

Discovery to Impact

What changed, and how do we know?

DENNIS GRIFFIN, EdD

LEARNING MEASUREMENT & EVALUATION



ACT I · THE MEASUREMENT ILLUSION

“The training was a success.”

Q3 PROGRAM REPORT

ILLUSTRATIVE

Completion 98%

Satisfaction 4.6 / 5

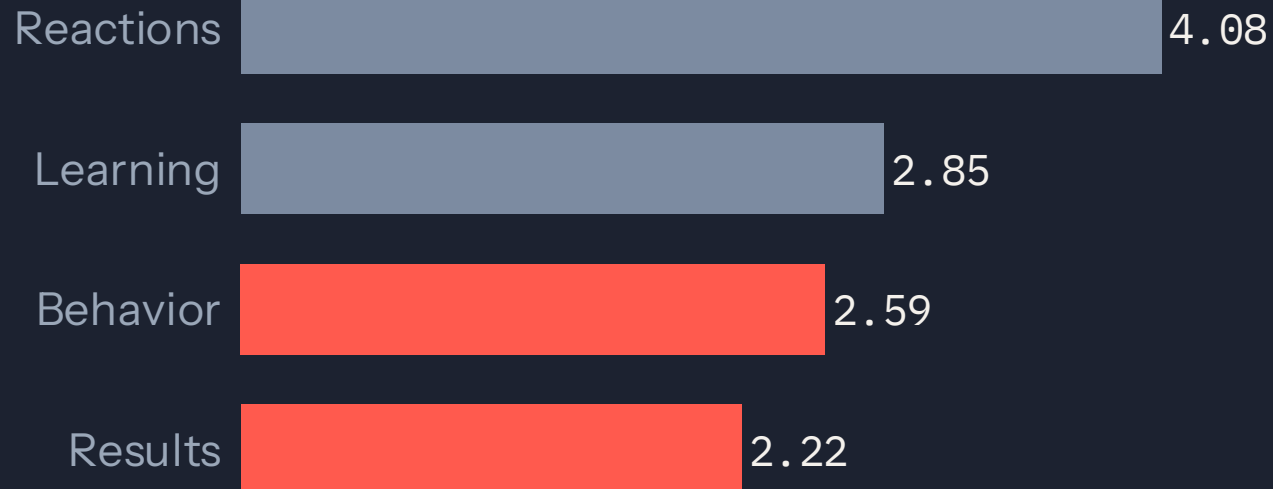
Assessment pass rate 87%

Recommendation · Scale to all regions

So what
changed?

We measure most what relates least to transfer

HOW OFTEN ORGANIZATIONS EVALUATE EACH LEVEL
1 = NEVER · 5 = ALWAYS



150 training professionals in Canada. Self-reported, correlational. Saks & Burke (2012)

RELATED TO REPORTED TRANSFER

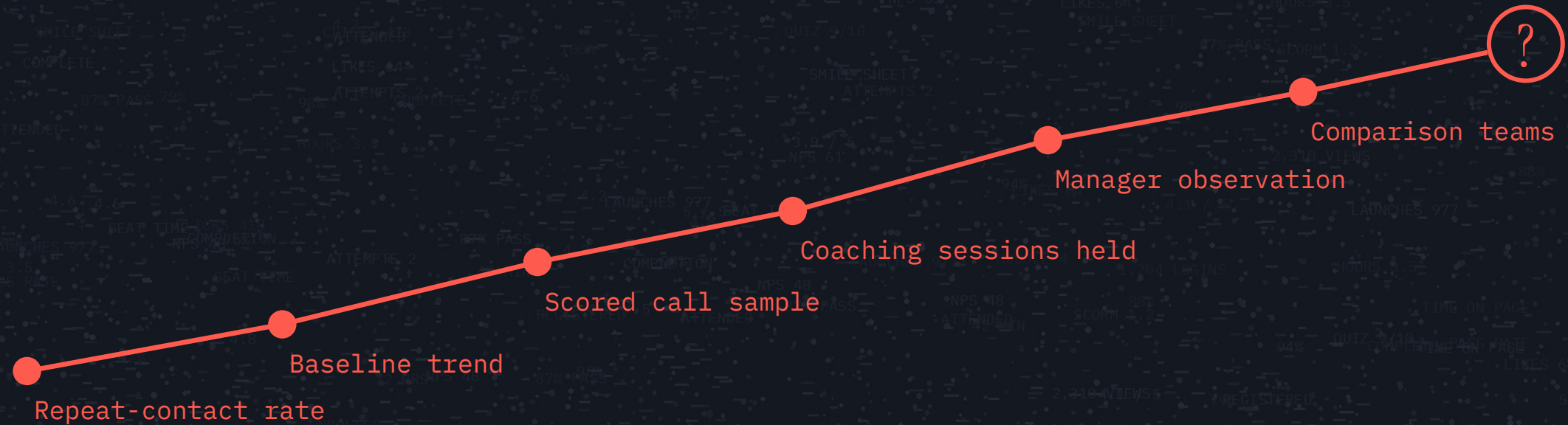
Only behavior and results

Evaluating them was linked to higher transfer right after training, at six months, and at one year.

CORRELATION AT ONE YEAR

Reactions	.02
Learning	.05
Behavior	.33
Results	.28

Which of these would help you make the call?



THE QUESTION THAT DEFINES THE SIGNAL

What changed in workplace performance, **what evidence supports that conclusion**, and what else could explain the results?

Three evaluations, three different questions

- TRAINING EVALUATION
Did they react, learn, and begin to apply it?
- TRANSFER EVALUATION
Did it become sustained, supported practice?
- BUSINESS IMPACT
How did it contribute to an outcome the business cares about?

DAYS TO WEEKS



L1 · L2 · EARLY L3

-3.9 / +5

MONTHS



SUSTAINED L3 + REQUIRED DRIVERS

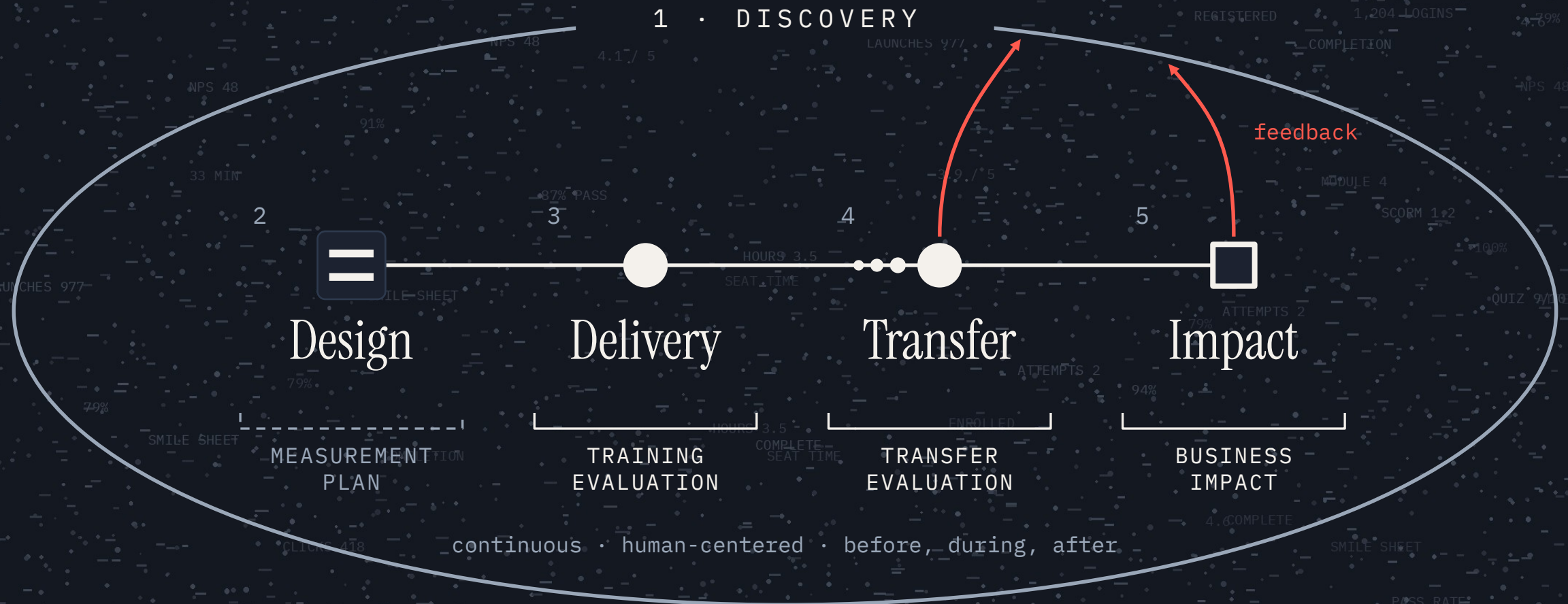
QUARTERS



L4 · CONTRIBUTION

MY INTERPRETATION Kirkpatrick's Level 3 splits in time. Early application is training evaluation. Sustained, supported practice is transfer. Each dashed link is assumed until shown.

The Discovery to Impact model



A practitioner model for organizing evaluation work, built on Kirkpatrick, Brinkerhoff, and transfer-of-training research.

Discovery runs the whole time

AFTER

What changed, for whom, and why?

Discovery
CONTINUOUS

BEFORE

What performance problem are we solving, and for whom?

DURING

What is getting in the way of people using this?

METHODS FROM UX RESEARCH

INTERVIEWS

OBSERVATION

CALL LISTENING

JOURNEY MAPS

SURVEYS

“Our agents need product training.”

12 HRS

SMILE SHEET

87% PASS
DOWNLOADS 51

THE REQUEST

REQUESTED BY

Contact center leadership

PROBLEM STATED

Repeat calls are climbing

SOLUTION REQUESTED

Product training for every agent by next quarter

SUCCESS MEASURE

Course completion

HOW I’LL TAG THE EVIDENCE



MEASURED

Evidence actually collected



ASSUMED

A proposition not yet adequately established



UNKNOWN

A question the existing evidence cannot answer

COMPOSITE SCENARIO · ILLUSTRATIVE · NOT DATA FROM A REAL ORGANIZATION

Discovery found a different gap



ASSUMED

~~Agents lack product knowledge~~



MEASURED

Repeat-contact rate, with a 12-month baseline



MEASURED • FROM DISCOVERY

Agents skip diagnostic questions to protect handle time

HOW WE LOOKED

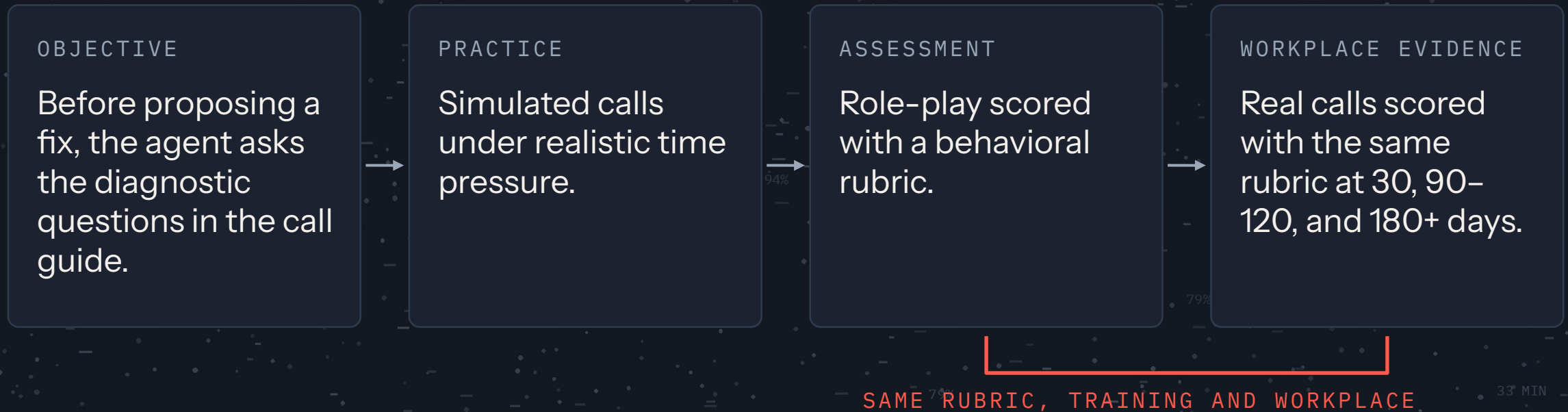
- Call listening
- Interviews with agents
- Supervisor interviews
- Workflow observation
- QA data review

CONTEXT • HANDLE-TIME TARGETS

WHAT CHANGED IN THE PLAN

Practice in diagnostic questioning • coaching for managers • a new look at handle-time reviews

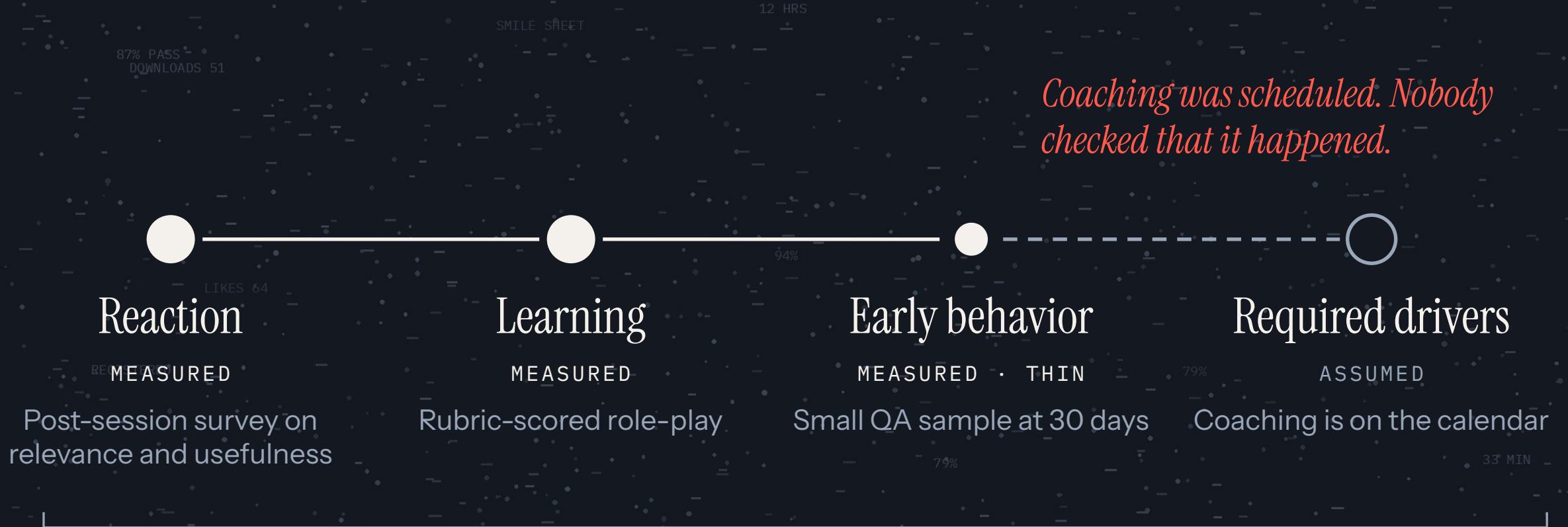
If you can't observe it, you can't evaluate it



FROM THE FIELD

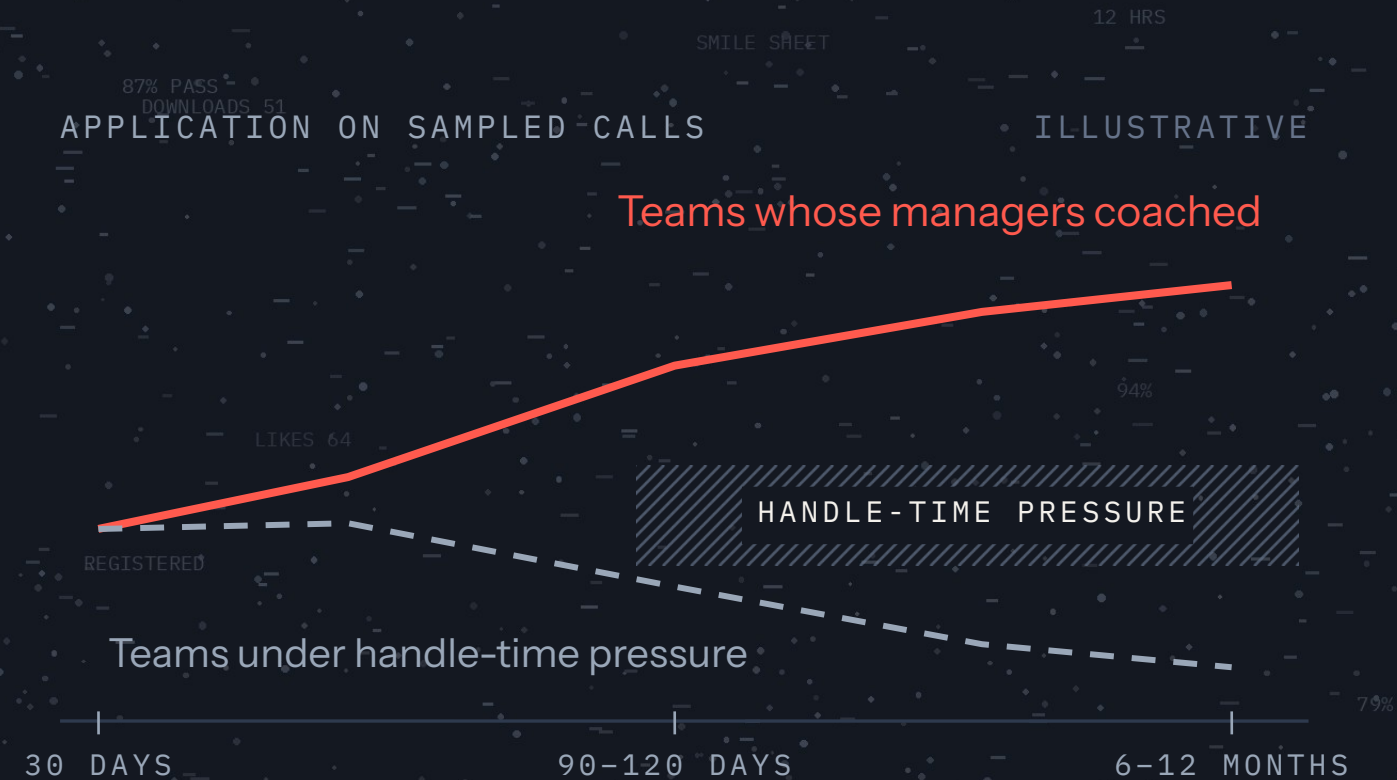
I've coached designers whose objectives didn't measure what their training taught. Good evaluation starts in design.

The first links look strong



TRAINING EVALUATION · KIRKPATRICK LEVELS 1 TO 3 AND REQUIRED DRIVERS

Transfer held where the work allowed it



WHAT EACH SOURCE CAN TELL YOU

- Self-report
What people believe they do
- Manager observation
One observer, one moment in time
- QA sampling, repeated
What people do on the job, over time

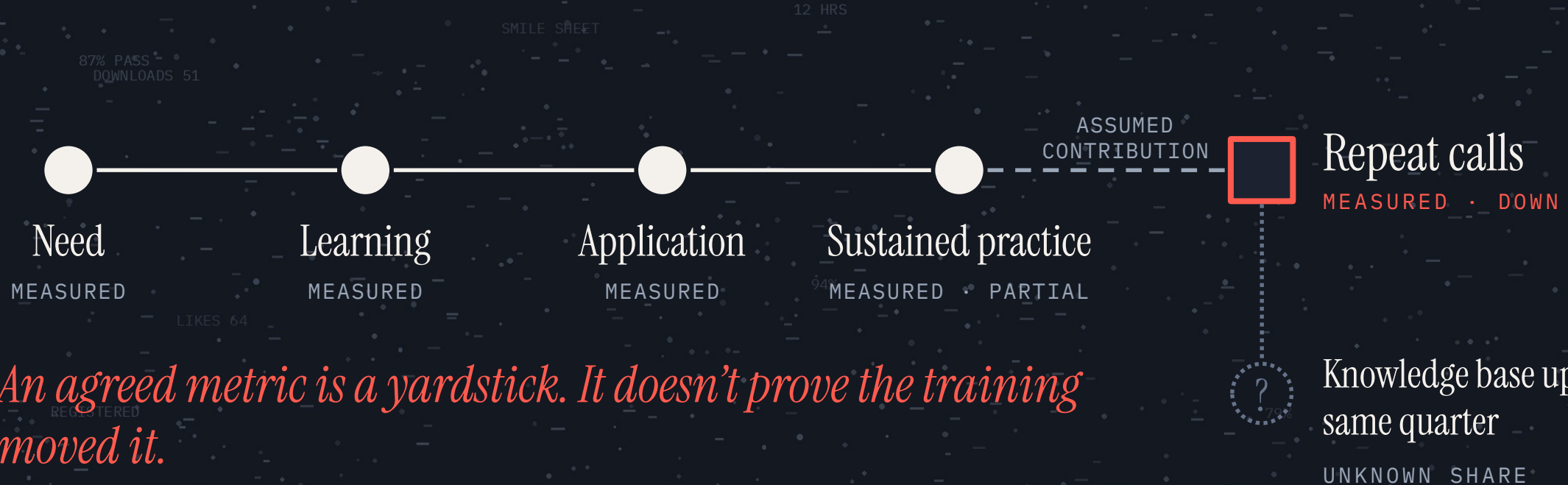
LEADING SIGNALS · IS TRANSFER LIKELY?

Coaching happening · chances to apply · job-aid use

LAGGING OUTCOMES · DID IT HAPPEN?

Sustained application · performance change

Repeat calls dropped. Now what can we claim?



RETURN ON EXPECTATIONS

Did we deliver what stakeholders agreed to in discovery?

Agreed expectations. A different thing from financial return.

SUCCESS CASE METHOD

How did it work where it worked, and what blocked it elsewhere?

Explains how impact happens. Can't estimate the average.

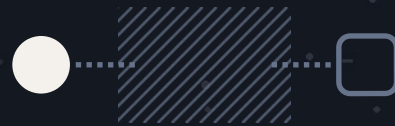
ROI

What was the financial return?

Only when a credible estimate is possible.

When the chain breaks, investigate

Learning improved. Behavior didn't.



INVESTIGATE

Opportunity to apply · manager support · reinforcement · competing priorities · workflow and systems · incentives

Behavior improved. The metric didn't.



The causal assumption · the time lag · whether the metric can detect a change this size

The metric improved. So did other things.

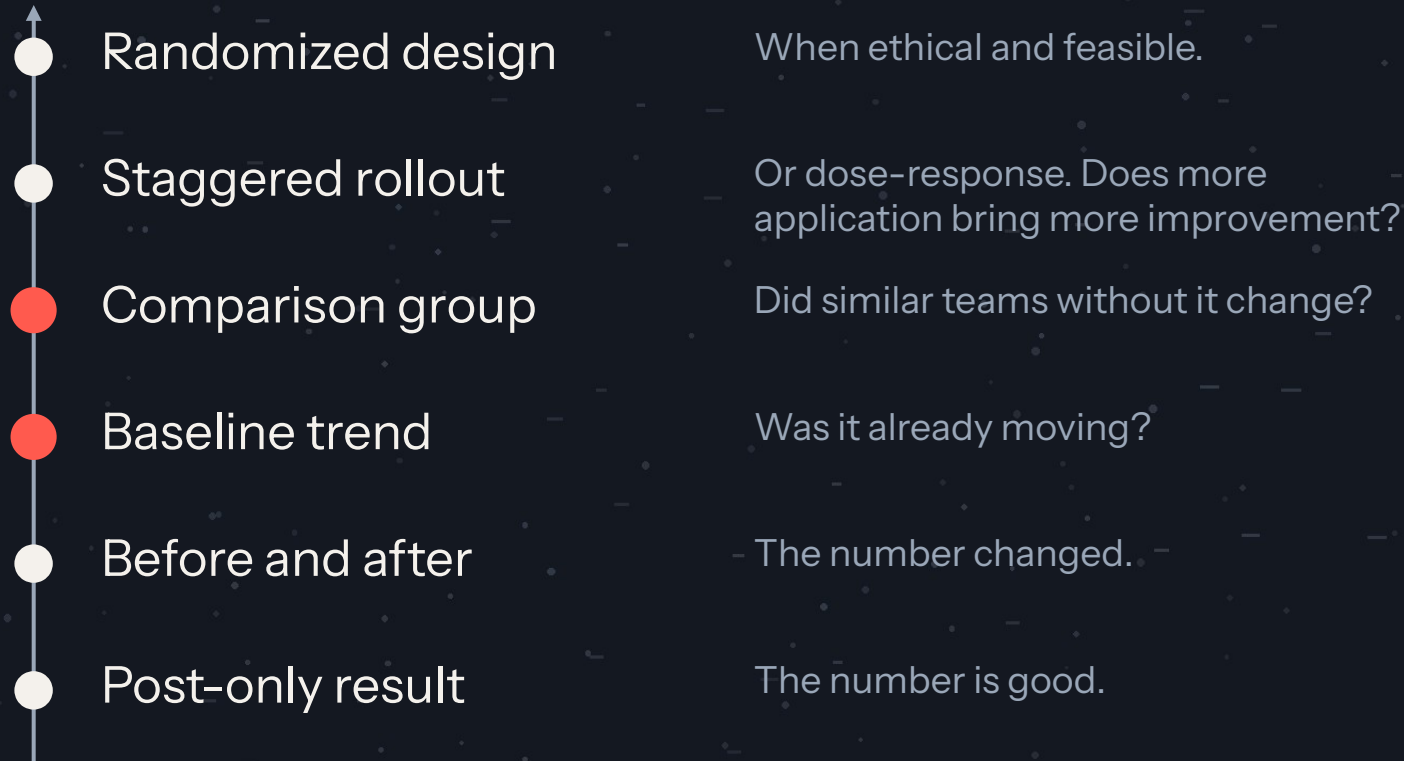


Other initiatives · seasonality · staffing · policy changes · a trend already underway

Good evaluation reduces uncertainty about performance.

Contribution is a matter of degree

STRONGER CLAIM



OUR SCENARIO

Baseline trend, plus teams with high and low application. A reasonable claim, still short of proof.

TRIANGULATE

QA SAMPLES

MANAGER OBSERVATION

BUSINESS METRIC



Where independent sources agree, confidence rises.

Same program. Better decision.

THE EXECUTIVE'S QUESTION

Roll the program out to every region?

WHAT THE REPORT SAID

- Completion 98% · satisfaction 4.6 / 5
- Repeat calls down

WHAT WE LEARNED

- Behavior held only where managers coached
- ⌚ Knowledge base updated the same quarter
- Handle-time review still rewards short calls

DECISION

~~Scale as is.~~

Scale with manager coaching. Change the handle-time review. Keep a comparison region for the next wave.

Discovery to Impact



Training effectiveness $\neq$ transfer

A great reaction score is not proof that behavior changed on the job.

Transfer $\neq$ business impact

Sustained behavior is not automatic proof that a business metric moved.

Evidence beats assumption

I compare relevant evidence and acknowledge what it can and cannot support.

Ask a better question

~~How successful was the training?~~

What changed in workplace performance, what evidence supports that conclusion, and what else could explain the results?

My goal is to understand what changed, what supported it, and what to improve next.

Think of a program your organization called a success. Draw its chain.

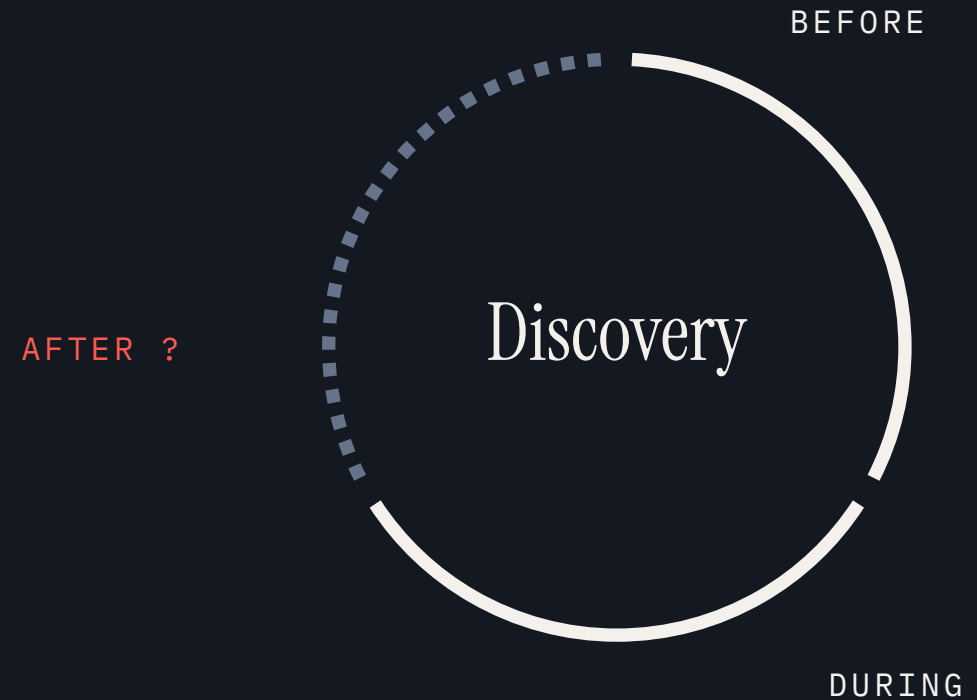
Tag each link measured, assumed, or unknown.



4 MIN · PAIRS

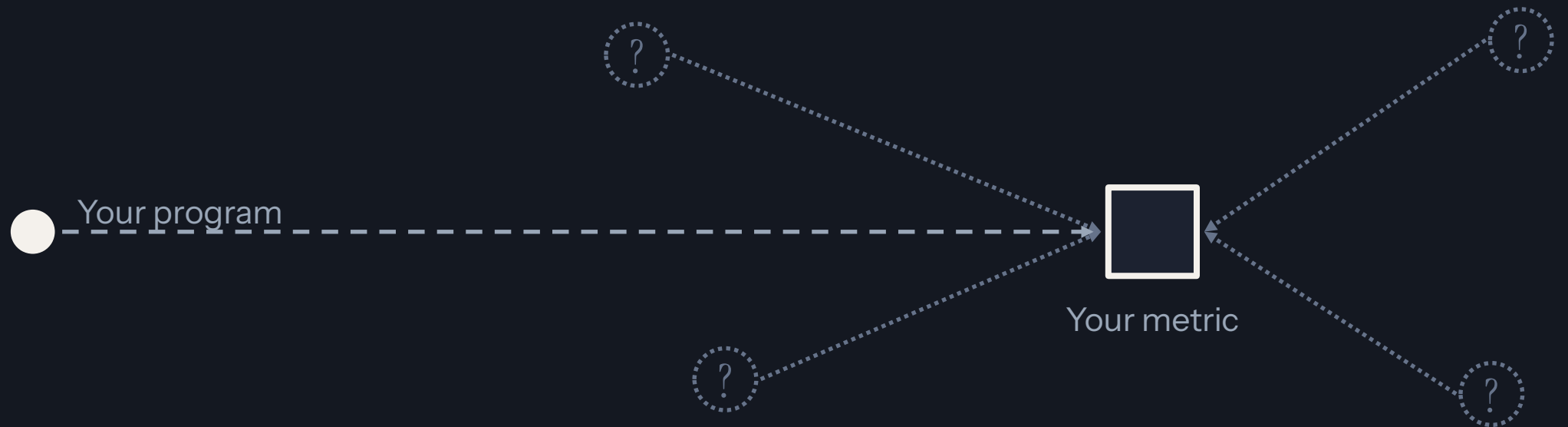
Where does discovery stop in your process today?

After launch, who is still talking to the people doing the work?



4 MIN · TABLE GROUPS

The last time a business metric moved after one of your programs, what else changed that quarter?

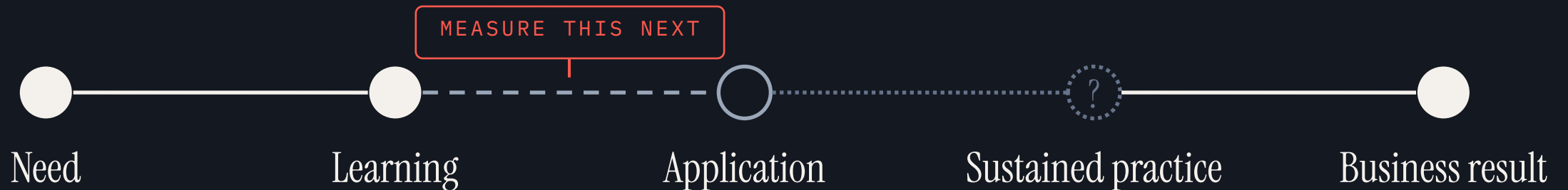


3 MIN · WHOLE ROOM

Would your organization welcome an evaluation that found training wasn't the main reason performance improved?

If not, what are your evaluations for?

Pick one program. Mark each link. Measure one assumed link next quarter.



The companion brief has a worksheet for this.

1 MIN · ON YOUR OWN

References

- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Jr., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50(2), 341–358. <https://doi.org/10.1111/j.1744-6570.1997.tb00911.x>
- Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, 41(1), 63–105. <https://doi.org/10.1111/j.1744-6570.1988.tb00632.x>
- Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, 36(4), 1065–1105. <https://doi.org/10.1177/0149206309352880>
- Brinkerhoff, R. O. (2003). *The success case method: Find out quickly what's working and what's not*. Berrett-Koehler.
- Grossman, R., & Salas, E. (2011). The transfer of training: What really matters. *International Journal of Training and Development*, 15(2), 103–120. <https://doi.org/10.1111/j.1468-2419.2011.00373.x>

References

Holton, E. F., III. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, 7(1), 5–21. <https://doi.org/10.1002/hrdq.3920070103>

Kirkpatrick, J. D., & Kirkpatrick, W. K. (2016). *Kirkpatrick's four levels of training evaluation*. ATD Press.

Kirkpatrick Partners. (2018). *Required drivers* [Job aid]. <https://www.kirkpatrickpartners.com/wp-content/uploads/2024/03/required-drivers.pdf>

Saks, A. M., & Burke, L. A. (2012). An investigation into the relationship between training evaluation and the transfer of training. *International Journal of Training and Development*, 16(2), 118–127. <https://doi.org/10.1111/j.1468-2419.2011.00397.x>

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

DISCOVERY TO IMPACT

Thank you

Dennis Griffin, EdD

LEARNING MEASUREMENT & EVALUATION

www.DennisRGriffin.com

Companion brief · Discovery to Impact: A Philosophy of Learning Measurement

