

Marking Rubric – 20 Points, Band by Band

Written so two markers reading the same work land within one mark of each other. A well-measured failure outscores an unmeasured success, and an effect claimed without a baseline is returned unmarked.

The three criteria, band by band

Twenty marks. Read the bands before you read the work, and mark against the description you can see evidence for.

The artifact 6 MARKS

Is the one change traceable to the costed problem, and was the same quantity measured the same way before and after, so a stranger could repeat both halves?

MARK	BAND	WHAT THE WORK LOOKS LIKE
0-1	Not yet	A claim of improvement stands where a measurement should be. There is no costed problem statement, or one carrying no number, no unit and no period. The change is described but the quantity behind it was never defined, so no before-figure and no after-figure exist to compare.
2-3	Emerging	Real measurement exists and the artifact opens. The problem carries a number but no cost per week or per year. Several things moved at once, a swap and a new shop and a new routine, so the after-figure cannot be attributed to any one of them, and the learner does not say so.
4-5	Secure	One named change, traceable to the largest line in the costed statement. Before and after share the same instrument, the same divisor, the same sources and the same times of day, all stated. Every figure carries a unit and a date. A stranger could repeat both halves from the artifact alone.
6	Exceptional	Secure, and the artifact carries a check the ten steps never asked for: an unchanged comparison household, a second isolation run, a repeat by another person. Every after-week is charted separately, including the one that went backwards, and the week the method slipped is marked on the chart itself.

What you look at

- The costed problem statement and the chosen change read side by side, checking that the change attacks the largest line on that page
- The baseline method line and the after method line set against each other word for word: instrument, divisor, sources, times of day, wording
- The chart, checking that every after-week appears separately rather than as a single mean
- Two figures picked at random from the artifact, traced back to a dated reading and a named source

Marking error to avoid

Assessors reward a big saving. A hand-drawn table showing a fall of four litres, measured the same way twice, outscores a designed poster claiming forty litres from a before-figure nobody can date. Mark the comparison, not the size of the number.

The process and its failures **6 MARKS**

Does the record show a change that was actually run, slipped somewhere, and was measured honestly anyway, rather than a plan written up after the result was known?

MARK	BAND	WHAT THE WORK LOOKS LIKE
0-1	Not yet	No dated log, or a log in which nothing fails. A submission recording no failures scores zero here, whatever the after-figure shows, because three weeks of measuring always slips somewhere. Entries carry one date, no reading is marked missed, and every discarded attempt has been tidied out.
2-3	Emerging	Failures appear but stay cosmetic: a phone died, a week was busy. Dates are spread and roughly in order. Missed readings are noted but never counted, and nothing states what each slip did to the comparison. A well-measured failure outscores an unmeasured success, and this record measures neither.
4-5	Secure	Two or more failures carry a date, what broke and what it cost the comparison: the night the change was forgotten, the reading taken an hour late, the guest who doubled the pot. Each is followed by a stated correction, and the spoiled rows stay in the log rather than being deleted.
6	Exceptional	Secure, and the change did not work. All three after-weeks were measured anyway, and the flat or worsened figure is reported with the care a win would get. The learner names the part of their own method they now distrust, and states the single thing they would test next and how.

What you look at

- The failure log read against the evidence log, checking that every missed or late reading appears in both
- The dates on the protocol and the prediction, checked against the date of the first after-reading
- The spoiled rows, discarded tally cards and abandoned isolation tests kept in the submission
- What the learner wrote on the week that went the wrong way, and whether that week reaches the chart

Marking error to avoid

Assessors mark the size of the effect and forget where the marks are. A learner whose swap changed nothing, who logged every one of the twenty-one days and names what they would test next, takes six here. A clean win with three unexplained gaps in the log takes two.

The defence 8 MARKS

Asked whether this would have happened anyway without the change, does the learner reason from their own before-and-after data, or repeat the result more slowly?

MARK	BAND	WHAT THE WORK LOOKS LIKE
0-2	Not yet	Answers repeat the submitted result close to word for word. Asked whether it would have happened anyway, the learner says the change caused it and offers nothing behind that. A figure from their own before-period cannot be found on request. I do not know arrives as a full stop.
3-4	Emerging	The learner explains their own arithmetic accurately and finds any figure in it. Asked the standing question, they answer from belief rather than data, or name one confounder and drop it unexamined. A fair challenge is agreed with and left there, and the result is restated unchanged.
5-6	Secure	The standing question is answered from their own record: the before-period and the after-weeks are compared on the things that moved, and the learner says which of those they held fixed and which they only wrote down. They separate what was measured from what was inferred, and concede a limit without dropping the finding.
7-8	Exceptional	Secure, and the learner names a confounder the assessor had not raised, says which way it would push the number and by roughly how much. They propose the measurement that would separate it from their change, state what three weeks cannot settle, and correct the assessor where the assessor is wrong.

What you look at

- The standing question asked in the learner's own terms: would this have happened anyway without your change, and how do you know
- One figure from the before-period asked for out of order, and how long it takes to reach it
- A confounder the learner did not list, put to them cold, and what they do with it
- What the learner does after saying they do not know: a named next measurement, or a stop

Marking error to avoid

Confidence reads as evidence. A quiet learner who says the load-shedding hours fell in the same fortnight, so part of this saving is not mine outscores a fluent one who insists the change worked because the household feels it did. Mark the reasoning against the data, not the delivery.

Two gates. Neither is scored, and neither can be waived.

The ten relationships

This is a gate, not a scored criterion. Ten logged relationships are required per project: 1 expert, 3 people the problem affects, 3 peers who reviewed the work before it was finished, 3 people who saw the finished work and replied. Until all ten are logged with evidence, the work is not marked at all and no score out of 20 is issued. The assessor returns it unmarked with the missing rows named.

WHO	WHAT COUNTS AS PROOF	HOW TO CATCH IT
1 expert in the field	A dated record of a two-way exchange: a reply email or message thread, a call with the date and length noted, or signed notes naming the expert, their role and where they work. The learner records one thing the expert said that changed the method or the number.	The learner cites a video, a podcast or a chatbot as the expert. Catch it by asking what the expert said that the learner did not expect. A one-way source cannot answer back, so nothing in the log will be specific to this measurement.
3 people the problem affects	Three separate dated records, each naming the person or a consistent initial, their relation to the costed problem, and at least one direct quote in their own words. One of the three must be whoever agreed to the change before it was made. On the No-Tech route, handwritten notes with a signature or thumbprint count.	Three relatives asked once, at the same table, giving the same answer. Catch it by comparing the three records for shared phrasing and identical dates. Ask which of the three objected to the change, and what they said about living with it.
3 peers who reviewed it before it was finished	Three dated comments made before the after-period closed, each naming something to change, plus the learner's note of what they did with each comment. At least one comment must be on the method or the arithmetic, and the version reviewed must predate the submitted artifact.	Reviews collected after the chart was drawn, so every comment is praise. Catch it by checking review dates against the protocol date and the artifact date, and by looking for one comment that led to a visible change. Praise with no requested change is not a review.
3 people who saw the finished work and replied	Three dated replies to the finished result, from named people, each responding to something in the work rather than to the learner. A screenshot, a written reply, or signed notes of what was said and when.	Shares counted without replies, or three replies of the same two words. Catch it by requiring that each reply names a number, a method or a limit from the work. Ask the learner what the three replies disagreed about.

The baseline

This is a gate, not a scored criterion. An effect claimed without the same quantity measured before the change is returned unmarked, because there is nothing to measure against. A baseline counts only when it carries a number with its unit, the dates it was taken on, the instrument or source behind it, and a written method that the after-measurement then repeats. Where any row below is true, the assessor issues no score out of 20, names the row, and sets the baseline again with the after-period rerun from it.

WHO	WHAT COUNTS AS PROOF	HOW TO CATCH IT
No before-figure at all	A dated before-number in the same unit as the after-number, taken across a stated period, sitting in the evidence log beside the readings it came from. Nothing else counts as a baseline, however good the after-period is.	A remembered figure, last year's bill, a national average or the shape of the problem described in words stands in for the before. Catch it by asking for the date and the instrument of the before-reading. Neither exists, so the answer is a paragraph rather than a number.
A before-figure taken after the change began	Every dated baseline reading falls before the dated line where the change starts, and the prediction was written and witnessed before the first after-reading. The protocol names the day the change went in.	The swap was already in the house for a fortnight, so the baseline measures the fix. Catch it by setting the earliest baseline date against the date the change was agreed with the household. Ask who was already doing the new thing before the first reading was taken.
A before and after measured by different methods	The two method lines match word for word on instrument, divisor, source, question wording, times of day and the length of the period. Where one thing had to move, the learner names it, dates it and states what it does to the comparison.	Weighed before and estimated by cupfuls after; one database for the baseline and a second for the after; a question reworded between the two rounds. Catch it by reading the two method lines side by side and marking every word that differs.

WHO	WHAT COUNTS AS PROOF	HOW TO CATCH IT
A before-figure with no stated method	A written method a second person could follow to reproduce the before-number: what was counted, with what, by whom, at what times, and how many days it ran. The number alone, however precise, is not a baseline.	A clean figure appears with no working behind it and no note of how it was taken. Catch it by asking the learner to reproduce their own baseline in front of you from their written method. Where the method exists only in their head, the after-period cannot match it.

Reading the bands against a declared level

Bands are read against the capability level the learner declared before starting, Starter, Practitioner, Researcher or Contributor, and not against one absolute standard for everybody. A Contributor who submits Starter-grade evidence scores badly, because the mark is set against the level they declared and not against the room. A Starter is never marked down for lacking what their level does not ask for, so an unjustified sample size costs a Researcher marks and costs a Starter nothing.

DECLARED LEVEL	WHAT "SECURE" EVIDENCE MEANS HERE
Starter	Secure means the written method was followed with no gap: same instrument, same times, every zero entered rather than left blank, and one stranger asked in person. Sample size and published comparison are not asked for here.
Practitioner	Secure means the learner chose what to measure and defends that choice, every borrowed figure is cited to its page or table, and one chart is read correctly by a stranger with nobody standing there explaining the axes.
Researcher	Secure means the protocol was written and dated before the first reading, three after-weeks are justified rather than assumed, and the error from the instrument, the divisor and any borrowed coefficient is quantified rather than mentioned.
Contributor	Secure means the figure is set against published work by name, the method survives a specialist who disagrees with it, and every raw reading and source is published openly so a stranger can recompute the result.

Which deliverable feeds which mark

#	DELIVERABLE	FEEDS	WHAT YOU CHECK IN IT
1	Question card	The artifact - whether the change is traceable to a costed problem	One named quantity with its unit and the household or place it belongs to, written down before any measuring starts.
2	Expert interview pack	Gate one - the expert row	Five questions sent before the reply arrived, and one recorded sentence from the expert that changed the method or the number.
3	Evidence log	Gate two - the baseline; and the artifact	Every before-figure with its date, instrument and source, taken before the change, and the identical columns filled in again afterwards.
4	Failure log	The process and its failures	Dated entries naming what slipped and what each slip did to the comparison, including every reading that was missed or taken late.
5	The build	The artifact	The measuring set-up and the before-and-after chart on one axis, with each after-week shown separately rather than averaged away.
6	One-page explainer	The artifact	The before number, the after number, the one change and one stated limit, on a page a neighbour reads in ninety seconds.
7	Identity card	The artifact	A name for the measurement and the one sentence meant to be repeated, carrying the number itself rather than an adjective about it.
8	Relations ledger	Gate one - all ten relationships	Ten names with roles, dates and what each person replied, with the three peer comments dated before the finished artifact.
9	Defence video	The defence	Under three minutes covering the baseline method, the one change, the result, and the answer to whether it would have happened anyway.

What the total means

TOTAL	VERDICT	WHAT HAPPENS NEXT
14-20	Pass - Foundation complete	Both gates are already cleared, so AI marks and a human certifies. Written feedback returns on every section within two weeks. The project enters the learner's portfolio with its before-figure, its after-figure and its confounder list attached, and the unverified label is removed. The learner starts their next project in a different lab.
10-13	Resubmit with specific notes	The assessor names the one criterion holding the mark down and the two changes that would lift it, with a date to resubmit by. The baseline and the after-readings stay as they are and are never recollected for a resubmission; only the named section is reworked. A resubmission is marked against the same bands by a different marker.
0-9	Not yet - rework one section	The assessor picks the single weakest section and sets it again, with an example of work at the Secure band attached at the learner's declared level. The learner keeps their relationships, their baseline and their log. Nothing is scored a second time until that one section is redone, so nobody restarts the project.

The gate is the assessment, never the fee.

No submission is refused for money. If a learner passes and cannot pay for certification, a school, a sponsor or the community fund covers it.

Calibration – six marks explained

Read these before marking a batch. If your instinct disagrees with one, that is the conversation to have with your second marker.

5/6 on the artifact

An Agriculture & Climate submission of The Sixty-Litre Dinner. One dish, one swapped ingredient, cooked on three dated evenings. The same scale, the same divisor by five and the same two published sources are used before and after, each named to its table. Four bars on one axis show the baseline and all three weeks. The sixth mark is missing because no second household cooked the dish as a comparison.

2/6 on the artifact

An Energy & Smart Systems submission of The Four-Hundred-Rupee Leak. The baseline week is complete and the readings are photographed. Then the learner moved the wash to nights, changed a fan setting and replaced a bulb in the same fortnight, so the fall of ninety units belongs to no single change. The write-up credits the fan setting, and the artifact carries nothing that could separate the three.

6/6 on the process

A Psychology & Behaviour submission of The Trigger. The learner removed the cue and logged fourteen more days: the count went from nine a day to nine a day. Every day is entered, including the three where entries were late, and the log says so. They name the cue definition as the part they now distrust, and state the two-cue test they would run next.

3/6 on the process

A Health & Preventive Medicine submission of The Sugar You Did Not Order. Two failures appear across the fortnight: one person away for two days, and a label figure copied from a different pack size. Both are dated and both are corrected. Neither says what it did to the household teaspoon total, and the missing two days are averaged into the after-week without a note anywhere.

7/8 on the defence

A Media & Deepfake submission of The Lie This Street Believes, where belief fell only from fourteen to thirteen of twenty. Asked whether that would have happened anyway, the learner names something the assessor had not raised: four of the twenty had already heard a radio correction that week, and says which direction that pushes the figure. They propose asking the twenty where they first heard it.

3/8 on the defence

A Sociology & Future Society submission of The Question Nobody Asked. The learner finds every figure in their own table and explains the timing arithmetic correctly. Asked whether the minutes would have dropped anyway, they say the neighbours worked together well, which nothing in the data shows. The assessor names the school holidays falling in the after-fortnight; the learner agrees and moves on.

Moderation – two markers, one mark

AI marks and humans certify, so a mark leaves the system only after two people have agreed on it. One marker alone drifts towards the learners they know, the language they read most easily, and the submission with the biggest reported saving.

1. 1. Gates first. Before any scoring, one marker checks both gates: ten logged relationships, and a baseline with a date, a unit, an instrument and a written method that the after-period repeats. A failed gate returns the work unmarked with the row named, and no score is recorded.
2. 2. First mark. The first marker scores all three criteria and writes one line of evidence per criterion naming what they looked at, including the two method lines they compared. Marks go in a sealed field, not in the shared sheet.
3. 3. Blind second mark. A second marker, who has not seen the first marks or the learner's name, school or city, scores the same work from the same bands at the same declared level. Anonymise the file before it is passed on.
4. 4. Compare and resolve. Open both sets together. Within one mark overall and two on any criterion, take the higher total. Beyond that, each marker names the observable feature behind their score; if they still differ, a third marker scores blind and the middle mark stands. Never average silently.
5. 5. Sample and record. Moderate every submission at 14 or above, every submission below 10, every submission reporting no effect, and one in five of the rest. Record both marks, the final mark, the criterion that split them and the descriptor line that settled it, and rewrite any descriptor that splits markers twice.

Tolerance

Two markers may differ by up to one mark out of 20, and by no more than two marks on any single criterion. Beyond either limit the mark is not issued: the pair reconcile against the descriptors and against the learner's declared level, and an unresolved split goes to a blind third marker whose score stands as the final mark.

EAD Labs - Early Age Development by technogroup.ai - free to copy and use in any school - earlyagedevelopment.com/EAD_Labs.html