 3.5 cm, closer to the middle, recorded as 3.46 cm with an uncertainty of ± 0.05 cm.

In the figure, the pencil tip sits between the 3.4 cm and 3.5 cm marks, a little past the middle, so the reading is 3.46 cm. The ± 0.05 cm uncertainty says exactly this: the true length is within half a division of the recorded value. If you only ever record to the nearest mark (3.4 cm or 3.5 cm), you throw away information your eye can genuinely give, and your uncertainty doubles to ± 0.1 cm.

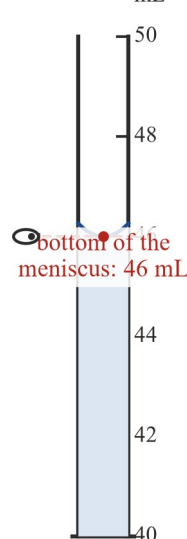
Reading at eye level: parallax and the meniscus. A reading is only as good as the angle you read it from. **Parallax** is the apparent shift of a pointer against a scale when your eye is not level with the pointer. The pointer sits a little in front of the scale, so looking from above or below makes it line up with the wrong mark.

Parallax at a pointer and scale



Only the eye at pointer level reads the true value.

Reading a measuring cylinder



Water climbs the glass a little, so the surface is curved. Read the bottom of the curve, at eye level.

Left: three eye positions reading a pointer against a scale; only the eye level with the pointer reads the true value 6.0, while the eye above reads 5.6 and the eye below reads 6.4. Right: a measuring cylinder with a curved water surface, read at eye level at the bottom of the meniscus, 46 mL.

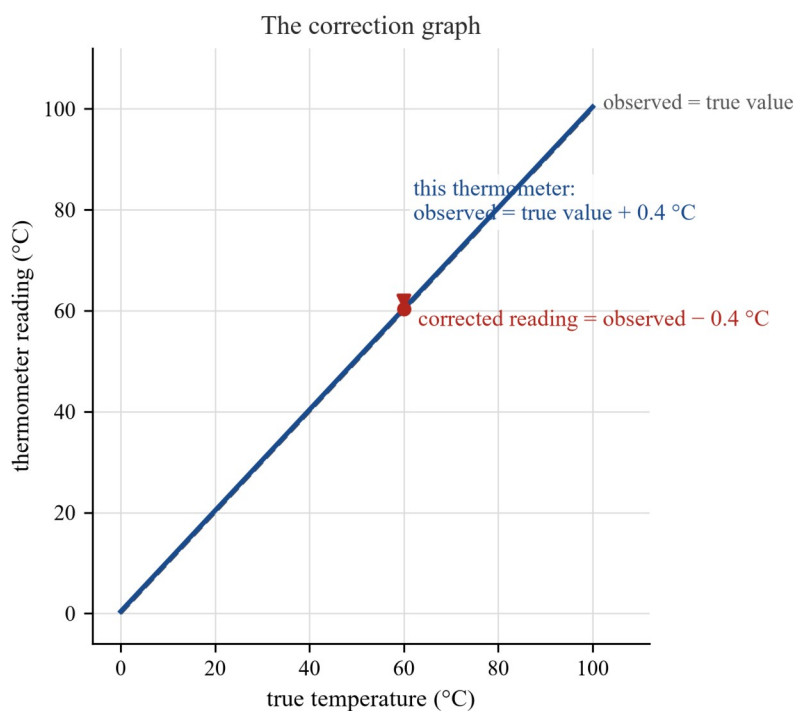
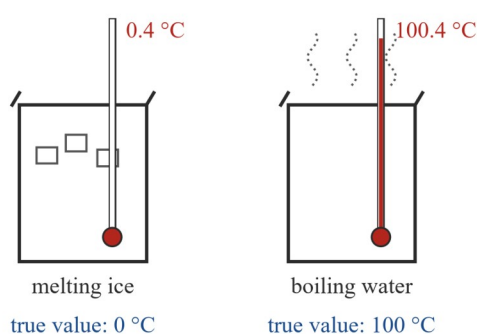
The left panel shows the effect: the eye above the pointer reads 5.6 (too low), the eye below reads 6.4 (too high), and only the eye level with the pointer reads the true 6.0. The rule is simple: **bring your eye level with the pointer or**

scale before you read. Some meters have a small mirror strip behind the pointer — line the pointer up with its own reflection and your eye is in the right place.

Liquids in glass add a second rule. Water climbs the glass a little, so the surface is curved; that curve is called the **meniscus**. Read a measuring cylinder at eye level, at the **bottom of the meniscus**, as in the right panel: 46 mL, not the higher value the edges of the curve would suggest.

Calibration: check the instrument before you trust it. Calibration is a check of an instrument against known values. A thermometer is checked at two **fixed points**: melting ice, which is 0°C by definition, and boiling water, which is 100°C at standard air pressure. A balance is checked against a standard mass — a mass whose value is trusted, kept for checking balances. If the instrument reads the known value, it is calibrated; if it does not, you know its offset and can correct readings or have the instrument adjusted.

Two fixed points check a thermometer
This thermometer reads 0.4°C too high at both points.

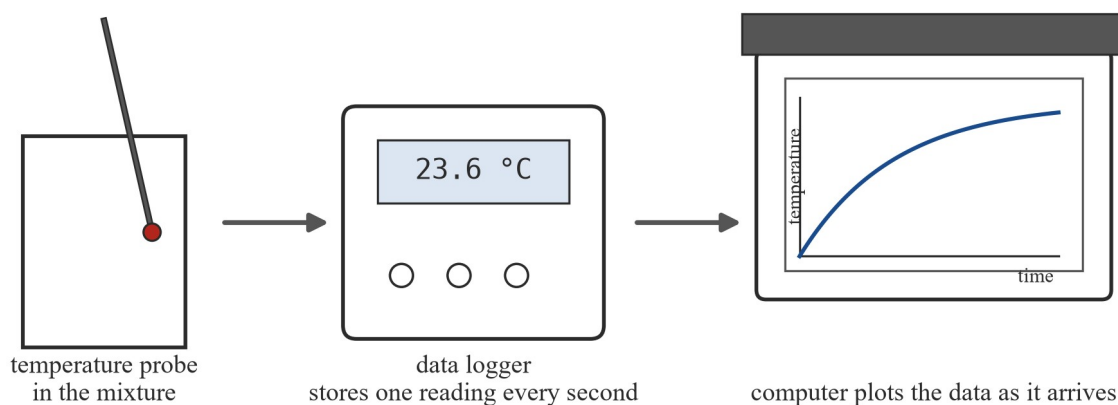


Left: a thermometer checked at two fixed points, reading 0.4°C in melting ice and 100.4°C in boiling water, so it reads 0.4°C too high at both points. Right: a correction graph of thermometer reading against true temperature, with the line observed = true value + 0.4°C and a corrected reading found by subtracting 0.4°C .

The thermometer in the figure reads 0.4°C in ice and 100.4°C in boiling water: 0.4°C too high at both fixed points. Every reading it gives is therefore 0.4°C too high, and the correction graph turns any observed reading into a corrected one: a reading of 60.4°C means a true temperature of 60.0°C . Calibration is not a one-off job. A sensible method **plans for recalibration between uses**: check the instrument at the start of an investigation, again at the end, and again any time it is knocked or its readings start to look odd. A drift caught early is a small correction; a drift caught never is a whole data set you cannot trust.

Digital tools: probes and data loggers. A **probe** (or sensor) connected to a **data logger** measures and stores readings without a person watching the scale. Set the logging rate — one reading every second, say — and the logger records for as long as you need, then the data downloads to a computer or tablet and plots itself.

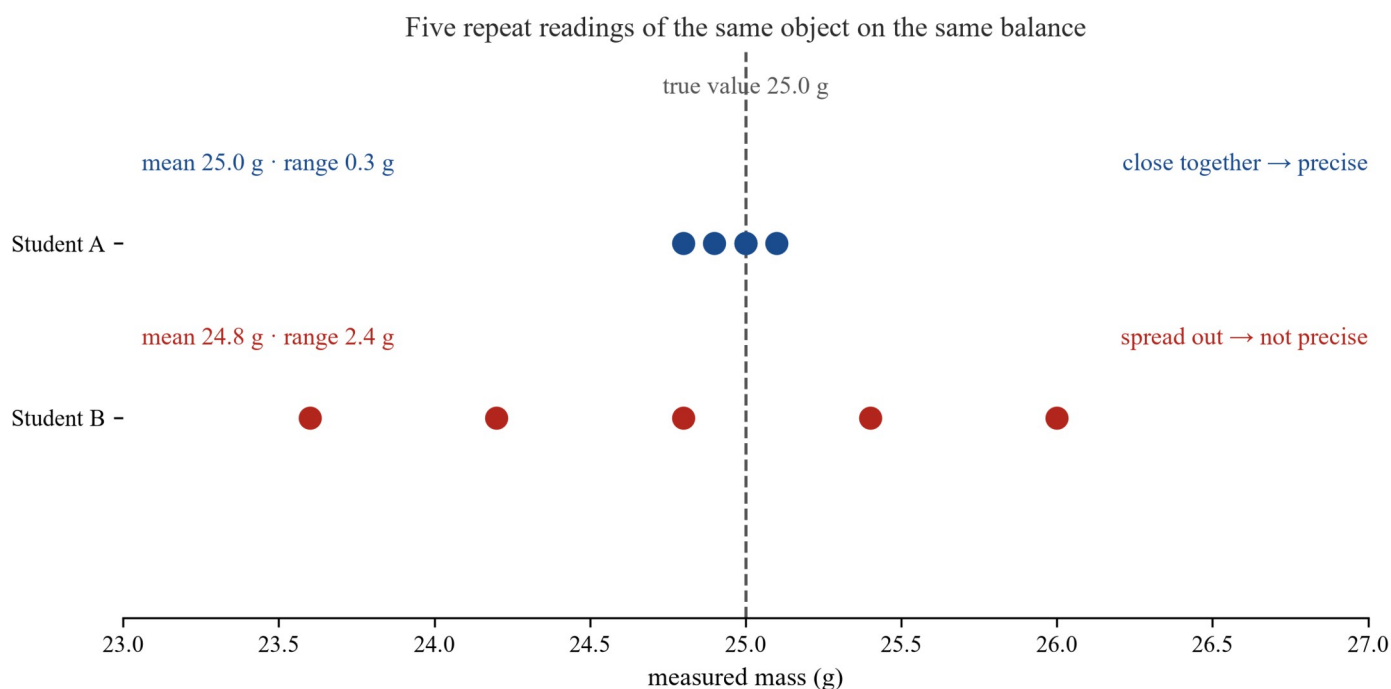
Generating data with digital tools: probe → logger → computer



A temperature probe in a mixture connected to a data logger displaying $23.6\text{ }^{\circ}\text{C}$, connected to a computer that plots temperature against time as the readings arrive.

Digital tools change what is possible in two ways. First, the reading is taken and stored the same way every time, at the same rate, with no parallax and no missed values — a person reading a thermometer every 30 s for ten minutes will miss values, glance late, or read at a slight angle, while a logger will not. Second, a logger makes **larger sample sizes** practical: hundreds of readings that would take a person an afternoon take the logger minutes. Its uncertainty is set by the smallest displayed digit — the $23.6\text{ }^{\circ}\text{C}$ on the display in the figure carries $\pm 0.1\text{ }^{\circ}\text{C}$.

Precision: how close repeat readings are to each other. Measure the same thing several times and the readings will not be identical. **Precision** describes how close repeat readings are to each other: a **precise** set of readings is tightly clustered, an imprecise set is spread out.

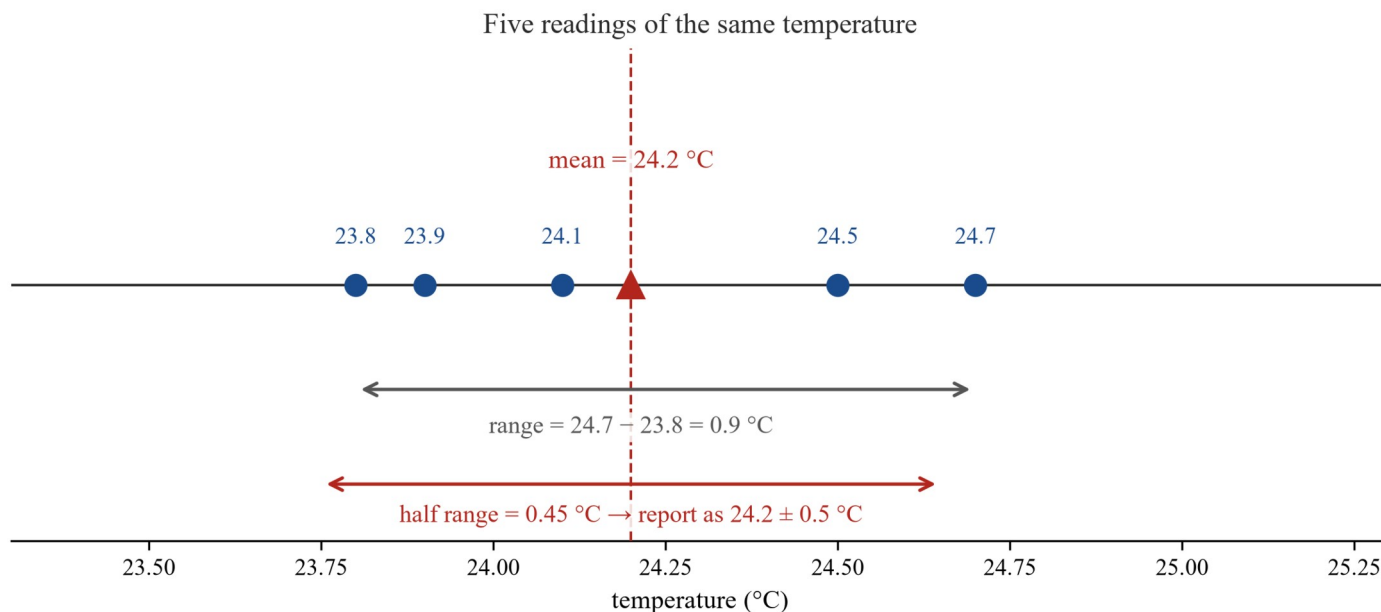


Dot plots of five repeat mass readings by two students on the same balance: Student A's readings cluster between 24.8 g and 25.1 g with a range of 0.3 g (precise), while Student B's readings spread from 23.6 g to 26.0 g with a range of 2.4 g (not precise); the true value 25.0 g is marked.

Both students in the figure measure the same 25.0 g object five times on the same balance. Student A's readings (24.8 , 25.0 , 25.1 , 25.0 , 24.9 g) are close together — precise. Student B's (23.6 , 25.4 , 24.2 , 26.0 , 24.8 g) are spread out — not

precise. The **range** (largest minus smallest) puts a number on the spread: 0.3 g for A, 2.4 g for B. The smaller the range of repeat readings, the more precise the data. (Whether a set of readings is also close to the true value is a separate question — the accuracy side of the story, which Subtopic 5A takes up. Here the job is precision.)

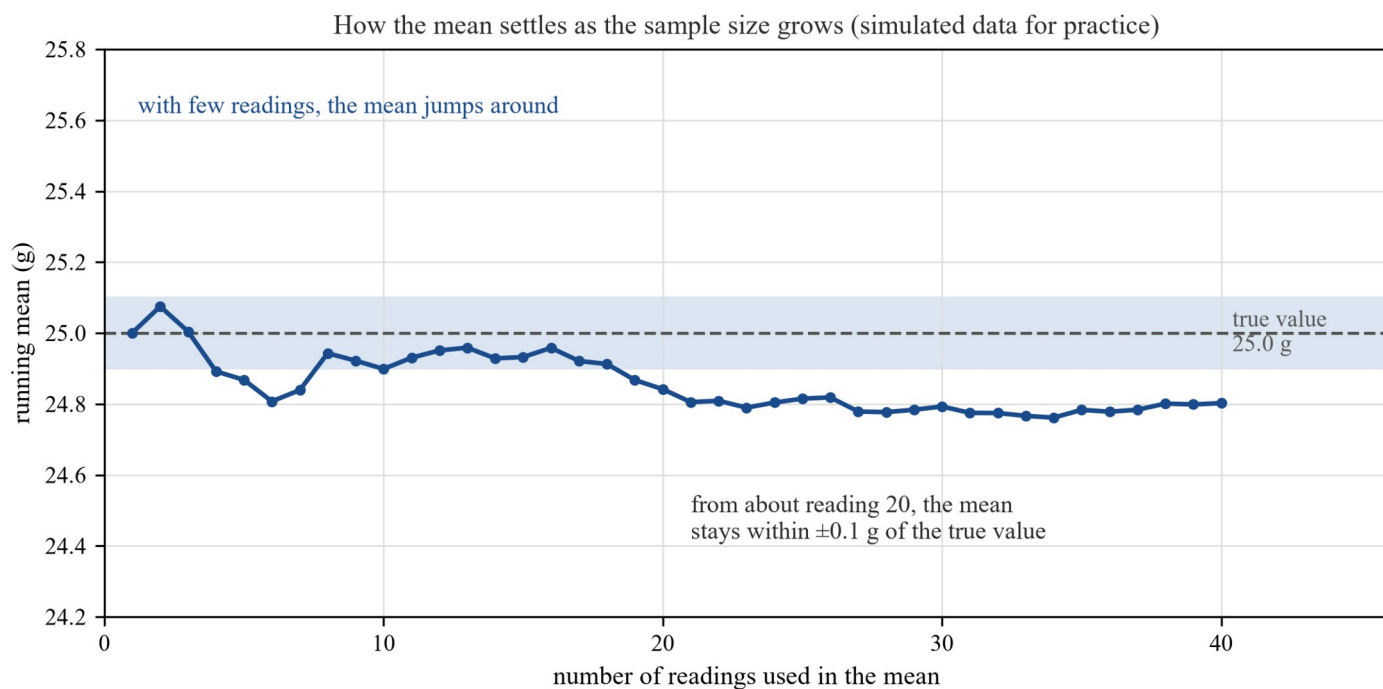
Reporting a set of readings: mean and half-range. With repeat readings, the best single value is the **mean**, and the spread gives the uncertainty: report **mean $\pm$ half the range**.



Five temperature readings on a number line (23.8, 23.9, 24.1, 24.5, 24.7 °C) with the mean 24.2 °C marked, a bracket showing the range 0.9 °C, and a bracket showing the half range 0.45 °C, rounded so the result is reported as 24.2 ± 0.5 °C.

For the five readings in the figure, the mean is $(23.8 + 24.5 + 24.1 + 23.9 + 24.7) \div 5 = 24.2$ °C, and the range is $24.7 - 23.8 = 0.9$ °C. Half the range is 0.45 °C, and since an uncertainty is given to one significant figure, that rounds to 0.5 °C. The result is reported as **24.2 ± 0.5 °C** — the mean with its uncertainty, the way a processed measurement should always be written. Tables and graphs of such data, and the choice of graph type, belong to Subtopic 4A.

Sample size: how many readings are enough? More readings give a more trustworthy mean — but each extra reading helps a little less than the one before. The running mean shows this directly: start with the first reading, then the mean of the first two, then the first three, and so on.

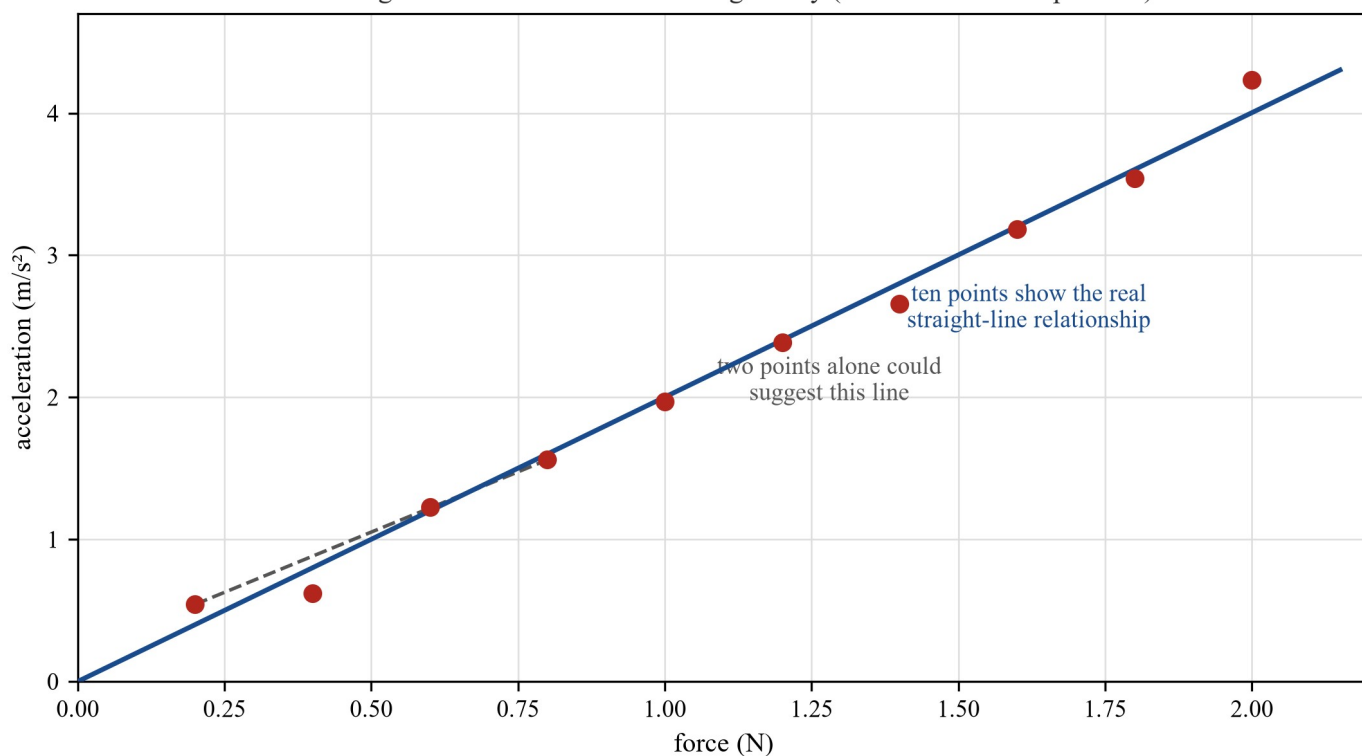


A line graph of the running mean of 40 simulated mass readings against the number of readings used in the mean: the mean jumps around with few readings and settles within 0.1 g of the true value 25.0 g from about reading 20.

With two or three readings, one unusual reading moves the mean a lot. By about reading 20 in the figure, the mean has settled: each new reading is a smaller and smaller fraction of the total, so the line steadies. That settling is the signal that the **sample size** — the number of readings in the data set — is large enough: taking more would change the mean very little. There is no magic number; the right sample size depends on how spread out the readings are and how sure you need to be. The running mean lets you *see* when enough is enough.

Sample size for relationships: more than two points. A sample size is not only about repeat readings; it is also about how many data points you need to see a pattern. Two points always lie on a straight line, so two points can never *show* a straight-line relationship — they can only suggest one.

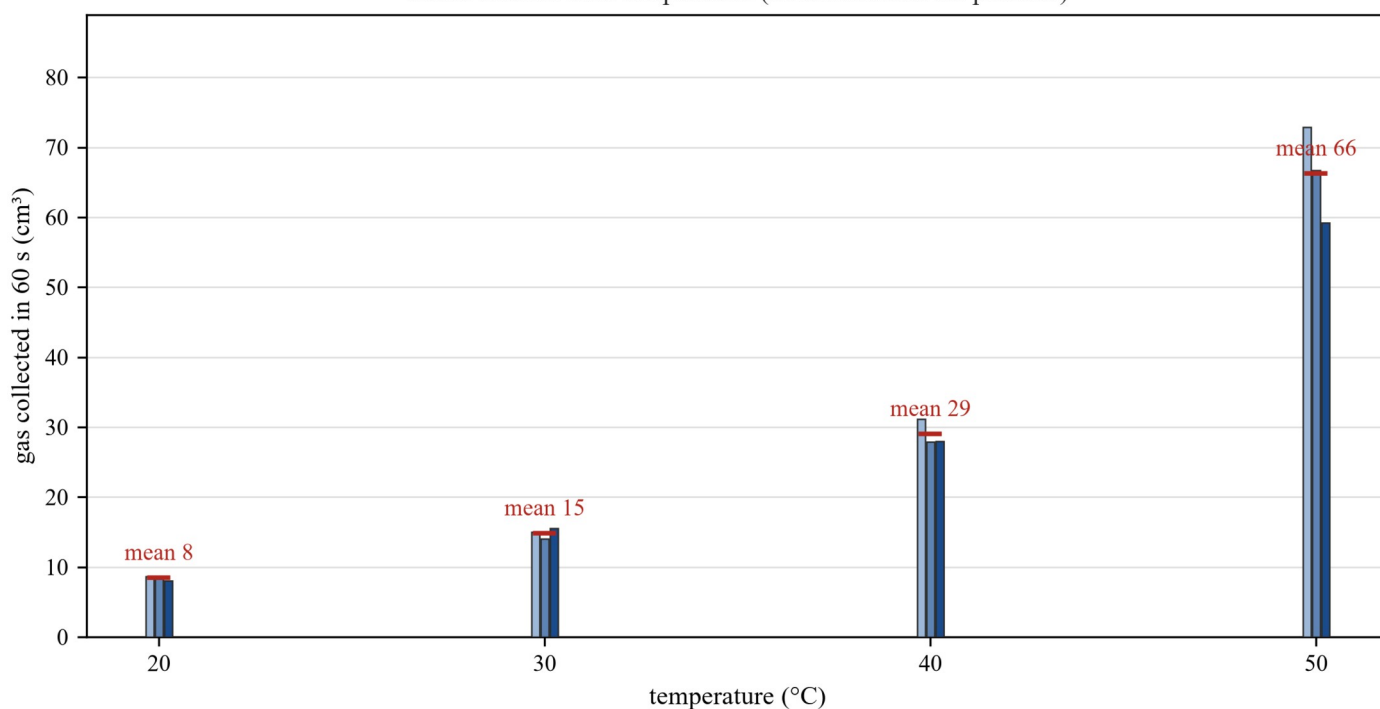
Force against acceleration for a 0.50 kg trolley (simulated data for practice)



A graph of acceleration against force for a 0.50 kg trolley: ten simulated data points scattered close to a straight line through the origin, with a dashed line showing the different, misleading line that two of the points alone could suggest.

The figure shows simulated data for a 0.50 kg trolley pulled with different forces. Take only the two points joined by the dashed line and you would draw the wrong line. Take all ten and the straight line through the origin is obvious: acceleration is proportional to force, and the slope of the line (2.0 m/s^2 per newton) is 1 divided by the mass, $1 \div 0.50 \text{ kg}$. Ten well-spaced points establish the relationship; two cannot. The same logic sets the number of trials in a rate investigation: zinc reacting with dilute sulfuric acid ($\text{Zn} + \text{H}_2\text{SO}_4 \rightarrow \text{ZnSO}_4 + \text{H}_2$) gives hydrogen gas, and the volume collected in a fixed time measures the rate.

Three trials at each temperature (simulated data for practice)

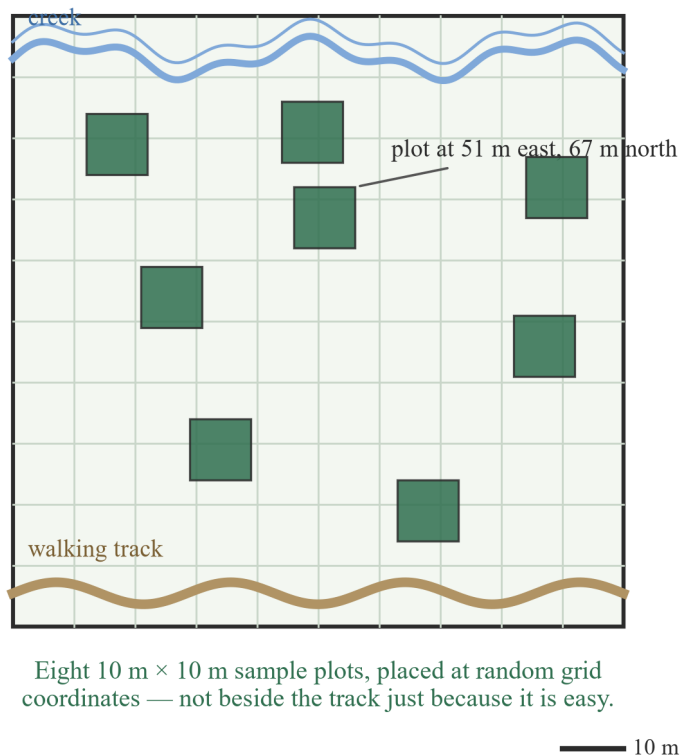


A bar chart of simulated volumes of gas collected in 60 s from zinc in dilute sulfuric acid: three trials at each of 20, 30, 40 and 50 °C, with group means of 8, 15, 29 and 66 cm^3 marked.

Three trials at each temperature give a mean for each group (8, 15, 29 and 66 cm³) and expose the spread within a group, so one odd trial cannot disguise the pattern: hotter acid, faster reaction, with the volume rising more steeply at higher temperature. One trial per temperature would give no feel for the spread at all; three is a workable minimum for a school investigation, and more is better when the spread is large.

Sample size for sampling a whole area or population. Some questions are about a whole system — every tree in a forest, every case of a disease in a population — and measuring all of it is too slow or impossible. Then you measure a **sample** and scale up, and two things decide whether the sample is useful: it must be **large enough**, and it must be **placed fairly**.

100 m × 100 m forest study area, divided into a grid



A 100 m × 100 m forest study area divided into a 10 m grid, with a creek along the top and a walking track along the bottom, and eight 10 m × 10 m sample plots placed at random grid coordinates, one labelled plot at 51 m east, 67 m north.

In the figure, a 100 m × 100 m forest is divided into a grid of 10 m squares, and eight sample plots (each 10 m × 10 m) are placed at **random** grid coordinates — chosen by random numbers, not by walking to a spot that looks nice. Counting the trees in each plot and scaling up gives an estimate for the whole forest; using more plots narrows the uncertainty of that estimate. This is the idea behind investigations of **carbon sequestration** — the storing of carbon in plants and soil — where ecologists count and measure plants in sample plots to estimate the carbon held by a whole ecosystem, and digital tools (a tablet recording plot counts straight into a shared spreadsheet, a drone image supplying extra plots) make larger sample sizes practical. The maths of sampling a population is the same as in VC2M9ST01/VC2M9ST02, where sample size and sampling method decide how much a survey's report can be trusted.

How much data makes a claim secure? The sample-size idea runs right through science. An **epidemic** is a disease spreading quickly through one region; a **pandemic** is the same disease spreading across many countries at once. Health authorities could not declare a pandemic from the case numbers of a single city — that would be a sample of one corner of the picture. The declaration waits on data from many populations, because the size and spread of the sample decide what conclusion the data can carry. In the same way, a scientific **law** — a short statement summarising how nature behaves, like the straight-line force–acceleration relationship above — is only declared when many independent investigations, across many situations, have supplied enough data points. A **theory** — the tested

explanation behind such patterns — likewise rests on a large body of evidence from different lines of investigation, never on one experiment. Small sample, small claim; large sample, strong claim.

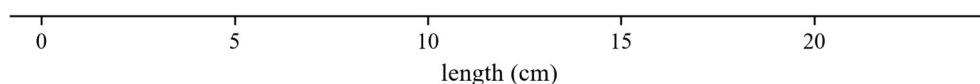
Absolute, relative and percentage uncertainty. The $\pm$ value written with a measurement is the **absolute uncertainty**, in the units of the measurement. The **relative uncertainty** is the absolute uncertainty divided by the measurement itself, and the **percentage uncertainty** is the relative uncertainty $\times 100$. These are the same ideas as absolute, relative and percentage error in VC2M9M04.

The same ruler (± 0.05 cm) measuring two different lengths

long piece: ± 0.05 cm $\rightarrow 0.25\%$
20.0 cm

short piece: ± 0.05 cm $\rightarrow 2.50\%$
2.0 cm

The absolute uncertainty is the same in both cases; the percentage uncertainty is ten times larger for the short piece, because the measurement is ten times smaller.



Two bars on a length axis measured with the same ruler (± 0.05 cm): a 20.0 cm piece with a percentage uncertainty of 0.25%, and a 2.0 cm piece with a percentage uncertainty of 2.5%, ten times larger because the measurement is ten times smaller.

The figure shows why the percentage form is useful for choosing equipment. The same ruler (± 0.05 cm) measures a 20.0 cm piece and a 2.0 cm piece:

- long piece: $(0.05 \div 20.0) \times 100 = 0.25\%$
- short piece: $(0.05 \div 2.0) \times 100 = 2.5\%$

The absolute uncertainty is identical, but the short piece's measurement is ten times less certain in relative terms. If the investigation needs the short length to about 1%, the ruler is the wrong tool and the caliper (± 0.01 cm, giving 0.5% on the short piece) is the right one. Comparing percentage uncertainties is how you decide whether equipment is fit for the job.

Activities

Activity 1 — Three instruments, one object (choosing equipment). Measure the same object — a pencil case or textbook — with a measuring tape, a millimetre ruler and digital calipers. For each instrument, record: the smallest division (or smallest displayed digit), the reading with its $\pm$ uncertainty, and the percentage uncertainty. Compare the three: which instrument gives the smallest percentage uncertainty, and is it the one you would choose for this object? Write one sentence justifying your choice. (*Keep the calipers on the bench with the jaws closed when not measuring, and keep all three instruments away from the bench edge.*)

Activity 2 — Checking a thermometer and a logger (calibration). Place a laboratory thermometer and a data-logger temperature probe in a beaker of ice–water, stir gently, and record both readings once they stop changing. The mixture should read 0 °C: any offset is the instrument's zero error. Then repeat in a beaker of warm tap water (below 50 °C), comparing the thermometer and probe against each other. Record each instrument's offset at the two checks, state the correction you would apply to its readings, and say when in a long investigation you would repeat the checks. (*Ice–water is safe to handle; dry your hands afterwards. Warm water is prepared by the teacher and kept below 50 °C to avoid scalds; the boiling-water fixed point (100 °C) is checked by technical staff, not in class.*)

Activity 3 — How many plots are enough? (sample size). Copy the 100 m × 100 m grid from the sampling figure above onto paper scattered with “trees” (dots) provided by your teacher. Use a table of random digits to pick eight pairs of grid coordinates, place a 10 m × 10 m plot at each, count the trees in each plot, and scale up to estimate the total for the whole forest. Then repeat the estimate using only your first four plots. Compare the four-plot and eight-plot estimates with the true total (your teacher’s value): which is closer, and by how much? Write one sentence on what happens to the estimate as the sample size grows. (*Scissors and paper only — the usual classroom care with scissors.*)

Worked examples

Uncertainties are given to one significant figure, and a processed value is reported to the same decimal place as its uncertainty. All data are simulated data for practice.

Example 1 — scaffolded

A student reads a volume of 36.5 mL from a measuring cylinder whose smallest division is 1 mL. Record the volume with its uncertainty, express the uncertainty as a percentage, and write the volume in litres.

Working	Comment
Uncertainty = ± 0.5 mL	Analogue scale: half the smallest division ($1 \text{ mL} \div 2$).
Volume = 36.5 ± 0.5 mL	Record the measurement with its absolute uncertainty.
$(0.5 \div 36.5) \times 100 = 1.37\% \approx 1.4\%$	Percentage uncertainty = absolute uncertainty $\div$ measurement $\times 100$.
$36.5 \text{ mL} = 36.5 \div 1000 = 0.0365 \text{ L}$	1 L = 1000 mL; the uncertainty scales with it ($\pm 0.0005 \text{ L}$).

The volume is 36.5 ± 0.5 mL, a percentage uncertainty of about 1.4%, or, in litres, 0.0365 ± 0.0005 L. Notice the estimated half-division digit: with 1 mL divisions, the “.5” in 36.5 is the estimated digit.

Example 2 — scaffolded

A student times ten swings of a pendulum four times, getting 15.2 s, 15.5 s, 15.1 s and 15.4 s. Report the mean time for ten swings with its uncertainty, and the time for one swing (the period) with its uncertainty.

Working	Comment
Mean = $(15.2 + 15.5 + 15.1 + 15.4) \div 4 = 61.2 \div 4 = 15.3 \text{ s}$	The mean is the best single value from repeat readings.
Range = $15.5 - 15.1 = 0.4 \text{ s}$; half-range = 0.2 s	The spread of the repeat readings gives the uncertainty.
Ten swings: $15.3 \pm 0.2 \text{ s}$	Report mean $\pm$ half-range.
Period = $15.3 \div 10 = 1.53 \text{ s}$; uncertainty = $0.2 \div 10 = 0.02 \text{ s}$	Dividing the total by 10 divides the uncertainty by 10 too.
Period = $1.53 \pm 0.02 \text{ s}$	The final reported value.

Timing ten swings and dividing by ten makes the period ten times more precise than timing one swing with the same stopwatch — a good reason to measure more than one swing.

Example 3 — unscaffolded

A thermometer reads 0.5°C in melting ice and 100.5°C in boiling water. During an investigation it shows 22.8°C . Find the corrected temperature, and describe a sensible recalibration plan for the investigation.

The thermometer reads 0.5°C too high at the ice point (0.5 instead of 0.0) and 0.5°C too high at the steam point (100.5 instead of 100.0). The offset is the same at both fixed points, so every reading is 0.5°C too high:

$$\text{corrected} = 22.8 - 0.5 = 22.3^\circ\text{C}.$$

A sensible recalibration plan: check the thermometer against the ice point before the investigation starts, check it again at the end, and check again in between if it is knocked or if readings suddenly look odd. Recording the checks in the method means anyone repeating the investigation knows the instrument was trustworthy — and the 0.5 °C correction can be applied to every reading if the offset is stable.

Example 4 — unscaffolded

A group investigates the rate of the reaction between zinc and dilute sulfuric acid at 40 °C by collecting the gas produced in 60 s. Their three trials give 31.2 cm³, 32.8 cm³ and 22.1 cm³. Identify the suspect reading with a reason, give the best mean the group can report from these trials, and say what the group should do next, with a sample-size reason.

The first two trials (31.2 and 32.8 cm³) agree to within 1.6 cm³, but 22.1 cm³ sits more than 9 cm³ below them — far outside the spread of the other two. It is the suspect reading (a leaky bung or a misread scale are typical causes), so the best mean from these trials uses the two that agree:

$$(31.2 + 32.8) \div 2 = 32.0 \text{ cm}^3.$$

But two readings are a small sample: with only two, the group has no way to know which of the three values was wrong, and the mean of two carries a large uncertainty. The group should repeat a fourth trial. If the fourth trial agrees with 31–33 cm³, the mean of the three agreeing values is reported; the sample size grows, the mean settles, and the conclusion rests on more than a pair of readings.

Practice questions

Scaffolded questions

Question 1 A millimetre ruler is used to measure a pencil, and the tip lies exactly on the 14.6 cm mark. (a) State the ruler's smallest division and the uncertainty of a reading with it. (b) Write the pencil's length in the form value $\pm$ uncertainty. (c) Calculate the percentage uncertainty of this measurement.

Question 2 A student measures a card with a millimetre ruler. The edge lies between the 7.3 cm and 7.4 cm marks, about two-tenths of the way from 7.3 toward 7.4. (a) Write the reading the student should record, including the estimated digit. (b) State the uncertainty of the reading. (c) Explain, in one sentence, why the last digit of the reading is called the uncertain digit.

Question 3 The left panel of the parallax figure shows a pointer read from three eye positions. (a) Write down the true value and the two wrong readings shown. (b) Explain, using the idea of a straight line of sight, why the eye above the pointer reads too low. (c) State the two rules for reading a measuring cylinder of water.

Question 4 A thermometer reads 0.4 °C in melting ice and 100.4 °C in boiling water. (a) State the thermometer's offset at the two fixed points. (b) During an investigation the thermometer shows 60.4 °C. Calculate the corrected temperature. (c) Give two occasions in a long investigation when the thermometer should be rechecked.

Question 5 Student A's five readings of a 25.0 g mass are 24.8, 25.0, 25.1, 25.0 and 24.9 g. (a) Calculate the mean of the five readings. (b) Calculate the range. (c) Report the result in the form mean $\pm$ half-range, with the uncertainty to one significant figure.

Question 6 A data logger displays a temperature of 23.6 °C, storing one reading every second. (a) State the uncertainty of one logged reading. (b) Give two reasons a data logger gives a better data set than a person with a thermometer for a ten-minute cooling experiment. (c) The logging rate is changed to one reading every 0.5 s for the same ten minutes. State what happens to the sample size, and one way a larger sample helps.

Unscaffolded questions

Question 7 The same ruler (± 0.05 cm) measures a 20.0 cm piece of timber and a 2.0 cm offcut. (a) Calculate the percentage uncertainty of each measurement. (b) The offcut must be known to about 1%. Explain whether the ruler is fit for the job, and name an instrument from the scales figure that would do better, with its percentage uncertainty on the offcut.

Question 8 A measuring cylinder has a smallest division of 2 mL. A student reads a volume of 27 mL. (a) State the absolute uncertainty of the reading. (b) Calculate the relative uncertainty as a decimal to two significant figures. (c) Calculate the percentage uncertainty.

Question 9 A student times ten swings of a pendulum four times: 15.6 s, 15.8 s, 15.4 s and 15.6 s. (a) Report the mean time for ten swings with its uncertainty. (b) Calculate the period (one swing) with its uncertainty. (c) Explain in one sentence why timing ten swings gives a more precise period than timing one swing.

Question 10 The running-mean figure shows the mean of simulated balance readings settling as the sample size grows. (a) Explain in your own words why the line jumps around at the start. (b) From the figure, estimate how many readings are needed for the mean to stay within ± 0.1 g of the true value. (c) A second balance gives readings with twice the spread. Predict, with a reason, whether its running mean would settle sooner or later than reading 20.

Question 11 The force–acceleration figure shows ten data points for a trolley, with a fitted straight line through the origin of slope 2.0 m/s^2 per newton. (a) Use the line to predict the acceleration when the force is 1.5 N. (b) Explain why ten data points establish the straight-line relationship while two points cannot. (c) Since $F = ma$, the slope of acceleration against force equals $1 \div m$. Calculate the trolley's mass.

Question 12 The zinc–acid figure shows three trials at each of 20, 30, 40 and 50 °C, with group means 8, 15, 29 and 66 cm³ of gas collected in 60 s. (a) Describe the pattern in the group means as the temperature rises. (b) At 50 °C the three trials spread from 59.2 to 72.9 cm³. Explain what this spread tells you about repeating trials at higher temperature. (c) Explain why three trials at each temperature give a stronger conclusion than one trial.

Question 13 In the sampling figure, eight 10 m $\times$ 10 m plots are placed in a 100 m $\times$ 100 m forest. A group counts trees in their eight plots and gets 11, 9, 14, 10, 12, 8, 13 and 11 trees. (a) Show that the eight plots sample 8% of the forest's area. (b) Calculate the mean trees per plot and use it to estimate the total number of trees in the whole forest. (c) Explain why the plots are placed at random grid coordinates instead of beside the walking track where they are easy to reach.

Question 14 Health authorities declare an **epidemic** when a disease spreads quickly through one region, and a **pandemic** when it spreads across many countries at once. (a) Explain why case numbers from a single city are not a large enough sample to declare a pandemic. (b) A scientific **law** is declared only after many independent investigations agree. Explain how this is the same sample-size idea.

Question 15 A balance is to be used for 30 weighings in a long investigation. A checked 100.0 g standard mass is available. (a) Describe a recalibration plan for the balance across the investigation. (b) Midway through, the balance reads 100.6 g for the standard mass. State the offset and the correction to apply to that day's readings.

Question 16 A method needs a length of about 5 cm measured to a percentage uncertainty of 1% or better. (a) Calculate the largest absolute uncertainty that is acceptable (1% of 5 cm). (b) Using the three instruments from the scales figure (± 0.5 cm, ± 0.05 cm, ± 0.01 cm), say which meet the requirement and which fail, with a calculation for each. (c) State which instrument you would choose, with one reason beyond the percentage uncertainty.

Answers

1 (a) 1 mm (0.1 cm); ± 0.05 cm. (b) 14.60 ± 0.05 cm. (c) $(0.05 \div 14.60) \times 100 \approx 0.34\%$.

2 (a) 7.32 cm (7.31–7.34 cm accepted). (b) ± 0.05 cm. (c) It is estimated between marks, so two careful readers may write different digits.

3 (a) True 6.0; eye above reads 5.6; eye below reads 6.4. (b) The line of sight from the high eye through the pointer continues down to a lower mark on the scale behind. (c) Eye level with the surface; read the bottom of the meniscus.

4 (a) $+0.4$ °C at both fixed points. (b) 60.0 °C. (c) Any two of: at the start; at the end; after a knock; if readings look odd.

5 (a) 24.96 g ≈ 25.0 g. (b) 0.3 g. (c) 25.0 ± 0.2 g.

6 (a) ± 0.1 °C. (b) Any two of: readings at even, exact intervals; no missed or late readings; no parallax or reading angle error; hundreds of readings are practical. (c) It doubles ($600 \rightarrow 1200$ readings); a larger sample makes the mean settle and exposes unusual readings.

7 (a) 0.25% and 2.5% . (b) $2.5\% > 1\%$, so the ruler fails; calipers: $(0.01 \div 2.0) \times 100 = 0.5\%$, which meets it.

8 (a) ± 1 mL. (b) $1 \div 27 \approx 0.037$. (c) $\approx 3.7\%$.

9 (a) 15.6 ± 0.2 s. (b) 1.56 ± 0.02 s. (c) The stopwatch's uncertainty is shared over ten swings, so the period's uncertainty is ten times smaller.

10 (a) With few readings, each new reading is a large fraction of the mean, so one unusual reading moves it a lot. (b) About 20 readings. (c) Later — with twice the spread, more readings are needed before the mean settles within the same band.

11 (a) $2.0 \times 1.5 = 3.0$ m/s². (b) Two points always fit a straight line, so they can only suggest one; ten points show whether the pattern really is a line. (c) $m = 1 \div 2.0 = 0.50$ kg.

12 (a) The mean volume increases with temperature, roughly doubling at each 10 °C step up to 40 °C, then more than doubling again from 40 to 50 °C. (b) The spread is wider at 50 °C, so repeats matter more there: the mean of the three trials is what can be trusted, not any single trial. (c) Three trials give a mean and a spread; one trial gives no feel for spread, so an odd value would go unnoticed.

13 (a) $8 \times (10 \times 10) = 800$ m²; $800 \div 10\,000 \times 100 = 8\%$. (b) Mean = $88 \div 8 = 11$ trees per plot; the forest holds 100 plots' worth, so $\approx 11 \times 100 = 1\,100$ trees. (c) Beside the track the sample would over-count the open, disturbed ground and under-count the deep forest; random placement samples the whole area fairly.

14 (a) One city is a small sample that cannot stand for the world's population; the spread across many countries must be shown by data from many populations. (b) A law rests on a large "sample" of evidence — many investigations across many situations — because a small sample of experiments could mislead, just as a small survey could.

15 (a) Zero the balance; check with the standard mass at the start, every ten weighings, and at the end; record each check. (b) Offset $+0.6$ g; subtract 0.6 g from each reading (or have the balance recalibrated and note it in the method).

16 (a) 1% of 5 cm = 0.05 cm. (b) Tape: 0.5 cm > 0.05 cm, fails (10%); ruler: 0.05 cm = 0.05 cm, meets it (1.0%); calipers: 0.01 cm < 0.05 cm, meets it easily (0.2%). (c) The calipers — comfortably inside 1% , leaving margin if the object's ends are hard to judge.

Fully-worked solutions

Question 1 (a) The smallest division is 1 mm = 0.1 cm, so the uncertainty is half of that: ± 0.05 cm. (b) With the tip on the 14.6 cm mark and one estimated digit beyond the smallest division, the length is recorded as **14.60 ± 0.05 cm** — the final 0 is the estimated digit and must be written, because it says the tip is within half a division of 14.60 . (c) Percentage uncertainty = $(0.05 \div 14.60) \times 100 = 0.342\% \approx \mathbf{0.34\%}$.

Question 2 (a) Two-tenths of the way from 7.3 to 7.4 gives an estimated digit of 2: **7.32 cm**. (Readings from 7.31 to 7.34 cm are all honest estimates.) (b) The uncertainty is half the smallest division: **± 0.05 cm**. (c) The last digit is estimated between two marks, so two careful people can record different final digits for the same edge — that digit carries the doubt.

Question 3 (a) The true value (eye level with the pointer) is **6.0**; the eye above reads **5.6** and the eye below reads **6.4**. (b) Light travels in straight lines. From the high eye, the line of sight through the pointer continues downward to the scale behind, so the pointer appears against a lower mark: the reading is too low. (c) Rule 1: put your eye level with the surface of the liquid. Rule 2: read the value at the **bottom of the meniscus**.

Question 4 (a) At the ice point the offset is $0.4 - 0.0 = +0.4$ °C; at the steam point, $100.4 - 100.0 = +0.4$ °C. The offset is the same at both fixed points. (b) Corrected = observed – offset = $60.4 - 0.4 = 60.0$ °C. (c) Any two of: before the investigation starts; at the end of the investigation; after the thermometer is knocked or dropped; part-way through if readings suddenly look odd.

Question 5 (a) Mean = $(24.8 + 25.0 + 25.1 + 25.0 + 24.9) \div 5 = 124.8 \div 5 = 24.96$ g, which rounds to **25.0 g** at one decimal place. (b) Range = $25.1 - 24.8 = 0.3$ g. (c) Half-range = 0.15 g, which rounds to 0.2 g to one significant figure, and the mean is reported to the same decimal place: **25.0 ± 0.2 g**.

Question 6 (a) The smallest displayed digit is 0.1 °C, so one reading carries **± 0.1 °C**. (b) Any two of: the logger records at exact, even intervals for the whole ten minutes; no readings are missed, late or glanced at an angle (no parallax); hundreds of readings are practical, which a person with a stopwatch could not manage; the data downloads and plots itself, avoiding copying mistakes. (c) Ten minutes at one reading per second is 600 readings; at 0.5 s intervals it is **1200 — the sample size doubles**. A larger sample makes the mean settle sooner and makes unusual readings easier to spot.

Question 7 (a) Long piece: $(0.05 \div 20.0) \times 100 = 0.25\%$. Short piece: $(0.05 \div 2.0) \times 100 = 2.5\%$. (b) The offcut's 2.5% is worse than the 1% the job allows, so the ruler is **not** fit for the job. The digital calipers (± 0.01 cm) give $(0.01 \div 2.0) \times 100 = 0.5\%$, which meets the requirement.

Question 8 (a) Half the smallest division: $2 \div 2 = \pm 1$ mL. (b) Relative uncertainty = $1 \div 27 = 0.037$ (two significant figures). (c) Percentage uncertainty = $0.037 \times 100 \approx 3.7\%$.

Question 9 (a) Mean = $(15.6 + 15.8 + 15.4 + 15.6) \div 4 = 62.4 \div 4 = 15.6$ s. Range = $15.8 - 15.4 = 0.4$ s, half-range 0.2 s. Ten swings: **15.6 ± 0.2 s**. (b) Period = $15.6 \div 10 = 1.56$ s; uncertainty = $0.2 \div 10 = 0.02$ s. Period = **1.56 ± 0.02 s**. (c) The stopwatch's reading uncertainty is spread over ten swings, so dividing by ten makes the period's uncertainty ten times smaller than timing a single swing.

Question 10 (a) At the start, each new reading is a large share of the total (the second reading is half of a two-reading mean), so one unusual reading moves the mean a lot. Later, each reading is a tiny share of a large total, so the mean moves very little. (b) From the figure, the mean stays within the ± 0.1 g band from about **20 readings** onward. (c) **Later**. With twice the spread, individual readings sit further from the true value, so more of them are needed before the mean settles within the same ± 0.1 g band.

Question 11 (a) The line gives acceleration = slope $\times$ force = $2.0 \times 1.5 = 3.0$ m/s². (b) Any two points always lie on *some* straight line, so two points can only suggest the relationship; with ten points scattered close to one line (and the dashed two-point line shown to be wrong), the straight-line pattern is established, not assumed. (c) Slope = $1 \div m$, so $m = 1 \div \text{slope} = 1 \div 2.0 = 0.50$ kg.

Question 12 (a) As the temperature rises, the mean volume of gas collected in 60 s increases — 8, 15, 29, then 66 cm³ — roughly doubling at each 10 °C step up to 40 °C, and more than doubling again from 40 to 50 °C. (b) At 50 °C the trials spread over $72.9 - 59.2 = 13.7$ cm³, a much wider spread than at 20 °C ($8.7 - 8.0 = 0.7$ cm³). The wider the spread, the more the mean of the trials — not any single trial — is what can be trusted, so repeats matter most where the spread is large. (c) Three trials give a mean and show the spread, so an odd trial can be seen and the mean settles; one trial shows no spread at all, so an odd value would pass unnoticed and the conclusion would rest on a single number.

Question 13 (a) Each plot is $10 \times 10 = 100$ m²; eight plots sample $8 \times 100 = 800$ m². The forest is $100 \times 100 = 10\,000$ m². $800 \div 10\,000 \times 100 = 8\%$. (b) Mean per plot = $(11 + 9 + 14 + 10 + 12 + 8 + 13 + 11) \div 8 = 88 \div 8 = 11$ trees. The

forest holds $10\,000 \div 100 = 100$ plots' worth of area, so the estimate is $11 \times 100 = \mathbf{1\,100\ trees}$. (c) Beside the track, the plots would sample disturbed, open ground that is not like the deep forest, so the estimate would be skewed one way or the other. Random coordinates give every 10 m square the same chance of being sampled, so the sample represents the whole forest.

Question 14 (a) A pandemic is a claim about the whole world's population. One city's case numbers are a small sample from one corner of that population — they cannot show how widely the disease has spread, so the declaration waits on data from many populations. (b) A law claims that nature behaves a certain way everywhere, so it rests on a large sample of evidence — many independent investigations in many situations. A couple of experiments could mislead, just as a couple of survey responses could; the large sample is what makes the claim secure.

Question 15 (a) Plan: zero the balance with nothing on it; weigh the 100.0 g standard mass before the first weighing, after every ten weighings, and at the end; record every check in the method so the checks travel with the data. (b) The balance reads 100.6 g for a 100.0 g mass: offset **+0.6 g**. Every reading taken that day is 0.6 g too high, so **subtract 0.6 g from each reading** (or have the balance recalibrated, and note the correction in the method).

Question 16 (a) 1% of 5 cm = $0.01 \times 5 = \mathbf{0.05\ cm}$ is the largest acceptable absolute uncertainty. (b) Tape: $\pm 0.5\ cm \rightarrow (0.5 \div 5) \times 100 = 10\%$ — fails. Ruler: $\pm 0.05\ cm \rightarrow (0.05 \div 5) \times 100 = 1.0\%$ — meets the requirement exactly. Calipers: $\pm 0.01\ cm \rightarrow (0.01 \div 5) \times 100 = 0.2\%$ — meets it with room to spare. (c) The **calipers**: at 0.2% they sit comfortably inside the 1% limit, leaving margin for the extra doubt of judging the object's ends, whereas the ruler uses up the whole 1% allowance on the instrument alone.

Where this leads: the accuracy side of measurement — how close a reading is to the true value, random and systematic error, and the full treatment of uncertainty — is VCE Science and is revisited in Subtopic 5A (validity, reproducibility, random and systematic error), reference-only for this book (TOC D9). Sources of error and bias in sampling and observation are taken up in Subtopic 3A; tables, graph choice and descriptive statistics in Subtopic 4A; and the reaction-rate patterns glimpsed here return in the chemical sciences chapters.

Test A

Topic test 1 — How science works, and planning an investigation

Nature and development of science, and science inquiry · Chapters 1–3 · Subtopics 1A, 1B, 2A, 2B, 3A, 3B and 3C · Codes VC2S10H01, VC2S10H02, VC2S10I01, VC2S10I02 and VC2S10I03

Total: 40 marks · Section A (multiple choice): 10 marks · Section B (short/structured response): 30 marks

28 of the 40 marks are science inquiry marks (codes VC2S10I01, VC2S10I02 and VC2S10I03).

Planned for about 50 minutes. This paper is untimed — there is no time to beat.

This is a school-designed paper. It is not a VCAA instrument. Achievement at Levels 9 and 10 is reported against the Victorian Curriculum achievement standard on teacher judgement; the marks here are one source of evidence towards that judgement.

Conditions

- Completion is **untimed**. The 50 minutes is planning guidance only; take as long as you need.
- **Reading support** is permitted. Your teacher may read any question out to you if you need it. Reading is not what is being looked at, and your teacher will not help with the answer.
- **No calculator is required**. Nothing on this paper needs one. No formula sheet. No reading time.
- Answer every question. For multiple-choice questions, circle the letter of the best answer. There is only one best answer for each question.

In this test you will:

- analyse how scientific knowledge is validated — replication, publication, peer review and consensus (Q1, Q2, Q11)
- examine the relationship between science, engineering and technologies (Q3, Q4, Q12)
- identify variables, write hypotheses, and judge claims of proof (Q5, Q6, Q13, Q14)
- identify bias in sampling, and plan safe, ethical investigations — including the checks on Country and Place (Q7, Q8, Q15, Q16)
- generate and record data that show precision — calibration, repeat readings and sample size (Q9, Q10, Q17)

Section A — Multiple choice (10 marks)

Each question is worth 1 mark. Circle the letter of the best answer.

Q1 (1 mark · VC2S10H01) You submit an article saying that a soil test produced a new result. Before publication, other scientists examine your methods, data and conclusion. What is the main purpose of this peer review?

A. To repeat every experiment themselves before publication
B. To prove that your claim will remain correct forever
C. To check whether your methods and reasoning support your conclusion
D. To select conclusions that agree with the most popular theory

Q2 (1 mark · VC2S10H01) You find that two separate laboratories, using the same procedure, get similar results about how temperature affects seed germination. What does this most directly show?

A. The discovery applies only in the first laboratory
B. No further testing can add useful knowledge
C. Temperature must be the only factor affecting germination
D. The result is reproducible, so confidence in it can increase

Q3 (1 mark · VC2S10H02) You study an Australian contribution to Wi-Fi. CSIRO researchers modelled how radio signals reflect from walls and interfere with one another indoors. What engineering improvement could this science support?

A. Making all radio waves travel in one straight path through walls
B. Coding signals so reflected radio waves cause fewer data errors
C. Changing radio signals into sound before they are transmitted
D. Storing internet data only in a single copper wire

Q4 (1 mark · VC2S10H02) You are studying an earthquake using recordings from many ground sensors. Why can a computer be useful?

A. It compares when shaking reached different sensors to help locate where the earthquake began. B. It predicts the exact time of the next earthquake. C. It makes the sensors more precise without needing calibration. D. It shows that every sensor vibration was caused by an earthquake.

Q5 (1 mark · VC2S10I01) You want to investigate whether walking speed changes your breathing. Which question names the variable you change and the result you measure?

A. Why does walking make people breathe faster? B. Do people like walking quickly? C. How does walking speed affect the number of breaths you take in one minute after walking? D. Which walking speed is healthiest?

Q6 (1 mark · VC2S10I01) You investigate crash impact by sending a 0.40-kilogram trolley into the same barrier at several speeds. The trolley has the same bumper and stopping distance each time. Which hypothesis should you test?

A. For this trolley and barrier, a higher impact speed will produce a higher peak force on the sensor. B. Crash force is important in road safety. C. A larger trolley mass will always produce a larger impact force. D. Peak force will make the trolley choose its starting speed.

Q7 (1 mark · VC2S10I02) You are planning to study a native animal on land cared for by a Traditional Owner group. What should you do before finalising your investigation?

A. Assume that public access means no consultation is needed. B. Collect data first and ask for feedback after publishing it. C. Ask only students which parts of the land are important. D. Consult the relevant Traditional Owner group and land manager about access, cultural protocols and appropriate methods.

Q8 (1 mark · VC2S10I02) If a computer simulation can answer your question about animal movement just as well as using live animals, which should you choose?

A. Use live animals because real animals are always more interesting. B. Use the computer simulation instead of live animals. C. Use both methods, even if one is unnecessary. D. Use live animals only when students want to see them.

Q9 (1 mark · VC2S10I03) You check a pH meter in a standard solution labelled pH 7.00, and it displays pH 6.70. What does this check most clearly show?

A. The pH meter has an offset and needs calibration before you test unknown samples. B. The standard solution must have changed to pH 6.70. C. Taking more readings from unknown samples will correct the meter. D. The pH meter is precise because it shows two decimal places.

Q10 (1 mark · VC2S10I03) You collect 40 soil cores for an estuary carbon estimate, but all are taken from just 2 small patches. What is the main sample-size concern?

A. Forty cores must always be exactly enough. B. Two patches make laboratory tests more precise. C. You have many readings but too few spread-out locations to represent the estuary. D. Each core automatically represents the entire estuary.

Section B — Short/structured response (30 marks)

Q11 (4 marks · VC2S10H01) Until the 1980s the settled view was that stomach ulcers came from stress and acid. In 1979 the Perth doctor Robin Warren noticed curved bacteria in stomach samples, and with Barry Marshall he argued that the bacteria caused ulcers. The claim was dismissed and early reports were rejected. Marshall swallowed a culture of the bacteria himself and developed an inflammation of the stomach lining. What changed minds was what came after: other labs grew the same bacteria, found it in ulcer patients, and trials showed that antibiotics cured ulcers. In 1994 a major health panel backed antibiotic treatment for ulcers.

Source: NobelPrize.org, “The Nobel Prize in Physiology or Medicine 2005 — Summary”, 2005, nobelprize.org — retrieved 2026-08-21. B. J. Marshall and J. R. Warren, *The Lancet*, vol. 323, pp. 1311–1315, 1984 — retrieved 2026-08-21.

(a) Give one reason the bacteria claim was dismissed at first. (1 mark)

- (b) Explain why Marshall's self-experiment, on its own, could not validate the claim. (1 mark)
- (c) Name the check of the validation cycle that the other labs supplied. (1 mark)
- (d) State the event in the record that shows a consensus had been reached. (1 mark)

Q12 (4 marks · VC2S10H02) For decades the standard crash-test dummy was built to the 50th-percentile man — the height and weight of a middle-sized man — and it has been the reference dummy since 1976. A smaller “female” dummy added in 2003 was a scaled-down version of the male design, built to a small body rather than to the average woman. A 2019 study of real crashes found that, in front-on crashes, belted women were at greater risk of injury than belted men in similar crashes at similar speeds. In 2023 a Swedish team led by Dr Astrid Linder presented the first dummy built to the average female body.

Source: J. Forman et al., “Automobile injury trends in the contemporary fleet: Belted occupants in frontal collisions”, Traffic Injury Prevention, 2019. <https://doi.org/10.1080/15389588.2019.1630825> — retrieved 2026-08-21. A. Carlsson, J. Davidsson, A. Linder and M. Y. Svensson, “Design and Evaluation of the Initial 50th Percentile Female Prototype Rear Impact Dummy, BioRID P50F”, Frontiers in Bioengineering and Biotechnology, 2021. <https://doi.org/10.3389/fbioe.2021.687058> — retrieved 2026-08-21.

- (a) Explain how the male-shaped standard dummy biased what crash science knew. (2 marks)
- (b) Explain how the average-female dummy improves the science of crashes, not just the engineering of dummies. (2 marks)

SCENARIO — invented for skills practice.

Q13 (4 marks · VC2S10I01) A group drops a ball from different heights and measures how high it bounces back after each drop.

- (a) Name the independent variable and the dependent variable in this investigation. (2 marks)
- (b) State one variable the group must keep the same, and say why. (1 mark)
- (c) Write a hypothesis for this investigation. Your hypothesis must name a relationship between the two variables and give a reason. (1 mark)

Q14 (2 marks · VC2S10I01)

- (a) State one difference between a hypothesis and a reasoned prediction. (1 mark)
- (b) A student writes: “Our results matched our prediction, so we have proved that the hypothesis is true.” Evaluate this claim. (1 mark)

SCENARIO — invented for skills practice.

Q15 (6 marks · VC2S10I02) A group wants to answer the question: “How do students most often travel to school?” Their plan is to stand at the school's front gate on one Monday morning between 8:15 and 8:30 am and record the first 20 students they see arriving.

- (a) State the meaning this book works to for *bias*. (1 mark)
- (b) Identify two features of the plan that could bias the result, and explain each. (2 marks)
- (c) Give one change to the plan that would make the sample fairer. (1 mark)
- (d) The group says its result answers the question for the whole school. Using the word *valid*, evaluate whether the result can carry that claim. (2 marks)

Q16 (4 marks · VC2S10I02) A class plans a field investigation at a local creek. The bank is steep and slippery, the ground is uneven, and the area is on Country that may contain registered Aboriginal cultural heritage.

- (a) Name two hazards at this site, and give one control for each. (2 marks)
- (b) State who checks whether the area has registered Aboriginal cultural heritage, and when that check happens. (1 mark)
- (c) If permission is needed, state from whom it should be sought. (1 mark)

All readings below are simulated data for practice.

Q17 (6 marks · VC2S10I03) A group weighs the same sample four times on a balance. The readings are 45.2 g, 45.1 g, 45.3 g and 46.4 g.

- (a) Identify the suspect reading and give a reason. (1 mark)
- (b) Calculate the best mean the group can report from these trials. (1 mark)
- (c) A group member says: “Two agreeing readings are enough — we can stop there.” Evaluate the claim using the idea of sample size. (1 mark)
- (d) Before the weighing, the balance read 0.2 g with nothing on the pan. State the name of this kind of error, and one thing the group should do about it. (2 marks)
- (e) The three agreeing readings run from 45.1 g to 45.3 g. State what the size of this range shows about the data. (1 mark)

Test B

Topic test 1 — How science works, and planning an investigation

Nature and development of science, and science inquiry · Chapters 1–3 · Subtopics 1A, 1B, 2A, 2B, 3A, 3B and 3C · Codes VC2S10H01, VC2S10H02, VC2S10I01, VC2S10I02 and VC2S10I03

Total: 40 marks · Section A (multiple choice): 10 marks · Section B (short/structured response): 30 marks

28 of the 40 marks are science inquiry marks (codes VC2S10I01, VC2S10I02 and VC2S10I03).

Planned for about 50 minutes. This paper is untimed — there is no time to beat.

This is a school-designed paper. It is not a VCAA instrument. Achievement at Levels 9 and 10 is reported against the Victorian Curriculum achievement standard on teacher judgement; the marks here are one source of evidence towards that judgement.

Conditions

- Completion is **untimed**. The 50 minutes is planning guidance only; take as long as you need.
- **Reading support** is permitted. Your teacher may read any question out to you if you need it. Reading is not what is being looked at, and your teacher will not help with the answer.
- **No calculator is required**. Nothing on this paper needs one. No formula sheet. No reading time.
- Answer every question. For multiple-choice questions, circle the letter of the best answer. There is only one best answer for each question.

In this test you will:

- analyse how scientific knowledge is validated — replication, publication, peer review and consensus (Q1, Q2, Q11)
- examine the relationship between science, engineering and technologies (Q3, Q4, Q12)
- identify variables, write reasoned predictions, and judge claims of proof and failure (Q5, Q6, Q13, Q14)
- plan safe, ethical investigations — consent, welfare rules and the hierarchy of controls (Q7, Q8, Q15, Q16)
- generate and record data that show precision — uncertainty, repeat readings and sample size (Q9, Q10, Q17)

Section A — Multiple choice (10 marks)

Each question is worth 1 mark. Circle the letter of the best answer.

Q1 (1 mark · VC2S10H01) You are learning that scientists use several climate models to study global warming. Which statement best explains how scientific consensus can form when models give somewhat different results?

A. Scientific consensus can form only when every climate model gives exactly the same result. B. Many measurements and studies point to the same overall conclusion, and other scientists repeat and review the work. C. Scientists choose the model that predicts the largest temperature increase and use only that model. D. Scientific consensus is decided by a vote, even if the evidence is limited.

Q2 (1 mark · VC2S10H01) You conduct an experiment that does not support your prediction. What action helps science most?

A. Report only the trials that supported your prediction B. Reject all earlier scientific ideas without checking your method C. Record the result, check your method and share reliable findings D. Change the data so that it matches your prediction

Q3 (1 mark · VC2S10H02) You are reading about cochlear implants developed from Australian research. Once these devices could deliver controlled electrical signals to the auditory nerve, how could they also help scientific research?

A. They could let researchers test how different electrical patterns are heard as sound B. They could show that electrical signals travel through air from the implant to the brain C. They could replace every hearing test with an electrical measurement D. They could show that louder electrical signals always create clearer sound

Q4 (1 mark · VC2S10H02) You use a computer model to test thousands of possible shapes for a protein molecule. What does faster computing most directly allow you to do?

A. Remove the need for laboratory testing of the protein. B. Make a protein in a real sample fold into the predicted shape. C. Guarantee that the model's rules and assumptions are correct. D. Evaluate more possible shapes in the same time and suggest candidates for laboratory study.

Q5 (1 mark · VC2S10I01) You will place a lamp at different distances from pondweed and count bubbles produced each minute. Which question clearly links the variable changed with the result measured?

A. Why does pondweed need light? B. How does distance from a lamp affect the number of bubbles produced by pondweed each minute? C. How does bubble count determine lamp distance? D. How do lamp distance and water temperature affect bubble production?

Q6 (1 mark · VC2S10I01) You release the same toy car from different heights on a ramp onto the same carpet. What hypothesis should you test?

A. If the release height increases, the car will travel farther across the carpet after leaving the ramp. B. The distance travelled will decide which ramp height you use. C. How does ramp height affect the distance travelled? D. If the carpet is rougher, the car will travel farther.

Q7 (1 mark · VC2S10I02) Free, prior and informed consent in research means:

A. Researchers inform communities only after the study has been published B. Verbal agreement alone is enough without explaining the research C. Communities agree to research after full understanding before it starts D. Researchers may proceed if no community member raises an objection

Q8 (1 mark · VC2S10I02) After three handling trials, a lizard repeatedly freezes and struggles to escape, but your plan originally included two more trials. What is the best action?

A. Ask another student to hold the lizard more tightly. B. Finish the trials more quickly so the data set is complete. C. Continue because all animals must complete the same number of trials. D. Stop that lizard's trials under the planned welfare rule and get advice before continuing.

Q9 (1 mark · VC2S10I03) You need to measure mass changes of about 0.8 grams in dissolving tablets. Balance X displays to 0.02 grams but its calibration status is unknown. Balance Y displays to 0.2 grams and was checked today. What is the best action?

A. Use both balances and average their results without checking balance X. B. Check or calibrate balance X with a known mass before deciding which balance to use. C. Use balance X immediately because more decimal places always mean better data. D. Use balance Y automatically, even if balance X could be calibrated.

Q10 (1 mark · VC2S10I03) You have measured soil carbon at 2 randomly selected sites in a mangrove area. Before you make a claim about the whole area, what should you do?

A. Choose only sites closest to the walking track. B. Repeat the calculation using the same 2 results. C. Sample more randomly selected sites using the same method. D. Discard one result to make the data simpler.

Section B — Short/structured response (30 marks)

Q11 (4 marks · VC2S10H01) In 1952 Rosalind Franklin and Raymond Gosling sent X-rays through a fibre of DNA and recorded the pattern of spots now known as Photo 51. The photograph was shown without Franklin's knowledge to James Watson, who was building models of DNA with Francis Crick. In 1953 Watson and Crick published their double-helix model in *Nature* — and in the same issue, Franklin and Gosling published their own paper, the X-ray evidence any model had to fit. Franklin died in 1958; the 1962 Nobel Prize went to Watson, Crick and Maurice Wilkins.

Source: J. D. Watson and F. H. C. Crick, "Molecular Structure of Nucleic Acids", *Nature*, vol. 171, pp. 737–738, 1953 — retrieved 2026-08-21. R. G. Franklin and R. G. Gosling, "Molecular Configuration in Sodium Thymonucleate", *Nature*, vol. 171, pp. 740–741, 1953 — retrieved 2026-08-21. NobelPrize.org, "The Nobel Prize in Physiology or Medicine 1962 — Summary", 1962, nobelprize.org — retrieved 2026-08-21.

- (a) State two things the pattern of spots showed about the shape of DNA. (2 marks)
- (b) Franklin and Gosling's paper appeared in the same issue of *Nature* as Watson and Crick's model. Explain why publishing the evidence beside the model matters. (1 mark)
- (c) State one reason the published record matters when people ask who contributed what. (1 mark)

Q12 (4 marks · VC2S10H02) ASKAP stands at the CSIRO Murchison Radio-astronomy Observatory in the outback of Western Australia. Its 36 antennas are each a dish 12 m wide. No single dish sees much, but a central computer combines the signals from all 36 so the array works as one much larger telescope, surveying the radio sky faster than any dish could.

Source: D. McConnell et al., "The Rapid ASKAP Continuum Survey I: Design and first results", *Publications of the Astronomical Society of Australia*, 2020. <https://doi.org/10.1017/pasa.2020.41> — retrieved 2026-08-21.

- (a) State what joins the 36 antennas into one telescope. (1 mark)
- (b) State what the array can do that no single 12 m dish can. (1 mark)
- (c) State which arrow of the content description this case best shows — "advances in technologies have enabled advances in science", or "science has contributed to developments in technologies and engineering" — and justify your choice in one sentence. (2 marks)

Q13 (4 marks · VC2S10I01) A group investigates the question: "How does the intensity of exercise affect heart rate?"

- (a) Name the independent variable and the dependent variable in this question. (2 marks)
- (b) Give one way the group could make "intensity of exercise" measurable — name the levels or values they would use. (1 mark)
- (c) State one variable the group should keep the same, and say why. (1 mark)

Q14 (2 marks · VC2S10I01)

- (a) A student writes: "The results did not support my hypothesis, so the investigation was a failure." Evaluate this claim. (1 mark)
- (b) State what should happen to a hypothesis when results do not support it. (1 mark)

SCENARIO — invented for skills practice. All counts are simulated data for practice.

Q15 (6 marks · VC2S10I02) A class asks: "How many vehicles pass the school gate between 8:15 and 8:30 am?" Two groups each run the survey three times. Group 1's method fixes every rule: the counting line is between the gate post and the power pole; count only vehicles moving towards the school; count bicycles but not parked vehicles; one observer stands at the same position each time. Group 2's method says only: "count traffic near the gate for fifteen minutes."

Group 1's three counts: 86, 90, 88. Group 2's three counts: 86, 42, 128.

- (a) Which group's results are reproducible? Use the data to justify your answer. (2 marks)
- (b) State the meaning this book works to for *reproducible*. (1 mark)
- (c) Identify two rules in Group 1's method that make its count reproducible, and explain each. (2 marks)
- (d) Even with Group 1's method, the counts come from three mornings only. State one reason the result might still not represent a typical school week. (1 mark)

Q16 (4 marks · VC2S10I02) A group plans to crush limestone chips with a hammer and then add dilute acid to the powder to watch the reaction.

- (a) Identify two hazards in this task. (2 marks)
- (b) For one of those hazards, give a control that is *not* personal protective equipment (PPE), and name its level in the hierarchy of controls. (1 mark)
- (c) Explain why "everyone wears goggles" alone is the weakest way to control a risk. (1 mark)

All readings below are simulated data for practice.

Q17 (6 marks · VC2S10I03) A student measures the width of a card with a millimetre ruler.

- (a) State the uncertainty of one reading with a millimetre ruler. *(1 mark)*
- (b) The edge of the card lies between the 5.6 cm and 5.7 cm marks, a little past the middle. Write the reading the student should record, including the estimated digit. *(1 mark)*
- (c) Five students each measure the same edge and record 5.65 cm, 5.66 cm, 5.64 cm, 5.65 cm and 5.67 cm. State what the closeness of these readings shows about them. *(1 mark)*
- (d) Report the result of the five readings in the form mean $\pm$ half-range, with the uncertainty to one significant figure. Show your working. *(2 marks)*
- (e) Explain why two data points can never establish a straight-line relationship between two variables. *(1 mark)*

Marking guide

This guide is written for a teacher marker — no specialist science background beyond Chapters 1–3 of this book is needed. Award a mark when the student’s answer matches an acceptable response below or says the same idea in other words. Do not award half marks. The marks are a record that evidence was seen; they are not used to rank students.

Test A — Section A — Multiple choice

Q	Answer	Code	Source
Q1	C	VC2S10H01	bank item VC2S10H01.E01 .S01.D1.V03
Q2	D	VC2S10H01	bank item VC2S10H01.E01 .S01.D1.V12
Q3	B	VC2S10H02	bank item VC2S10H02.E01 .S01.D2.V05
Q4	A	VC2S10H02	bank item VC2S10H02.E02 .S01.D1.V03
Q5	C	VC2S10I01	bank item VC2S10I01.E01 .S01.D1.V02
Q6	A	VC2S10I01	bank item VC2S10I01.E02 .S01.D2.V02
Q7	D	VC2S10I02	bank item VC2S10I02.E01 .S01.D2.V11
Q8	B	VC2S10I02	bank item VC2S10I02.E01 .S01.D1.V16
Q9	A	VC2S10I03	bank item VC2S10I03.E01 .S01.D1.V02
Q10	C	VC2S10I03	bank item VC2S10I03.E02 .S01.D2.V10

All ten items are drawn from the MCQ bank under the codes assessed on this paper; none is drawn from a neighbouring code. Items mix easy (D1) and medium (D2) bank difficulties. Items are printed with their stem and options word for word; only the positions of the options have changed for this paper, so that the correct answers are spread across A, B, C and D (the printed positions are $A \times 3$, $B \times 2$, $C \times 3$, $D \times 2$).

Why each answer is correct, in brief:

- **Q1 C** — peer review is a check before publication: expert reviewers ask whether the method is sound, the data support the claims and the conclusions follow. It does not prove anything forever, repeat every experiment, or favour popular theories.
- **Q2 D** — the same procedure giving similar results in two separate laboratories is reproducibility, and reproducible results earn growing confidence.
- **Q3 B** — CSIRO’s radio-astronomy maths of untangling indoor echoes (the fast Fourier transform) supported coding signals so reflected waves cause fewer data errors — the basis of fast indoor wi-fi.
- **Q4 A** — a computer compares when the shaking reached different sensors and works backwards to the start point; it does not predict earthquakes or make sensors precise.
- **Q5 C** — the question names the variable changed (walking speed) and the result measured (breaths in one minute); “why”, “healthiest” and “like” are not measurable as written.
- **Q6 A** — the hypothesis names the two variables actually in the test (impact speed and peak force) and is testable; mass is not being changed here, and the other options are not testable statements.

- **Q7 D** — before finalising an investigation on land cared for by a Traditional Owner group, the relevant Traditional Owner group and land manager are consulted about access, cultural protocols and appropriate methods.
- **Q8 B** — the three Rs are applied in order: Replace first. If a simulation can answer the question just as well, it is used instead of live animals.
- **Q9 A** — a check against a known standard that reads differently shows the meter has an offset; it must be calibrated before unknown samples are tested.
- **Q10 C** — forty readings from two patches is a large count but a poorly placed sample; a sample must be large enough *and* spread fairly to represent the whole estuary.

Test A — Section B — Short/structured response

Q11 (4 marks · VC2S10H01) (a) any one of: it contradicted the settled view that ulcers came from stress and acid; early reports were rejected. 1 mark (b) one person is a sample of one — with no repeat by another group (and no control for comparison) the swallow cannot be replication. 1 mark Accept any response that says a single person's result is not enough. (c) replication. 1 mark (d) the 1994 decision by a major health panel backing antibiotic treatment for ulcers. 1 mark Accept “the 2005 Nobel Prize to Marshall and Warren” — both events sit on the consensus side of the cycle; the panel is the stronger answer because a panel backs a treatment only once the reviewed and replicated evidence has piled up. Marker note: the case is taught in Subtopic 1A (elaboration VC2S10H01 . E01); the source line above the question carries the real record.

Q12 (4 marks · VC2S10H02) (a) 2 marks — 1 mark for the bias: while the standard dummy was built to the 50th-percentile man, crash tests measured how male bodies respond, so the injury data carried that shape; 1 mark for the cost: the science of crashes never measured female bodies properly, and the 2019 finding — belted women at greater risk of injury than belted men in similar front-on crashes — is exactly the kind of result male-shaped testing could not see. (b) 2 marks — 1 mark for the instrument link: science is built on what its instruments measure, so a new instrument changes what can be measured; 1 mark for the consequence: with an average-female dummy, crash tests produce injury data for female bodies for the first time, so the science of crashes — what cars should protect against — changes, not only the dummy. Marker note: this is elaboration VC2S10H02 . E05; the case is taught in Subtopic 1B with the same two sources.

Q13 (4 marks · VC2S10I01) (a) independent variable: the drop height; dependent variable: the bounce height (how high the ball bounces back). 2 marks (1 mark each; the name and the role must match). (b) any one with its reason, for example: the same ball (a different ball bounces differently); the same surface (a different floor changes the bounce); released, not thrown (a throw adds speed the test is not about). 1 mark The variable and a reason must both be there. (c) for example: “The greater the drop height, the greater the bounce height, because the ball is moving faster when it hits the floor.” 1 mark Accept any statement that names both variables, says how they change together, and gives a reason. A bare guess (“I reckon it bounces higher”) does not earn the mark — a hypothesis carries a reason.

Q14 (2 marks · VC2S10I01) (a) any one correct distinction: a hypothesis is a proposed answer written generally, before the data are in; a reasoned prediction says what will be observed in one particular investigation if the hypothesis holds. 1 mark (b) the claim is too strong: results that match the prediction *support* the hypothesis, but they do not prove it — a later test could still find otherwise. 1 mark Accept “supported, not proved” in any wording.

Q15 (6 marks · VC2S10I02) (a) a push built into a method that shifts results one way, the same way every time. 1 mark Accept any response that captures a method built so results lean one way. (b) any two with explanations, for example: standing only at the front gate misses everyone who arrives another way (a side gate, the bus stop), so walkers and front-gate arrivals are over-counted; recording only the first 20 students seen is a convenience sample, not a fair one, so early arrivers are over-counted; one Monday morning is a single snapshot and may not be like other days. 2 marks (1 mark each: the feature and how it pushes the result must both be there). (c) any one, for example: sample students chosen at random from the roll; count at every entrance at once; repeat across several mornings; survey the whole school, not the first 20 at one gate. 1 mark (d) 2 marks — 1 mark for the judgement: the result cannot carry the whole-school claim as it stands; 1 mark using *valid* correctly, for example: a valid investigation tests what it set out to test, and this sample does not fairly represent every student, so it does not validly answer the whole-school question. Accept “not valid for the whole school because the sample is biased” with the idea of fair representation.

Q16 (4 marks · VC2S10I02) (a) any two hazard–control pairs, for example: steep, slippery bank — keep back from the edge and no climbing; uneven ground — strong shoes and walk, never run; water — no entering the creek, work from the bank; sun and UV — hat, sunscreen, water and shaded breaks; wildlife — keep distance, never touch. 2 marks (1 mark per pair; the hazard and a matching control must both be there). (b) the teacher; before fieldwork begins. 1 mark Both parts needed. Marker note: the student never works this check out alone — that is the subtopic's amber rule. (c) the Traditional Owners for that area, through the Registered Aboriginal Party for that area. 1 mark Accept “the Registered Aboriginal Party” or “the Traditional Owners”. Marker note: the framework behind (b) and (c) is the *Aboriginal Heritage Act 2006* (Vic) and the *Aboriginal Heritage Regulations 2018* (Vic); the instruments are named in Subtopic 3B and are not unpacked further here. No cultural knowledge is asked of any student.

Q17 (6 marks · VC2S10I03) (a) 46.4 g; it sits far above the other three readings, which agree to within 0.2 g. *1 mark* (b) mean of the three agreeing readings: $(45.2 + 45.1 + 45.3) \div 3 = 135.6 \div 3 = 45.2$ g. *1 mark* (c) the claim is weak: two readings are a small sample — with only two, there is no way to know which reading is wrong when they disagree, and the mean of two carries a large uncertainty; more trials let the mean settle and expose unusual readings. *1 mark* Accept any response that says more readings make the mean more trustworthy. (d) *2 marks* — 1 mark for the name: an offset (zero error) of +0.2 g; 1 mark for the action: zero or recalibrate the balance before weighing, or subtract 0.2 g from every reading, and record the correction in the method. (e) the range is small ($45.3 - 45.1 = 0.2$ g), so the repeat readings cluster tightly — the data are precise. *1 mark* The small range and the word *precise* (or its meaning) must both be there.

Test B — Section A — Multiple choice

Q	Answer	Code	Source
Q1	B	VC2S10H01	bank item VC2S10H01.E02 .S01.D1.V01
Q2	C	VC2S10H01	bank item VC2S10H01.E01 .S01.D1.V08
Q3	A	VC2S10H02	bank item VC2S10H02.E01 .S01.D1.V07
Q4	D	VC2S10H02	bank item VC2S10H02.E02 .S01.D2.V06
Q5	B	VC2S10I01	bank item VC2S10I01.E01 .S01.D2.V08
Q6	A	VC2S10I01	bank item VC2S10I01.E02 .S01.D1.V02
Q7	C	VC2S10I02	bank item VC2S10I02.E02 .S01.D1.V01
Q8	D	VC2S10I02	bank item VC2S10I02.E01 .S01.D2.V13
Q9	B	VC2S10I03	bank item VC2S10I03.E01 .S01.D2.V14
Q10	C	VC2S10I03	bank item VC2S10I03.E02 .S01.D1.V06

All ten items are drawn from the MCQ bank under the codes assessed on this paper; none is drawn from a neighbouring code. Items mix easy (D1) and medium (D2) bank difficulties. Items are printed with their stem and options word for word; only the positions of the options have changed for this paper, so that the correct answers are spread across A, B, C and D (the printed positions are $A \times 2$, $B \times 3$, $C \times 3$, $D \times 2$).

Why each answer is correct, in brief:

- **Q1 B** — consensus forms when independent lines of evidence and many reviewed studies point the same way; it is not a vote, does not need identical models, and never picks the most dramatic model.
- **Q2 C** — a result that does not support the prediction is still real evidence: record it, check the method, and share what is reliable. Changing or hiding data, or rejecting everything else, are the wrong moves.
- **Q3 A** — once the implant could deliver controlled pulses to the auditory nerve, it became an instrument for research: testing how different electrical patterns are heard as sound.
- **Q4 D** — faster computing lets the model evaluate more possible shapes in the same time and point to candidates for laboratory study; it does not fold real proteins, guarantee the model's rules, or remove the need for lab work.
- **Q5 B** — the question names the variable changed (distance from the lamp) and the result measured (bubbles per minute); the reversed and two-variable versions do not match the plan.
- **Q6 A** — the hypothesis names the variable changed (release height) and the predicted response (distance travelled); the carpet is kept the same, so roughness is not in this test.

- **Q7 C** — free, prior and informed consent means communities agree after full understanding, before the research starts; “no” is a complete answer.
- **Q8 D** — the planned welfare rule comes first: stop that animal’s trials and get advice; the data set never outranks the animal’s welfare.
- **Q9 B** — the balance is checked against a known mass before deciding; more decimal places alone do not make data better, and an unchecked instrument is not trusted.
- **Q10 C** — two sites is a small sample; before claiming for the whole area, more sites are sampled at random with the same method. Choosing sites beside the track would be a convenience sample.

Test B — Section B — Short/structured response

Q11 (4 marks · VC2S10H01) (a) any two of: the X shape showed the molecule is a spiral; the spacing of the dark bands fixed the rungs at 0.34 nm apart; the symmetry of the pattern fitted two strands wound around each other. 2 marks (1 mark each). (b) it puts the evidence any model had to fit on the public record beside the model itself, so every reader can check the model against the pattern — and the record fixes who supplied the evidence. 1 mark Accept any response that says the evidence sits beside the claim so both can be checked. (c) publication puts contributions on the record with names and dates — the record shows Franklin and Gosling supplied the X-ray evidence the model depended on, whatever happened later with prizes. 1 mark Accept any response that says the published record fixes who did what. Marker note: the case is taught in Subtopic 1A (elaboration VC2S10H01.E04); no student is asked for any knowledge beyond the record given.

Q12 (4 marks · VC2S10H02) (a) a central computer combines the signals from all 36 antennas. 1 mark (b) the array works as one much larger telescope and surveys the radio sky faster than any single dish could. 1 mark (c) 2 marks — 1 mark for the arrow: advances in technologies have enabled advances in science; 1 mark for the justification, for example: the linked array and its fast computing produce observations and survey data no single dish could, which is new evidence for radio astronomy. Accept the other arrow only with a matching justification (for example: radio-astronomy science drove engineers to build the array) — the case runs both ways, and the mark is for the pairing, not the choice. Marker note: this is elaboration VC2S10H02.E02; the case is taught in Subtopic 1B with the same source.

Q13 (4 marks · VC2S10I01) (a) independent variable: the intensity of exercise; dependent variable: heart rate. 2 marks (1 mark each; the name and the role must match). (b) any one workable set of levels, for example: rest, walk, jog, run; or walking speeds of 3, 5 and 7 km/h; or step rates on the spot. 1 mark The levels must be named, not just “different intensities”. (c) any one with its reason, for example: the same person (different bodies have different resting heart rates); the same time spent at each level (otherwise the heart has different time to respond); the same time of day, or the same recovery time before counting (so the counts are comparable). 1 mark The variable and a reason must both be there.

Q14 (2 marks · VC2S10I01) (a) the claim is wrong: a result that does not support the hypothesis is not a failure — it still tells you something real, and it is the only way a hypothesis or model gets better. 1 mark Accept any response that says unsupported results are valuable, not failures. (b) check the method first — was it a fair test? — and if the method was sound, revise the hypothesis (or the model behind it) and test again. 1 mark Accept “revise and try again” with the method check in any order.

Q15 (6 marks · VC2S10I02) (a) Group 1. 1 mark Justification: its three counts (86, 90, 88) agree closely, while Group 2’s (86, 42, 128) swing wildly — the same method run again gives agreeing results only for Group 1. 1 mark (b) reproducible means someone else can follow the written method and get results that agree. 1 mark (c) any two with explanations, for example: a fixed counting line (gate post to power pole) — every observer counts the same vehicles, because the line is the same each time; direction rule (only vehicles moving towards the school) — observers cannot disagree about which vehicles count; the bicycle/parked-vehicle rule — parked vehicles are never counted by mistake; one observer at a fixed position — two observers cannot double-count or miss the same vehicle differently. 2 marks (1 mark each: the rule and how it keeps the count the same must both be there). (d) any one, for example: those three mornings may not be like every morning — a wet day, an excursion day or an event changes traffic; mornings later in the week could differ. 1 mark Accept any response that says three mornings is a small sample of the week’s variation. Marker note: this is the traffic-survey skill of elaboration VC2S10I02.E06, taught in Subtopic 3A, in fresh numbers.

Q16 (4 marks · VC2S10I02) (a) any two of: flying limestone chips while hammering (eye or skin injury); splashes of dilute acid (to skin or eyes); dust from the crushed powder (inhaled); broken chips or spilled liquid on the bench. 2 marks (1 mark each; each hazard must say what could be harmed and how). (b) any one with its level, for example: crush the chips in a closed container or behind a screen — isolation (or engineering); use dilute acid only, in small volumes, in trays — substitution; a written rule that only one person hammers at a time, bags over the chips — administrative. 1 mark The control must sit above PPE in the hierarchy, and the named level must match. Do not accept goggles here — goggles are part (c). (c) goggles are PPE, the last line of the hierarchy: they only stand between the hazard and the eye — the chips still fly, and everything the goggles miss stays exposed. Controls higher in the hierarchy deal with the hazard itself. 1 mark Accept any response that says PPE wraps the person instead of fixing the

hazard. Marker note: the risk-assessment routine is taught in Subtopic 3B (elaboration VC2S10I02 . E07); this task applies it to a fresh context.

Q17 (6 marks · VC2S10I03) (a) ± 0.05 cm — half the smallest division. 1 mark (b) 5.65 cm. 1 mark Accept any honest estimate from 5.64 to 5.66 cm; the estimated digit beyond the millimetre marks must be there (5.6 cm alone loses the mark). (c) the readings are precise — repeat readings cluster tightly, with a range of only 0.03 cm. 1 mark Accept “close together” with the idea of precision. (d) 2 marks — 1 mark for the mean: $(5.65 + 5.66 + 5.64 + 5.65 + 5.67) \div 5 = 28.27 \div 5 = 5.654$ cm, which reports as 5.65 cm at the uncertainty’s decimal place; 1 mark for the uncertainty and form: range = $5.67 - 5.64 = 0.03$ cm, half-range = 0.015 cm, which rounds to 0.02 cm to one significant figure; result: **5.65 $\pm$ 0.02 cm**. (e) any two points always lie on some straight line, so two points can only suggest a relationship; several well-spaced points are needed to show whether the pattern really is a line. 1 mark

Tracing the items to the achievement standard

Every item traces to the achievement-standard extract of the subtopic that owns its code (§4.1 of the plan of record). The extracts are quoted as the authority file gives them. VC2S10I01 is split-delivered: 2A owns the questions clause and 2B the hypotheses-and-predictions clause; VC2S10I02 is split-delivered: 3A owns the valid, reproducible limb and 3B the safe, ethical limb. Each pair cites the same extract.

Code	Subtopic	Achievement-standard extract	Test A items	Test B items
VC2S10H01	1A	“students analyse the importance of different scientific methods, critique, replication, publication and peer review in the development of scientific knowledge.”	Q1, Q2, Q11	Q1, Q2, Q11
VC2S10H02	1B	“They examine the relationship between science, engineering and technologies.”	Q3, Q4, Q12	Q3, Q4, Q12
VC2S10I01	2A and 2B	“Students formulate and refine questions and hypotheses to make reasoned predictions, test relationships and develop explanatory models when investigating scientific questions, problems and claims.”	Q5, Q6, Q13, Q14	Q5, Q6, Q13, Q14
VC2S10I02	3A and 3B	“They plan a range of valid, reproducible and safe scientific investigations and explain how they have addressed any ethical and cultural considerations when	Q7, Q8, Q15, Q16	Q7, Q8, Q15, Q16

Code	Subtopic	Achievement- standard extract	Test A items	Test B items
VC2S10I03	3C	generating or using primary and secondary data.” “They select and use equipment to generate and record data, ensuring the use of suitable sample sizes and assessing the precision of multiple measurement readings.”	Q9, Q10, Q17	Q9, Q10, Q17

Inquiry weighting: 28 of the 40 marks on each paper sit on the inquiry codes VC2S10I01, VC2S10I02 and VC2S10I03 (Section A Q5–Q10 and Section B Q13–Q17) — well above the required minimum of 6 inquiry-strand marks per topic test (D7).

Mark record — Test A

Question	Code	Worth	Given
Q1 — the purpose of peer review	VC2S10H01	1	
Q2 — two labs, one result	VC2S10H01	1	
Q3 — radio-astronomy maths to wi-fi	VC2S10H02	1	
Q4 — sensors, computers, earthquakes	VC2S10H02	1	
Q5 — a question names its variables	VC2S10I01	1	
Q6 — choosing a testable hypothesis	VC2S10I01	1	
Q7 — consultation before fieldwork	VC2S10I02	1	
Q8 — replace before animals	VC2S10I02	1	
Q9 — checking a pH meter	VC2S10I03	1	
Q10 — sample size and fair placement	VC2S10I03	1	
Q11 — Marshall and Warren: rejected, then validated	VC2S10H01	4	
Q12 — crash-test dummies and bias	VC2S10H02	4	
Q13 — variables and a hypothesis	VC2S10I01	4	
Q14 — hypothesis, prediction, proof	VC2S10I01	2	
Q15 — bias in a survey plan	VC2S10I02	6	
Q16 — safe and ethical fieldwork	VC2S10I02	4	
Q17 — precision, calibration, sample size	VC2S10I03	6	
Total		40	

Mark record — Test B

Question	Code	Worth	Given
Q1 — consensus with different models	VC2S10H01	1	
Q2 — when results do not support	VC2S10H01	1	
Q3 — the implant feeds the science back	VC2S10H02	1	
Q4 — faster computing, more shapes	VC2S10H02	1	
Q5 — a question names its variables	VC2S10I01	1	
Q6 — choosing a testable hypothesis	VC2S10I01	1	
Q7 — free, prior and informed consent	VC2S10I02	1	
Q8 — the welfare rule comes first	VC2S10I02	1	
Q9 — which balance, and checked how	VC2S10I03	1	
Q10 — a small sample, fairly taken	VC2S10I03	1	
Q11 — Photo 51: publication and credit	VC2S10H01	4	
Q12 — ASKAP: one telescope from 36	VC2S10H02	4	
Q13 — exercise and heart rate	VC2S10I01	4	
Q14 — failure claims and revised hypotheses	VC2S10I01	2	
Q15 — reproducible traffic counts	VC2S10I02	6	
Q16 — hazards and the hierarchy of controls	VC2S10I02	4	
Q17 — uncertainty, precision, sample size	VC2S10I03	6	
Total		40	