POSITION PAPER · EDITION 1

The Commercial Gap.

One overlooked measurement in a complicated system.

The start-up support system can report how many businesses it engaged. It cannot report whether any of them became more capable of winning and keeping customers. This paper argues that the missing instrument is tractable, and states what it does not claim.

EDITION	AUTHOR	PUBLISHER	METHOD VERSION
1, August 2026	Charles Talbot	Closing Foundry Ltd	v1.4

PERMANENT ADDRESS

fac16.com/papers/2026-08/the-commercial-gap/

CONTENTS

SECTION 0 · CONFLICT OF INTEREST

We build and sell a commercial capability measure. This paper argues that such a measure is missing from the start-up support system, and the one we would point you at is our own.

That interest should be weighed when reading everything below. Our view is that declaring it plainly is more useful than burying it, and that the reason we can see the gap at all is that we have spent several years trying to fill it and finding out what is hard about it.

> We are not asking anyone to adopt our instrument. We are asking that whatever gets adopted meets a standard that can be inspected, and we have published ours so it can be attacked.

EXECUTIVE SUMMARY

One cause sits underneath a great many of the others, and it is the one the support system measures least well.

Companies fail for many reasons. Undercapitalisation, the wrong team, a market that was not there, timing, bad luck, a founder who ran out of energy before the thing worked. Any account of business failure that reduces it to one cause is wrong, and this paper is not making that mistake.

But one cause sits underneath a great many of the others, and it is the one the support system measures least well: whether a company can generate interest, convert it into intent, turn intent into commitment, and then deliver enough value that the customer stays and buys again.

That capability is not currently measured anywhere in a way that can be compared between companies, aggregated across a portfolio, or observed changing over time. The Office for National Statistics measures management practice and does not look at it. Programme evaluations reach it only through what founders say about themselves in retrospect. Routing into support is decided by a mixture of self-reported symptom, adviser familiarity and whatever is currently funded.

The result is a system that can report how many businesses it engaged but not whether any of them became more capable of winning and keeping customers.

This paper proposes one thing: a short, free, reproducible reading of commercial capability that names what is getting in the way and decides which support a company is sent to. Not a replacement for capital, operators, framework thinking, or the many other things a company needs. One missing instrument, in one overlooked place, which happens to be a place where a lot of otherwise viable companies come apart.

We do not claim it makes companies succeed. We claim it lets a support system see something it currently cannot see,

and route on evidence rather than on guesswork. That is a smaller claim than the sector is used to hearing, and we think it is the largest one anybody can honestly make today.

SECTION 1

What this paper is not arguing.

Worth stating first, because the argument is narrower than the subject matter invites.

Not that commercial capability is the only thing that matters. A company with an excellent commercial engine and no product, no capital and no team fails, and deservedly. Building a business is a system with many interacting parts, and anyone selling a single-variable explanation of why companies live or die is selling something.

Not that measurement improves companies. It does not. What a company does after a reading is what changes anything, and most companies do very little after most interventions. That is the base rate and no instrument alters it.

Not that the existing support system is failing. Most of it is staffed by capable people doing sensible work with limited budget and limited time. The gap described here is structural rather than a matter of anyone's competence, and it is not solved by trying harder.

Not that our instrument is the answer. It is an attempt, it is published, and it is improvable in public.

Not that our instrument is the answer. It is an attempt, it is published, and it is improvable in public.

The argument is only this. There is a specific capability that decides a great deal, it is not measured, the absence of the measurement has knock-on effects that are visible throughout the system, and building the measurement is tractable.

SECTION 2

What we know.

Four findings from public evidence, none of them ours.

Access to markets is the largest reported barrier, and has been for most of a decade. The ScaleUp Institute's 2025 annual review[1] puts it ahead of talent and leadership at 55 per cent and finance at 42 per cent, with access to markets cited by 58 per cent. Its first report in 2014 named talent, markets, finance, leadership and infrastructure. Eleven years, and the list is the same list. Where finance is the live problem it is usually visible and usually addressed, because there is an entire apparatus for identifying and supplying it. The commercial one has no equivalent apparatus.

Management practice quality predicts firm performance, the state measures it, and the measure does not reach the customer. The Office for National Statistics runs the Management and Expectations Survey[2], scoring firms from 0 to 1 across continuous improvement, use of key performance indicators, use of targets, and employment practices. The mean was 0.55 in 2023, rising with firm size from 0.51 at 10 to 19 employees to 0.68 at 250 or more. The relationship with labour productivity is significant, and the experimental literature suggests much of it is causal.[3]

Those four dimensions measure whether a firm has discipline around metrics and people. They are agnostic about what the metrics are for. Nothing in the instrument asks whether the firm can generate interest, convert it, close, or retain. The survey is also aggregate and periodic, published as national statistics, with no individual firm told its own score or given anything to act on.

The same result holds outside official statistics and closer to the size of company this paper is about. McKenzie and Woodruff's business practices index, built across seven countries, finds a one standard deviation improvement in measured practice associating with roughly 35 per cent higher labour productivity in small firms.[6]

So the principle is established and it is established twice. Practice can be scored, the scores mean something, and the scoring has been pointed at general management rather than commercial capability, and at the economy rather than the firm.

Founders are not reliable narrators of their own capability. The World Management Survey[8], which scores management quality through structured interviews rather than questionnaires, ends by asking managers to rate their own establishment. Most over-estimate it, having just spent an hour discussing the exact practices they are being asked to score. The same body of research finds that not knowing is itself a common reason firms fail to improve.

This matters more than it first appears, and it cuts at us as well as at everyone else. Section 4 deals with that directly.

The best available UK trial leaves the central question open. The Growth Vouchers Programme ran in 2014 and 2015. More than 20,000 firms completed a diagnostic and chose one of five advice themes, and a randomly selected group received a voucher covering half the cost of advice up to £2,000. Analysis published by the Centre for Economic Performance[4] found turnover 8.2 per cent higher among firms that received and used a voucher, with no employment effect and no effect persisting beyond a year. The turnover effect appeared only among firms taking sales and marketing advice.[5]

It is tempting to read that as proof that the commercial dimension is where the return sits. It is not proof, and we will not present it as such.

Firms chose their own theme, so the firms that picked marketing may simply have been the firms already set up to grow. The trial did randomise the diagnostic itself, online against personal, and that produced differences in drop-out and take-up but not differences in outcomes, which is the one comparison in the study that bears on whether more thorough diagnosis produces better results, and it does not support the case made here. Most voucher users were in non-tradable sectors, so some of the gain may have come at the expense of unsupported neighbours. Only one firm in three offered a voucher used it.

What the trial establishes is narrower and still useful. The commercial theme is where the only detected effect appeared, we cannot tell whether that is the advice or the selection, and after a decade and twenty thousand firms nobody can settle it, because there was no objective reading at the point of triage to settle it with. The open question is the finding.

Support moves practice slowly, and this cuts against the whole sector including us. McKenzie's review of the business training literature puts the average effect on profits at around 10 per cent[7], and explains the field's many null results by noting that the programmes shift business practices only modestly.[6] That is the uncomfortable half of the evidence and we would rather print it than have it printed at us.

It has a direct consequence for how any measure of this kind should be read. If interventions move practice slowly, then a company showing large movement between two readings ninety days apart should be treated as suspicious rather than celebrated. The most likely explanation is that the respondent has learned the vocabulary, not that the business has changed. Any commissioner adopting a capability measure should expect small numbers, and should be told so before they buy rather than after.

SECTION 3

What the absence costs.

The support system already has a triage step almost everywhere. It is not a measurement.

The Business Growth Service launched in 2025 as a single national front door[9], taking businesses through a short set of questions and directing them onward. Growth Hub contracts specify a diagnostic assessment of business need. Accelerators and university programmes run intake forms. Something diagnostic-shaped sits at the front of nearly every route into support.

FIGURE 2 · COUNTED AT THE FRONT, COUNTED AT THE BACK

AT THE FRONT	BETWEEN THEM	AT THE BACK
The attendance register	**Nothing comparable**	**The accounts**
Applications, attendees,	No per-company reading of	Revenue, headcount, survival,

Applications, attendees, businesses supported. Counted at the door. It tells you the room was full.	No per-company reading of whether a company can win and keep customers, what is stopping it, or whether it got better.	Revenue, headcount, survival. Real, a year late, and silent on why.

The middle column is the only place anyone can still act, and it is the one place nothing is recorded.

Three consequences follow, none of which is anyone's fault.

The route is decided partly by the wrong input. A company is directed by some mixture of what it says it needs, what the adviser happens to know well, and what is currently funded and available. Founders describe symptoms, usually this month's most painful one. Pipeline is thin, so the answer is marketing. Deals stall, so the answer is a better deck. Sometimes the symptom and the cause coincide. Often they do not, and nothing in the process catches it.

Nothing aggregates. Two hundred adviser-led diagnostics produce two hundred incomparable judgements, each of which may be individually excellent. They cannot be added, compared across places, tracked over time, or used to see which parts of a portfolio are underserved. Commissioners are making portfolio decisions without portfolio data, and they know it.

Progression cannot be evidenced, so participation is reported instead. The independent evaluation of Scotland's flagship start-up programme, published in February 2026[10], shows the problem clearly and to its credit does not conceal it. Stage at joining was categorised by the delivery partner. Stage now was described by the member. The evaluators note that the additionality figure rests on members' own views of the contribution and is necessarily subjective.

That is a well-conducted evaluation doing the most defensible thing available to it. The programme was never asked to produce a per-company capability reading, so no such reading existed for the evaluators to use, and no evaluator in the country could currently do better with any comparable programme. The gap is in what was specified at the point of commissioning, not in the delivery and not in the evaluation.

And a commissioner who cannot observe progression will commission against throughput, after which the whole sector spends a decade apologising for throughput.

SECTION 4

The objection to answer first.

If founders over-estimate their own practice, how can a self-completed questionnaire measure anything?

It is the strongest objection to what follows, it applies directly to our own instrument, and any version of this paper that did not meet it head on would deserve to be put down.

Four things are true.

The objection is correct about absolute accuracy. A founder answering questions about their own company will, on average, answer generously. Our own method review says so in blunter terms than this paragraph does. Anyone claiming a self-completed instrument produces an objectively true score of commercial capability is overselling, and we are not claiming it.

The bias damages the level and largely spares the ranking. This is the substantive answer and it decides the design. Upward bias is roughly uniform within a respondent: a founder who inflates everywhere still inflates their weakest area least, because the bottom of a scale is genuinely uncomfortable to claim about yourself. So the ordering of a company's own strengths and weaknesses survives the bias even where the absolute number does not.

The diagnosis uses the ordering. It keys off the weakest stage and the weakest root cause rather than the average, which produces a deliberate asymmetry: the headline score is a level and is therefore the soft part, and the read is a ranking and is substantially immune. The number is the weaker half of the output. The named limitation is the stronger half. Any change that made the diagnosis depend on absolute level rather than relative position would forfeit that property, which is why it will not be made.

Item design moves answers toward facts. There is a large difference between asking a founder to rate their sales process and asking whether they reuse discovery questions that have worked before, or whether they write down why a deal was lost and let it change the next one. The second kind is closer to a fact about behaviour than an opinion about quality. That is a design discipline rather than a solved problem, and it is where most of the work in the instrument has gone.

Determinism gives consistency even where it does not give accuracy. Identical answers always produce an identical result. That is a weaker property than truth and a much stronger one than adviser judgement, which varies between advisers, between days, and between how well the conversation went.

There is a limit to how far the movement argument can be pushed, and section 8 states it. Within-respondent bias cancels well. Response shift, where a founder answers differently at the second reading because the first reading taught them what good looks like, does not cancel, and it inflates apparent improvement. That is the principal reason we say large short-run movement should be distrusted.

> THE HONEST SUMMARY
>
> This measures a founder's structured account of their own commercial practice, taken the same way every time, ordered so that the weakest part is identified reliably even when the overall level is generous. That is less than a laboratory measurement and considerably more than the sector has now. An instrument that overstates what it is will not survive contact with a risk function, so we would rather state the limit ourselves.

SECTION 5

What a measure would need to be.

Requirements before products, and with the contestable ones marked as contestable.

Free to the company. Charging the measured party suppresses volume, and without volume nothing can be compared. The parties who need the comparison should pay for it.

Deterministic and reproducible. The same answers always give the same result. A routing decision that cannot be reproduced cannot be audited, appealed, or defended, and no public body should deploy one that cannot.

Published in full. Questions, arithmetic, thresholds, version history, and enough test cases for an independent party to build a conforming implementation and check it. An institution with a risk function cannot adopt what it cannot inspect. Ours is at fac16.com/method/v1.4/, with the test vectors at fac16.com/conformance/.

Repeatable. One reading describes. Two readings measure change, and change is the only thing a commissioner can honestly count inside a budget period.

honestly count inside a budget period.

Firewalled from money. It routes, it never rations. No role in eligibility, grant decisions, or investment gating. A capability measure that becomes a funding filter gets gamed within a cycle and deserves to be.

Completable without an adviser. Contestable, and we should say so. A skilled adviser with an hour and good questions can probably read a company more accurately than any questionnaire. What the questionnaire has is consistency, cost, and reach: it works at a scale advisers cannot, and it produces something comparable, which adviser judgement does not. That is a trade of accuracy for coverage and comparability. We think it is the right trade for a triage layer and the wrong trade for the support that follows. Someone may reasonably disagree, and we would rather have that argument in the open than pretend this requirement is neutral.

SECTION 6

What we have built against that.

Described concretely so the argument is not abstract, and published so it can be tested.

The Growth Score reads a company across sixteen cells: four stages of the commercial job, against four root causes. The grid is at appendix A, and you can run the instrument yourself at fac16.com/run/.

The four stages follow the buyer rather than the seller's org chart.

Each is a transition the buyer makes, not an activity the seller performs.

Interest. Unaware to interested.

Intent. Interested to actively evaluating.

Commitment. Evaluating to committing money.

Value. Committed to realising the value they paid for, which is what decides whether they stay and buy again.

The four root causes are **Product, People, Process and Repeatability**. The stage is where the problem shows up. The root cause is why the buyer failed to make that transition. Product covers the claim being made and the thing being sold. People covers who owns the work and whether they can execute it. Process covers the routine and whether it holds under pressure. Repeatability covers whether what worked once gets captured and reproduced.

That framing is deliberate. Naming the stages after buyer states rather than sales functions means the same sixteen cells apply to a business selling to consumers, a founder-led enterprise sale, a product-led motion or a channel, without forking into incomparable versions. It also puts retention inside the measured job rather than outside it, because a company that wins customers and loses them has a commercial problem, not merely a customer service one.

It reports capability, what a company has built and can evidence, separately from momentum, whether the engine is producing output now. It returns a score from 0 to 100 in four bands and names one primary limitation. It takes about eight minutes, it is free to the company and always will be, and there is no generative model anywhere in the scoring path, which is why every result is reproducible and every routing decision can be audited. The report a company receives is at fac16.com/specimen/report/.

Question wording adapts to how a company sells. The cells and the arithmetic never change, which is what makes two very different companies comparable at all.

One consequence worth stating. What is measured is whether a company can win and keep the customers it is actually trying to win, not whether it resembles a venture-scale business. A durable eight-person company can read higher than a funded one. Companies are not all trying to become the same thing, and a capability measure should not assume they are.

The route falls out of the reading. The named limitation decides which support a company is sent to, rather than availability, adviser familiarity, or the founder's guess. That makes the decision legible in advance to the company, the adviser and the commissioner, and it makes it possible to say what a company was not sent to and why. A support system that never declines is a menu, and a menu hands the allocation problem back to the person who came in because they could not solve it.

The measure does not require our support to work. Movement between two readings can be produced by anyone: another provider, a university, a peer group, an adviser, or nobody at all. That is deliberate. It makes the instrument useful to whoever is already delivering, which is the only version of this that is worth building.

SECTION 7

What changes, and for whom.

For a commissioner. The unit becomes the reading rather than the programme. You can state what you are buying in terms a founder would recognise: this many companies measured, this distribution of limitations found, this proportion routed, this movement at the second reading, this proportion that did not move. Expect the movement to be modest and expect a substantial share of companies not to move at all, for the reasons set out in section 2. A supplier promising large short-run movement on an instrument of this kind is promising something the evidence does not support. Portfolio gaps become visible, because sixteen cells across a cohort is a picture and two hundred adviser notes are not.

For an existing provider. This is the group with most to gain and most reason to be suspicious, so it is worth being direct. A measurement layer does not replace delivery and does not judge it. It gives a provider something they currently cannot get: a defensible account of what changed in the companies they worked with, produced by an instrument they did not build and therefore cannot be accused of marking their own homework with. Every provider in the country is currently asked to prove impact using tools that cannot prove it. This is a way out of that, not an attack on the work.

For an evaluator. Progression stops being self-described at the point of evaluation and becomes something recorded at the time, twice. That does not solve attribution, and nothing does. It does separate the question of whether a company improved from the question of who caused it, which are currently tangled together in every evaluation in the sector.

For a founder. A named problem, in eight minutes, at no cost, with a route attached, and a second reading later to see whether anything moved. Most founders will do nothing with it. Some will.

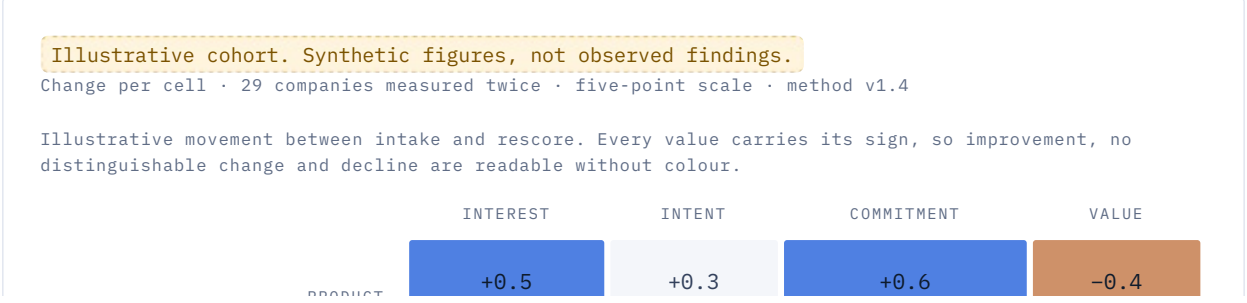
What does not change. The need for capital, for operators who have done it before, for introductions only a well-connected person can make, and for the framework thinking that helps a system reason about stages and needs at all. None of that is displaced. The argument is only that allocating any of it without a reading is guesswork with a budget attached.

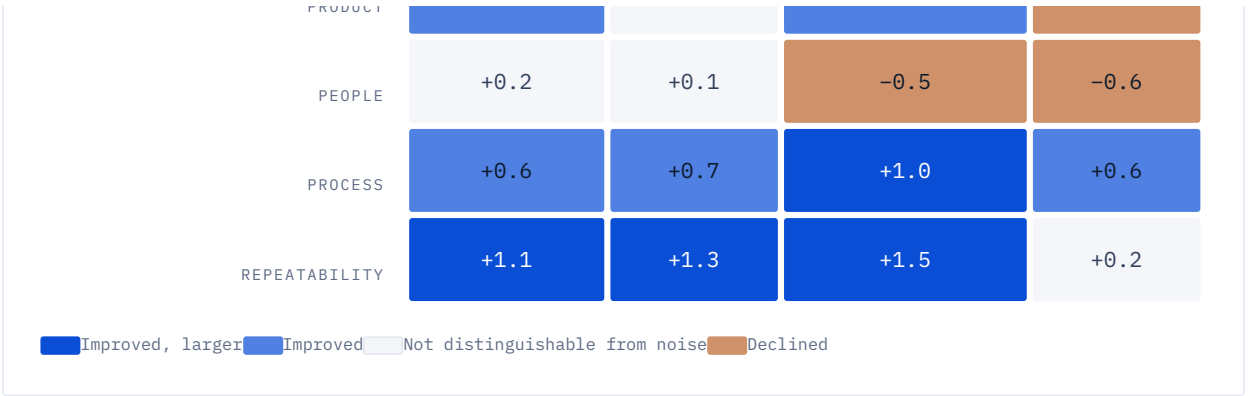
SECTION 8

Known limits, and what would prove this wrong.

A method that cannot be wrong is not a method. These are printed because the paper asks to be inspected, and an instrument that hides its assumptions has no business asking that of anyone.

FIGURE 3 · WHAT A MOVEMENT REPORT PRINTS ABOUT ITSELF





The legend is the argument. A report that can print “not distinguishable from noise” and “declined” about its own cohort is a report that can be believed when it prints an improvement. The full specimen is at fac16.com/specimen/cohort/.

Assumptions we have made and not proved

All sixteen cells are weighted equally. No weighting has been derived from data. Treating every part of the commercial job as equally important is a deliberate choice made for explicability, and it is almost certainly not true of any individual company. It is defensible as a stated choice and indefensible if left unstated.

Capability and momentum are blended equally in the headline number, which means each of the five momentum items carries several times the weight of each of the thirty-two capability items. Five self-rated answers therefore decide a large share of the score. This is one of the reasons we direct attention to the named limitation rather than the number.

The instrument is self-rated. The canonical instruments in this field are binary and behavioural, which is what makes them resist drift when a company is measured again. Moving to that format would improve the instrument and would break comparability with every reading taken so far. That trade is documented and unresolved rather than quietly avoided.

Ties are broken by a fixed rule. A company with no material weakness is still handed a named limitation. That is a known artefact of forcing a single answer, and it means a strong reading should be interpreted as strong rather than as having found something.

What would falsify it

Four conditions. Any of them holding would mean this is wrong, however internally consistent it looks.

Two competent observers of the same business produce materially different readings. Then the items are not measuring what they claim to.

The same company scores materially differently depending on which version of the question set it is given. Then adapting the wording to how a company sells has not preserved the underlying construct, and the comparability claim fails.

Readings do not predict anything about what subsequently happens to companies. Then the frame is wrong. Establishing this either way needs years and large numbers, and it has not been established.

Scores rise reliably after short interventions while the businesses visibly do not change. Then the instrument is measuring vocabulary rather than practice.

We do not currently have the evidence to rule out the third. We are stating the test so that someone can eventually apply it.

SECTION 9

What we are deliberately not claiming.

We are not claiming the Growth Score improves company performance.

We are not claiming to predict which companies will succeed, or that a reading should influence a funding decision. It should not, and we will say so in any room where it is suggested.

We are not claiming an objectively accurate measurement of capability. See sections 4 and 8.

We are not claiming to have proved anything at scale. The instrument is published, versioned and test-covered. The volume behind it is early and we will not dress it up.

We are not claiming the sixteen cells are the only defensible construct. They are ours, they are published, and they are published so that someone can propose a better set and be checked.

We are not aware of a longitudinal, comparable, per-company record of commercial capability held anywhere in the UK,

and we have looked. That is a statement about our searching, not a proof that none exists.

The payoff we do claim is narrow. Certain things currently do not exist and would exist: a per-company reading of commercial capability, a named limitation behind every routing decision, a reproducible record of that decision, and a measurement of change between two points in time. That is the whole claim.

SECTION 10

The ask.

Two things, of anyone commissioning or delivering support for growing companies.

Require a measurement standard behind what you fund. Not necessarily ours. Any instrument that is free to the company, deterministic, published in full, repeatable, and firewalled from funding decisions. Specify it as you would any other reporting requirement, and let providers compete on what they do with the reading rather than on how well they write a case study.

Engage with this one, including by attacking it. The method and arithmetic, the version history and the conformance test cases are published, and the instrument itself runs in your browser. If the cells are wrong they are wrong in public and can be corrected. An instrument that invites checking is worth more than one that does not.

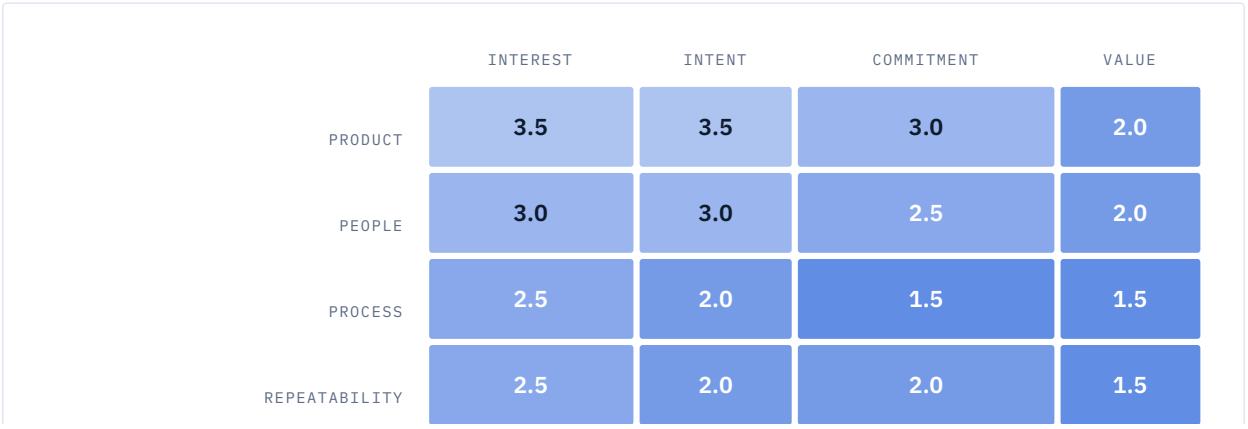
Companies will still fail. Most of the reasons will have nothing to do with anything measured here. But a system that can see which companies are struggling to turn interest into intent, or intent into commitment, or commitment into value the

customer actually receives, is a system that can put help in the right place more often than one that cannot. That is a modest improvement in one part of a complicated machine, and it is available now.

APPENDIX A

The sixteen cells.

FIGURE 1 · FOUR STAGES, FOUR ROOT CAUSES



Values shown are the worked example published in the method, on a five-point scale. Darker cells indicate lower scores, and every cell carries its number.

Two capability items per cell. Capability and momentum reported separately. The stage is where the problem shows up, the root cause is what is underneath it, and a company can be weak in one stage across every cause or weak in one cause across every stage. Those are different problems and they need different support. The published method sets out the items, the arithmetic, the bands and the rules by which the primary limitation is named.

APPENDIX B

Sources.

1ScaleUp Institute, **ScaleUp Annual Review 2025**, and the 2014 ScaleUp Report barrier list. scaleupinstitute.org.uk

2Office for National Statistics, **Management practices in the UK**, 2023 release and corrected dataset. ons.gov.uk

3Office for National Statistics and ESCoE, **Management practices in Great Britain 2016 to 2020**, on the relationship with productivity and the experimental evidence for causality.

4Centre for Economic Performance, LSE, **Does subsidising business advice improve firm performance? Evidence from a large RCT**, Discussion Paper 1977, January 2024. cep.lse.ac.uk

5What Works Centre for Local Economic Growth, **Lessons for business support schemes from a large RCT in the UK**,

February 2024. whatworksgrowth.org

6D. McKenzie and C. Woodruff, **Business Practices in Small Firms in Developing Countries**, Management Science, 63(9), 2017, pages 2967 to 2981, on the association between measured business practices and productivity in small firms across seven countries, and on the modest movement in business practices that explains the field's null results. doi.org/10.1287/mnsc.2016.2492

7D. McKenzie, **Small business training to improve management practices in developing countries: re-assessing the evidence for 'training doesn't work'**, Oxford Review of Economic Policy, 37(2), 2021, pages 276 to 301, on the average effect of training on profits. doi.org/10.1093/oxrep/grab002

8World Management Survey, **The World Management Survey at 18**, IZA Discussion Paper 14146, on manager over-estimation and on lack of knowledge as a barrier to improvement. docs.iza.org

9Department for Business and Trade, **SME Action Plan 2025 to 2028**, on the Business Growth Service. gov.uk

10Scottish Government and EKOS, **Early Evaluation of the Techscaler Programme 2022 to 2024**, February 2026. gov.scot

ABOUT THIS EDITION

Editions behave the way the method behaves. This edition keeps its dated address permanently and is never amended in place. Superseded editions stay live indefinitely, and fac16.com/papers/the-commercial-gap/ always resolves to the current one.

This edition is the argument. It publishes no distribution of readings, because the volume behind the instrument is early and a distribution drawn from it would mislead. A later edition will publish that distribution once the volume threshold set out in the method is met, and not before. Stating the threshold in advance is deliberate, so nobody has to wonder whether we waited until the numbers looked flattering.

Corrections to this edition are published as a dated note here rather than as a silent edit. If you think the argument is wrong, the method is at fac16.com/method/v1.4/ and the test cases are at fac16.com/conformance/.