

HUMANS ARE NOT METRICS.



THE MISSING COLUMN

A field guide to useful numbers, blind spots, incentives and the people who live beyond a dashboard.

16

METRICS

04

CASES

02

WORKSHEETS

01 / ORIENTATION

Measure, but do not reduce.

Statistics illuminate part of a system. They do not contain the whole system, and they never define the worth of a person.

A well-chosen metric can reveal an overlooked problem, improve public accountability, detect exclusion, and help people coordinate. The failure is not counting. The failure is treating a narrow observation as a comprehensive judgment.

THREE DISTINCTIONS TO CARRY

SIGNAL IS NOT GOAL

Attendance can tell you who came. It cannot tell you who could have come, how welcome they felt, or what changed for them.

ACCURACY IS NOT COMPLETENESS

A pass rate can be calculated perfectly and still omit transferable learning, peer support or students who left before testing.

IMPROVEMENT IS NOT CAUSATION

A number may rise after a policy change without proving the policy caused the change. Other conditions may have shifted too.

THE OPERATING PRINCIPLE

Never ask only: "Did the metric go up?" Also ask: "What did it measure? For whom? At what cost? What would contradict our interpretation?"

All numerical illustrations and organisational stories in this publication are intentionally fictional teaching examples. They are not estimates, benchmarks or predictions. Methods referenced at the end support the general reasoning, not the invented figures.

02 / ANATOMY

The six parts of a metric

Before discussing the value of an indicator, reconstruct what it actually represents.

01**CONSTRUCT**

What underlying thing do we wish to understand? “Learning” and “exam passing” are not identical constructs.

02**PROXY**

What observable variable stands in for it? Why should that variable track the desired outcome?

03**POPULATION**

Who is eligible, sampled, excluded, absent or unable to respond? A rate without a population is a riddle.

04**TIME WINDOW**

A day, quarter or cohort? Did the definition, window or collection method change?

05**INCENTIVES**

Who can influence this measure, and how does being judged by it change their behaviour?

06**DECISION RULE**

What action will the number trigger, and what independent evidence can pause or reverse that action?

TEST THE FRAME

One missing element can change the meaning of the same displayed number. Keep the definition, denominator, collection limits and decision rule close to the chart.

03 / THE PORTABLE AUDIT

Twelve questions before a verdict

Use the first six for a quick review. Use all twelve when a measure shapes pay, access, education, services or reputation.

- ☐ 01 Exactly what event or outcome does this number record?
- ☐ 02 Who is in the numerator and the denominator?
- ☐ 03 Who is absent, unreachable or excluded?
- ☐ 04 Is the definition stable across time, teams and groups?
- ☐ 05 What behaviour could the target reward accidentally?
- ☐ 06 What relevant outcome can improve while this number falls?
- ☐ 07 What could worsen while the reported number improves?
- ☐ 08 How uncertain is the estimate or comparison?
- ☐ 09 Does an average hide different experiences or harms?
- ☐ 10 Could a changing mix of people explain the trend?
- ☐ 11 What other method would challenge or corroborate this story?
- ☐ 12 Can affected people contest the interpretation without penalty?

Use this as a discussion guide, not a points system or moral test. More boxes checked is not proof of a better institution. A single material blind spot may deserve more attention than the rest.

04 / INSTRUMENT LIBRARY / 01 OF 08

01-02 | Signal and shadow

Each indicator is useful inside a definition. Pair it with evidence that covers a different part of the goal.

01 Grades

EDU / 01

WHAT IT CAN TELL YOU

A grade can summarise performance on a particular set of assignments against a rubric. It can direct useful feedback when criteria are clear.

WHAT IT OMITS

It omits the quality of help a student received, prior opportunity, curiosity beyond the assigned task, and things the rubric was not designed to observe.

WHAT TO PAIR WITH IT

Show the task and rubric, assess learning over time, and let students demonstrate understanding in more than one way. Do not translate a grade into a statement about human worth.

ASK IN THE ROOM

Did the assessment measure the learning goal?

02 Productivity

WORK / 02

WHAT IT CAN TELL YOU

A count of completed tasks helps a team plan capacity or spot a workflow bottleneck—if tasks are comparable and counted consistently.

WHAT IT OMITS

It may miss invisible coordination, mentoring, preparation, prevention of errors, task complexity and the costs of hurried rework.

WHAT TO PAIR WITH IT

Combine throughput with quality, rework, complexity, team learning and the experience of people doing the work. Interpret slow periods before judging them.

ASK IN THE ROOM

Are we counting outputs or useful outcomes?

04 / INSTRUMENT LIBRARY / 02 OF 08

03-04 | Signal and shadow

Each indicator is useful inside a definition. Pair it with evidence that covers a different part of the goal.

03 Followers

MEDIA / 03

WHAT IT CAN TELL YOU

Follower counts reveal the rough scale of a platform audience. A rising count may matter for distribution and commercial opportunities.

WHAT IT OMITS

It cannot reveal depth of attention, agreement, trusted connection, audience composition or whether the work is useful offline.

WHAT TO PAIR WITH IT

Read substantive responses, repeat engagement and direct feedback. Some important work may remain valuable with an unusually small audience.

ASK IN THE ROOM

Is reach becoming relationship?

04 Test scores

EDU / 04

WHAT IT CAN TELL YOU

A standardised test can compare performance on specified items at a given time. With thoughtful design, it can expose meaningful gaps.

WHAT IT OMITS

A score does not contain all of learning, creativity, resources, language context or untested skills. Group comparisons also depend on sampling and comparability.

WHAT TO PAIR WITH IT

Check the construct, scoring scale, uncertainty and selection of test-takers; pair with other learning evidence. See the OECD PISA readers' guide for scale interpretation.

ASK IN THE ROOM

What does the scale actually represent?

04 / INSTRUMENT LIBRARY / 03 OF 08

05–06 | Signal and shadow

Each indicator is useful inside a definition. Pair it with evidence that covers a different part of the goal.

05 Revenue per worker

ECON / 05

WHAT IT CAN TELL YOU

This ratio can help describe output per employee under a given accounting definition and support some business comparisons.

WHAT IT OMITS

It does not isolate an individual's effort. Prices, capital, outsourcing, business model, hours worked and who is counted as a worker all change the ratio.

WHAT TO PAIR WITH IT

Define revenue, time period and labour denominator. Contextualise with costs, equipment, job quality and distribution of gains—not just the headline ratio.

ASK IN THE ROOM

Whose contributions disappear into the denominator?

06 Response time

SERVICE / 06

WHAT IT CAN TELL YOU

Time to first response can reveal neglected requests and help teams honour reasonable service commitments.

WHAT IT OMITS

A quick acknowledgement is not a solution. The clock often overlooks complexity, repeated hand-offs, emotional labour and reopened tickets.

WHAT TO PAIR WITH IT

Track resolution, reopens, case severity and satisfaction in appropriate contexts. Avoid penalising an agent for spending enough time with a difficult case.

ASK IN THE ROOM

Did faster mean better resolved?

04 / INSTRUMENT LIBRARY / 04 OF 08

07-08 | Signal and shadow

Each indicator is useful inside a definition. Pair it with evidence that covers a different part of the goal.

07 Retention

PEOPLE / 07

WHAT IT CAN TELL YOU

Staying rates can indicate continuity and reveal an abrupt organisational change when definitions remain stable.

WHAT IT OMITS

Retention alone cannot distinguish trust and satisfaction from constrained choices, scarce alternatives or fear of change.

WHAT TO PAIR WITH IT

Compare voluntary and involuntary exits, mobility opportunities, job conditions and confidential employee perspectives. Respect privacy and avoid inferring motives from a rate.

ASK IN THE ROOM

Do people stay because they want to?

08 Attendance

CIVIC / 08

WHAT IT CAN TELL YOU

Attendance helps organisers plan capacity and spot when people cannot reliably access a program.

WHAT IT OMITS

It omits travel costs, caregiving, accessibility, competing commitments, participation quality and people who could never register.

WHAT TO PAIR WITH IT

Pair sign-ins with access barriers, qualitative participation, learning or benefit, and the voices of people missing from the room.

ASK IN THE ROOM

Who had the option to be present?

04 / INSTRUMENT LIBRARY / 05 OF 08

09–10 | Signal and shadow

Each indicator is useful inside a definition. Pair it with evidence that covers a different part of the goal.

09 Graduation rates

EDU / 09

WHAT IT CAN TELL YOU

Completion rates help institutions notice lost opportunities and track whether participants finish a defined programme within a stated period.

WHAT IT OMITS

They may hide changed admissions standards, transfers, partial progress, meaningful learning and reasons people could not continue.

WHAT TO PAIR WITH IT

Publish the cohort definition and follow-up window; look at access, transfer students, completion quality and respectful accounts of barriers.

ASK IN THE ROOM

Who started, who was excluded, and what did completing mean?

10 Average salary

PAY / 10

WHAT IT CAN TELL YOU

A mean salary is one compact way to describe the labour costs or earnings of a defined group.

WHAT IT OMITS

A few high earners can pull the mean upward. It does not reveal the median, pay distribution, hours, job security or cost of living.

WHAT TO PAIR WITH IT

Report median and ranges, relevant roles, hours and contract types; examine gaps with proper safeguards for privacy and small groups.

ASK IN THE ROOM

Does the average represent the typical person?

04 / INSTRUMENT LIBRARY / 06 OF 08

11–12 | Signal and shadow

Each indicator is useful inside a definition. Pair it with evidence that covers a different part of the goal.

11 Satisfaction scores

EXPERIENCE / 11

WHAT IT CAN TELL YOU

A clearly asked satisfaction question can help identify service failures and monitor a specific experience over time.

WHAT IT OMITS

Scores may miss nonrespondents, fear of consequences, changing expectations, question wording and the reasons behind a rating.

WHAT TO PAIR WITH IT

Record response rates, question wording and uncertainty; invite optional concrete examples and provide an independent route for critical feedback.

ASK IN THE ROOM

Who answered, and did they feel safe answering?

12 Engagement rate

MEDIA / 12

WHAT IT CAN TELL YOU

Defined engagement events per impression can help compare how published work attracts actions on a particular platform.

WHAT IT OMITS

It may equate a tap with understanding, reward outrage and miss private conversations, thoughtful reading or effects outside the platform.

WHAT TO PAIR WITH IT

Define numerator and denominator, segment by format and follow substantive comments, audience needs and actual outcomes where observable.

ASK IN THE ROOM

What counts as an interaction and who saw the work?

04 / INSTRUMENT LIBRARY / 07 OF 08

13–14 | Signal and shadow

Each indicator is useful inside a definition. Pair it with evidence that covers a different part of the goal.

13 Research citations

RESEARCH / 13

WHAT IT CAN TELL YOU

Citation counts can indicate that work has entered scholarly conversations or is being used by others.

WHAT IT OMITS

A citation may affirm, challenge or merely mention a paper. Fields differ greatly in publication and citation habits; new work needs time.

WHAT TO PAIR WITH IT

Read the citing context, evaluate methods and contributions, and consider field-normalised patterns without making a count a verdict on a researcher.

ASK IN THE ROOM

What kind of attention is a citation evidence of?

14 Defect counts

QUALITY / 14

WHAT IT CAN TELL YOU

Recorded defects help product teams see failure patterns, prioritise repairs and reduce harm when severity and detection are consistent.

WHAT IT OMITS

Reporting culture, test coverage, user access, issue severity and hidden failures can change the count independently of underlying quality.

WHAT TO PAIR WITH IT

Combine counts with severity, independent testing, reporting ease, repair time and patterns among people most affected by failures.

ASK IN THE ROOM

Does fewer reports mean fewer actual problems?

04 / INSTRUMENT LIBRARY / 08 OF 08

15–16 | Signal and shadow

Each indicator is useful inside a definition. Pair it with evidence that covers a different part of the goal.

15 Volunteer hours

CIVIC / 15

WHAT IT CAN TELL YOU

Hours contributed can support resource planning and demonstrate the scale of donated time in a community project.

WHAT IT OMITS

Hours alone miss skills, outcomes, unpaid coordination, emotional strain and whether volunteers had meaningful choice.

WHAT TO PAIR WITH IT

Pair hours with beneficiary-defined outcomes, volunteer well-being, quality and fairness of workload; do not romanticise exhaustion.

ASK IN THE ROOM

Did time spent translate to valued help?

16 Conversion rate

BUSINESS / 16

WHAT IT CAN TELL YOU

The share of visitors taking a defined action can help test the clarity and usability of an experience.

WHAT IT OMITS

Conversion does not establish long-term benefit, informed consent, customer fit, returns or hidden friction introduced after the click.

WHAT TO PAIR WITH IT

Track cancellations, complaints, returns, accessibility, informed choice and longer-term satisfaction; reject manipulative design that improves only the click.

ASK IN THE ROOM

Was the action worthwhile and freely chosen?

05 / FICTIONAL CASE FILE

01 | A school mastered the test

An imagined scenario, not an account of an actual institution or a causal study.

THE REAL GOAL

Help students learn knowledge they can use beyond a familiar exam.

WHAT THE DASHBOARD SHOWS

The share of students passing a standardised final test rises after intensive targeted practice.

WHY THE METRIC COULD HELP

Focused practice may clarify expectations, support struggling learners and reveal where teaching needs improvement. A rising pass rate could represent meaningful progress.

HOW INCENTIVES COULD DISTORT IT

If the passing threshold becomes the dominant reward, teaching may narrow to repeated item formats. Open-ended projects and individual feedback may be squeezed, even if the measured pass rate is accurate.

WHAT THE NUMBER LEAVES OUT

Transfer to novel problems; curriculum breadth; which learners received support; teacher workload; test participation and exclusion.

A BETTER EVIDENCE DESIGN

Show pass rates by appropriately sized groups; add a small sample of unfamiliar problem-solving tasks; review work with teacher and student input; monitor who was not tested.

ASK THIS BEFORE DECIDING

Can we demonstrate learning on tasks that were not rehearsed, including among students who initially struggled?

WHAT MIGHT CHANGE OUR MIND: If novel-task performance rose with the pass rate and students reported broader opportunity, the narrow-curriculum hypothesis would be weaker.

05 / FICTIONAL CASE FILE

02 | The fastest unresolved request

An imagined scenario, not an account of an actual institution or a causal study.

THE REAL GOAL

Help service users resolve problems promptly and respectfully.

WHAT THE DASHBOARD SHOWS

A help desk cuts its median first-response time from an illustrative 18 minutes to 4 minutes.

WHY THE METRIC COULD HELP

Faster acknowledgement can prevent uncertainty, signal that urgent requests were seen and reduce abandoned enquiries.

HOW INCENTIVES COULD DISTORT IT

Agents may send acknowledgements immediately to satisfy the clock while postponing difficult resolutions. Transfer between queues can distribute work without fixing the underlying need.

WHAT THE NUMBER LEAVES OUT

Reopen rates; actual time to resolution; case complexity; hand-offs; customer effort; whether staff have sufficient focus.

A BETTER EVIDENCE DESIGN

Keep response time, but add severity-aware resolution measures, reopens, an escalation review and brief voluntary feedback from people with complex cases.

ASK THIS BEFORE DECIDING

For people with the hardest cases, did the total burden shrink or merely move to a different queue?

WHAT MIGHT CHANGE OUR MIND: If independent case reviews show fewer reopens and better outcomes for complex requests, faster response likely accompanied genuine improvement.

05 / FICTIONAL CASE FILE

03 | An audience, not a community

An imagined scenario, not an account of an actual institution or a causal study.

THE REAL GOAL

Help original work find interested people without making exploration unsafe.

WHAT THE DASHBOARD SHOWS

A studio displays rapid follower growth and increasing platform engagement.

WHY THE METRIC COULD HELP

More reach can open revenue, collaboration, access and distribution opportunities. Audience size is not meaningless.

HOW INCENTIVES COULD DISTORT IT

Content choices may skew toward familiar, reliably clickable formats, reducing room for unusual experiments or dissent inside the collective.

WHAT THE NUMBER LEAVES OUT

Depth of reading or listening; substantive exchanges; creative risks attempted; durable relationships; participation of quieter members.

A BETTER EVIDENCE DESIGN

Report reach with thoughtful repeat feedback, creative experiments that were pursued, opportunities for critique and member reflections on decision freedom.

ASK THIS BEFORE DECIDING

Are we reaching more people while preserving the ability to make difficult or less immediately popular work?

WHAT MIGHT CHANGE OUR MIND: If the collective continues to experiment and participants report meaningful dialogue, high reach might be enabling rather than crowding out depth.

05 / FICTIONAL CASE FILE

04 | Retention tells two stories

An imagined scenario, not an account of an actual institution or a causal study.

THE REAL GOAL

Create a workplace where people can do sustainable, worthwhile work and choose to stay.

WHAT THE DASHBOARD SHOWS

A workplace celebrates 94% retention in a fictional yearly snapshot.

WHY THE METRIC COULD HELP

Longer tenure can reflect trust, experience, good working conditions and continuity of knowledge.

HOW INCENTIVES COULD DISTORT IT

Some may stay because outside options are scarce, personal constraints make transitions hard, or morale problems are not safe to report.

WHAT THE NUMBER LEAVES OUT

Job quality; role mobility; workload; career prospects; choice; voluntary or involuntary exits; fear of reprisals.

A BETTER EVIDENCE DESIGN

Pair the retention rate with confidential qualitative feedback, opportunities to move roles, exits by relevant categories and safeguards against retaliation.

ASK THIS BEFORE DECIDING

Can people describe credible reasons they want to stay, and can they leave or speak up without punishment?

WHAT MIGHT CHANGE OUR MIND: If people freely report high agency, fair workloads and opportunities, retention may be a genuinely positive signal rather than a trap.

06 / WORKED EXERCISE

The denominator changes the story

Two perfectly accurate percentages can invite the wrong comparison if you skip who was eligible or what success meant.

PROGRAM A	90%	9 of 10
PROGRAM B	75%	75 of 100

READ THE FIGURES IN TWO DIRECTIONS

- Rate question: What fraction of participants met the stated criterion? A leads: 90% versus 75%.
- Count question: How many people met it? B leads: 75 versus 9.
- Policy question: Who was eligible, selected and supported, and what outcomes mattered? Neither figure answers this.

TRY THE ARITHMETIC

Programme A: $9 \div 10 = 0.90 = 90\%$. Programme B: $75 \div 100 = 0.75 = 75\%$. Changing the denominator changes the question, not necessarily the underlying truth. Note that these cohorts are different sizes; a causal comparison would require much more information.

WHAT TO TAKE BACK

Ask for the count next to every rate and a clear definition of both numerator and denominator. If you cannot reconstruct the fraction, you cannot interpret the headline confidently.

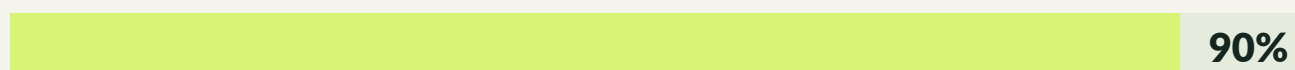
DISCUSSION PROMPT: Imagine Programme A serves only a small highly selected group while B accepts nearly anyone. Now reverse that assumption. How does your evaluation change? The counts alone cannot decide between these stories.

07 / WORKED EXERCISE

A survey with a bright green score

Large percentages can accurately describe respondents while poorly describing everyone the decision affects.

Imagine a fictional group of 100 members. Exactly 50 liked an event and 50 did not. A survey reaches 50 respondents: 45 people who liked it, and 5 who did not.

LIKED IT / ALL 100 MEMBERS**LIKED IT / 50 RESPONDENTS****WHAT HAPPENED?**

- Among respondents: $45 \div 50 = 90\%$ reported liking the event.
- Among all members: $50 \div 100 = 50\%$ actually liked it in this invented, fully known population.
- Satisfaction and response were related; positive members were more likely to respond. More responses would not automatically eliminate that selection pattern.

WHAT A CAREFUL REPORT WOULD INCLUDE

- The exact survey question and when it was asked.
- Response rate and outreach procedure.
- Known differences between respondents and nonrespondents.
- Uncertainty and limits on generalising beyond respondents.
- An invitation to improve access to future feedback.

ONE SENTENCE TO KEEP

A survey result is evidence about the responses you obtained. Generalising it requires an account of who answered and who did not.

08 / WORKED EXERCISE

When the target changes the job

These are possible mechanisms, not inevitable outcomes. Metrics can focus attention; targets can also change what people choose to do.

SELECTIVITY

A school may stop entering borderline students for an exam to improve the measured pass rate without helping those students learn more.

SUBSTITUTION

A help desk may count an acknowledgement as a response, replacing resolution with quick contact when the response clock dominates.

SHORTENED HORIZON

A content team may favour rapid reach over projects that need months of experimentation.

SUPPRESSED SIGNALS

A team judged on low defect counts may feel less safe reporting defects, reducing the count without reducing the underlying risk.

DESIGN RESPONSE

Design incentives for learning rather than display: reward honest reporting, allow exceptions, review unintended consequences and keep a route to revise a target.

Classic frameworks often associated with these issues include Goodhart's law and Campbell's law. Their practical lesson is a caution about incentive sensitivity, not a claim that every target corrupts every measure.

09 / INTERPRETATION

One average. Many possible lives.

An average is a summary, not a map of a distribution. Similar means can conceal different experiences.

FICTIONAL TEAM A**FICTIONAL TEAM B**

Both teams have a mean of 10.

Team A has values 10, 10, 10, 10, 10: total 50 / five people. Team B has 2, 2, 2, 4, 40: also total 50 / five people. They share a mean while differing in median, concentration and variation.

QUESTIONS WORTH ASKING

- What is the median alongside the mean?
- What do the lower and upper parts of the distribution look like?
- Do groups or contexts differ in meaningful ways?
- Could disaggregated reporting expose or identify individuals?
- Are we comparing people with genuinely comparable opportunities?

HUMAN CONTEXT

Distribution is necessary but not sufficient. Numbers about different groups deserve privacy safeguards and context, not automatic stereotypes.

10 / PRINTABLE WORKSHEET

Make a metric legible

Use this one-page sheet during a review. Write the unknowns as unknowns rather than inventing reassuring answers.

1. WHAT HUMAN OUTCOME DO WE ACTUALLY WANT?

2. WHAT PRECISE THING IS COUNTED (NUMERATOR, DENOMINATOR, WINDOW)?

3. WHAT RELEVANT EXPERIENCES OR OUTCOMES ARE INVISIBLE?

4. WHICH BEHAVIOUR MIGHT THIS MEASURE REWARD UNINTENTIONALLY?

5. WHAT COMPANION EVIDENCE AND VOICE WILL WE CONSULT?

11 / PRINTABLE WORKSHEET

Write a decision memo

Convert a dashboard reading into a provisional, reviewable decision. Every box asks for an explanation, not a score.

DECISION UNDER CONSIDERATION:

WHAT THE METRIC SUPPORTS US IN SAYING:

WHAT THE METRIC DOES NOT ESTABLISH:

WHOSE PERSPECTIVE HAS NOT BEEN REPRESENTED?

WHAT EVIDENCE WOULD CHANGE THIS DECISION?

12 / CONVERSATION CARDS

Six questions that open a room

Choose language that is curious and precise, not accusatory. The purpose is a better decision, not to win an argument.

WHEN A KPI LOOKS WONDERFUL

"This is encouraging. What else improved at the same time, and who might not be represented here?"

WHEN SOMEONE BECOMES A NUMBER

"What did this measure actually observe about their work, and what should we avoid inferring from it?"

WHEN AN AVERAGE HIDES EXPERIENCE

"Could we inspect the spread, the median and carefully chosen groups before drawing a conclusion?"

WHEN YOU SUSPECT A DISTORTED INCENTIVE

"If we rewarded only this number, what sensible behaviour might become harder to do?"

WHEN DATA ARE INCOMPLETE

"Which uncertainty would most affect the decision, and what is a proportionate way to check it?"

WHEN YOU NEED A FAIRER PROCESS

"How can people affected by this conclusion correct it, without exposing their private information or facing punishment?"

A useful habit: give the strongest legitimate case for the original metric first. Then ask what it leaves open. Contextual critique is more persuasive when it does not pretend the observed number has no value.

13 / EVIDENCE DESK

Read further. Keep questioning.

These sources inform general statistical and measurement concepts. None validates the fictional outcomes in this field guide.

01 / FURTHER READING

OECD: How's Life? 2024

Well-being and the usefulness of distributional views beyond a headline mean.

02 / FURTHER READING

OECD: Guidelines on Measuring Subjective Well-being (2025)

Question design, survey interpretation and multidimensional evidence.

03 / FURTHER READING

OECD: PISA 2022 Reader's Guide

Scales, definitions and interpreting educational assessment data.

04 / FURTHER READING

Our World in Data: Work and Employment

Why definitions and measurement choices matter across statistical sources.

05 / FURTHER READING

Bradley et al. (2021): Unrepresentative big surveys

How selection effects can matter even in very large samples.

REUSE NOTE

Share, print and adapt the questions with attribution to “Humans Are Not Metrics — Sahil Tayal” and a link to sahiltayal.com. This guide is educational, not a validated assessment or a substitute for context-specific statistical review.

RETURN TO THE EXPERIMENT

humansnotmetrics.sahiltayal.online