TULSER

No. 07 · © 2026 Tulser

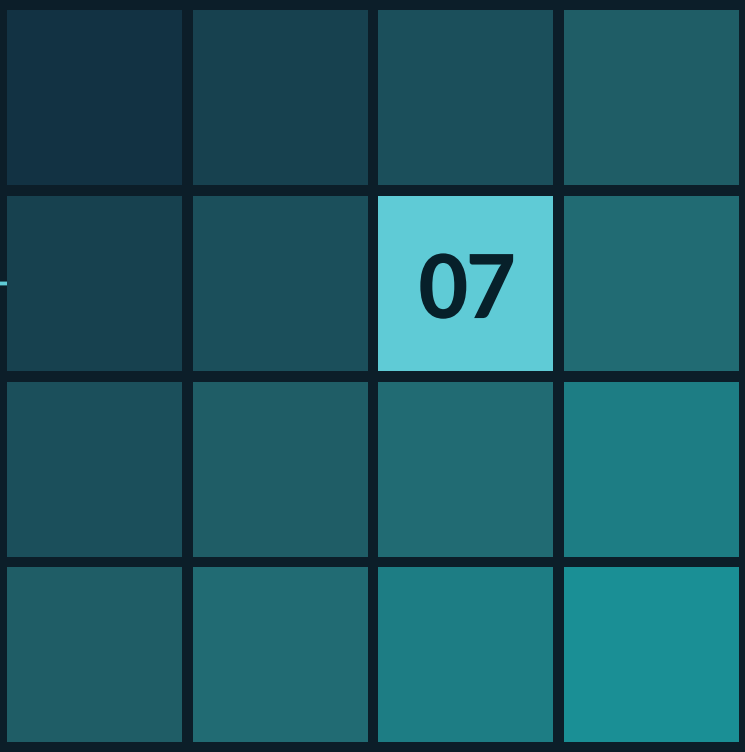
WHITE PAPERS · PERFORMANCE CONSULTING

BUSINESS PROBLEM · QUESTION 3

Somebody in this organization already gets the result. Nobody has asked them how.

The exemplary *performer*

Why the exemplary performer is an instrument rather than a role model, how to find one without holding a popularity contest, and how what they have gets given to everybody else.



JOS ARETS

with Vivian Heijnen

IN THIS PAPER

You cannot distribute a personality.
You can distribute what made one possible.

THE QUESTION — *Somebody is already producing the result under the same conditions. Who are they, what do they actually have that the others lack, and how much of it can be given to the rest of the unit?*

FOR — Anyone who owns a result or is asked to fix one: line managers and directors, operations, quality and continuous improvement, OD, HR business partners and L&D.

READ — 22 minutes · with a worked case and a one-page instrument.

01	The experiment that is already running	Same job, same rules, two results, and Gilbert's ratio for how much is at stake.
02	Expert or exemplary performer	Not the same person, not the same use, and one of them is easy to find for the wrong reason.
03	Finding one without a popularity contest	The metric comes first. Then the names, then the comparison group.
04	They cannot tell you what they do	Why the interview fails, what the research shows goes missing, and the four ways to get a record instead.
05	Reading the difference	Four qualities that look like character and are conditions of the work.
06	From one person to the average of a unit	The step almost every attempt skips, and why copying behavior fails.
07	The lending officer who sent one email	A worked case, and the part of it that stayed unexplained.
08	What this does not claim	Five limits, because the rest is not credible without them.
.	Run it yourself	The comparison, on the problem in front of you.

The exemplary performer is not a model to copy.
They are a reading of what everybody else is missing.

The experiment
that is already running

In almost every unit, some people get the result and others do not: under the same rules, with the same systems, the same customers and the same manager. That is as close to a controlled comparison as an organization ever produces, it has already been run, and nobody has read the result.

Gilbert built his method on it, and gave the comparison a number. Then he did the thing that matters more: he subtracted. Take what the exemplary performer does and has, take away what the typical performer does and has, and read what is left.

What is left is almost never a personality. It is a list of conditions the organization gave to one person and withheld from the others, usually without deciding to. That list is the finding, and unlike a personality it can be handed out. The number comes first, though, because it decides whether the subtraction is worth doing at all.

> **WHY THIS QUESTION SITS THIRD AND NOT FIRST**
>
> It is tempting to start here, because it is the pleasant question and everyone enjoys it. It sits third because it needs the first two: without a measure you cannot say who the exemplary performer is, and without a gap you do not know how much of the difference matters. Asked first, "who is good at this?" collects opinions about people. Asked third, it collects evidence about the work.

One caution before anything else. This is a comparison, not an experiment. Nobody randomized the territories, the customers, the tenure or the manager. Establishing that the two groups are genuinely comparable is the first task of the analysis, not a formality, and section 08 says what happens when they are not.

Gilbert's number, *and what it is for*

The performance improvement potential is the oldest instrument in this discipline and one of the most misused. It is a ratio, it takes two minutes to compute, and its purpose is the opposite of what its name suggests to most readers.

Gilbert defined it as the ratio between exemplary and typical performance on the same measure (Gilbert, 1978). Not between the best and the worst: between the best and the ordinary, because the ordinary is where the volume is. A PIP of 2 says the exemplary performer produces twice what the typical one produces. A PIP of 1 says there is nothing to find here and you should go and look somewhere else.

THE MEASURE	TYPICAL	EXEMPLARY	PIP
Decision time on a loan file lower is better, so invert	17.2 days	8.1 days	**2.12**
Complete roadside reports higher is better	40%	71%	**1.78**
Two-zone orders inside promise higher is better	0%	96%	**undefined**

Three computations from cases elsewhere in this series. The third is the one worth dwelling on: when the typical performer scores zero the ratio breaks, and the finding is that nobody has ever done this except one person (a different and larger discovery).

Two mechanics matter and both are routinely got wrong. **Direction.** Where a low number is good (days, errors, cost, complaints) the ratio is typical over exemplary, not the other way round. Reported the wrong way a PIP of 2.12 arrives as 0.47 and reads as a problem rather than an opportunity. **The comparison group.** Gilbert's denominator is the typical performer, not the mean of everybody including the exemplar, and not the worst. Using the worst inflates the number and makes the whole finding easy to dismiss.

What the ratio *cannot tell you*

A number this easy to compute invites more weight than it can carry. Four cautions, and then the way to use it that survives all four.

THE CAUTION	WHAT GOES WRONG
It says nothing about cause	A PIP of 3 tells you the distance and not one thing about why it exists. Treated as a finding rather than as a reason to start looking, it produces confident nonsense.
It inherits the measure's faults	Computed on a measure that counts the wrong thing, it gives a precise number for an irrelevant difference. Establish the measure first, that is what number 05 is for.
It can be an artifact of conditions	An easier territory, a newer system, a quieter branch. Then the PIP is measuring the environment's unevenness, which is a real finding but a different one.
Not all of it is available	The ratio describes what one person achieves, not what the unit could achieve. Part of the difference will turn out to be undistributable, and the honest version of the number is the part that is not.

Four ways a two-minute calculation misleads. None of them is a reason to skip it; all of them are reasons to report it with its conditions attached.

Using the ratio
so that it survives the room

None of those four cautions is a reason to skip the calculation. All of them are reasons to report it with its conditions attached, and to be careful about the sentence it arrives in.

So use it as a sorting instrument rather than a conclusion. Compute it across the jobs that touch the result, take the largest, and spend the analysis there. Report it as a range rather than a point ("between 1.8 and 2.1, depending on whether the two newest advisers are included") because a range invites the argument about comparability that the number needs to survive anyway.

And state it in the form that carries the argument. Not "our top person is twice as good", which is a sentence about a person and provokes a defense of everybody else. But: **"the distance between what this unit produces and what one of its own members produces is a factor of two, under the same conditions."** That sentence is about the organization, which is where the finding actually lives.

> **WHAT GILBERT WANTED IT FOR**
>
> Not to rank people. To rank *jobs*. Compute the PIP across several roles and the largest one tells you where an analysis will pay for itself: the place where the distance between the ordinary and the achievable is widest. His argument was always about lifting the mass in the middle: a large PIP is not a compliment to the exemplar, it is a measure of how much the organization is currently leaving on the table with everybody else, which is why the number belongs in a conversation about the unit and not in anybody's appraisal.

Expert
or exemplary performer

These are routinely treated as the same person and they are not. The difference decides who you interview, and interviewing the wrong one is the commonest way this question is wasted.

	THE EXPERT	THE EXEMPLARY PERFORMER
Defined by	Recognition. Someone whose judgement is accorded authority and standing by peers or by the public.	Output. Someone who consistently produces the outcome better than a standard performer, under the same conditions.
Visibility	High. Everybody can name them, inside the organization and often outside it.	Frequently low. Organization-level indicators aggregate individual contribution away, so nobody has noticed.
What they give you	A standard: what good looks like, and where the quality and safety limits are.	A condition: what the work actually requires, discovered by someone who found it themselves.
The risk	They may describe the job as it is supposed to be done rather than as they do it.	They may be unable to tell you what they do at all, because it has become automatic.

An expert may also be an exemplary performer. The reverse holds far less often, and the question decides which one you need.

Ask a manager for “the expert” and you will be given the most visible person: the one who runs the induction, answers the questions, sits on the working group. Ask for “whoever produces the outcome best, on the measure” and you will sometimes be given the same name and often not. The second question is the one this analysis needs.

And a word *about deliberate practice*

The popular move at this point is to turn exemplary performers into experts through deliberate practice. How much practice explains is an open debate in the research, and it is not close to being settled, which does not leave a practitioner without guidance.

The debate in one sentence: Hambrick, Macnamara and Oswald put the variance it explains at about 14 percent overall and roughly 1 percent in professions (Hambrick et al., 2020); Debatin and colleagues reply that the meta-analyses understate it by pooling studies in which the practice was not individualized (Debatin et al., 2021). Both sides are arguing partly about measurement, so what would move the argument is not another pooled estimate but studies that define the practice tightly and observe it in professional work rather than in chess and music. That research is worth doing and has not been done at any scale.

IT APPLIES WHERE	IT DOES NOT WHERE
The task is the same every time, so improvement accumulates on one thing	**The task changes with the customer, the case and the week**
The standard is fixed and public: a time, a score, a piece of music	**“Good” is contested and depends on who is judging**
Feedback arrives in seconds, so the correction attaches to the attempt	**Feedback arrives in a quarter, attached to an outcome with many other causes**
Practice can be separated from production, and failing costs nothing	**Every attempt is live work with a customer at the end of it**

Chess, music and sport meet all four. Most professional work meets none.

WHAT IT IS WISE TO TAKE ACCOUNT OF MEANWHILE

Three conclusions hold whichever way the debate goes. That deliberate practice works on one person at a time and on one of the six conditions, while this question exists to find what can be handed to a whole unit at once. That the four tests above decide whether your setting resembles the ones the research was done in. And that a practice program proposed *before* the subtraction in section 05 is a remedy chosen without a diagnosis, whatever the evidence eventually says about the remedy. So the question to put to the proposal is not “does the evidence support deliberate practice?” It is: **has anybody done the subtraction yet?**

> *If you pit a good performer against a bad system, the system will win almost every time.*

GEARY A. RUMMLER & ALAN P. BRACHE

Improving Performance, 3rd ed., 2013, p. 11

Finding one
without a popularity contest

The failure mode here is nomination. Ask who the good ones are and you will be handed the well-liked, the long-serving and the loud. The method that avoids it is not complicated, and its first step is the one that gets skipped.

Establish the metric before you ask for names. Which number distinguishes key from standard performance in this job, and what value does each of them sit at? Only then do you ask a manager for names, and you ask them against the metric rather than against an impression. Four criteria have held up across our own analyses: the person produces the outcomes in question markedly better than a standard performer, has met the metric consistently for more than two quarters, is recognized by their own team as well as by the organization, and can say what the organization is trying to achieve. The third criterion filters out the performer whose numbers come at someone else's expense; the fourth filters out the one who cannot explain a single choice they make.

THE DECISION	WHAT THE METHOD SAYS
How many	One to three percent of the workgroup, with two percent as the target. In a group of a thousand, twenty to thirty. The sample is purposive, not random: widening it drags the sample down the curve and dilutes exactly what you came for.
When to stop	When the data repeats. If the same behaviors and decisions are recurring by the time you are half to two-thirds through the list, stop and record that you did.
If nobody does all of it	Select several, each for the part of the outcome they are exemplary at, and interview each one only about that part.
If nobody has ever achieved it	Take those closest, record why the manager chose them, and treat the whole finding as weaker. A desired level nobody has reached is a design problem, not a benchmark.
The step people skip	Select standard performers too. Without them there is no comparison, and without a comparison you have a profile of an admirable colleague.

Selection and sampling rules from our own practice, not from a published study. The last row is the one that decides whether this is an analysis or an appreciation.

The exemplary performer *cannot tell you what they do*

This is the point on which the whole question turns, and it is the point most approaches get wrong. You cannot copy what somebody achieves by asking them how they achieve it. Not because they are unwilling, because they do not have access to the answer.

The reason is how skill is stored. Work that is repeated thousands of times stops being a sequence of decisions and becomes a single automated routine. The steps are still executed; they are no longer represented in a form the person can retrieve and put into words. What returns when you ask is not the procedure but a reconstruction: a plausible account assembled at the moment of asking, out of what the person believes must have been going on.

The size of the omission is now documented. A systematic review of cognitive task analysis in clinical and health services research states it plainly: experts may leave out **up to 70 per cent** of the components and decision steps needed to perform a task successfully, with the consequence that protocols and training built on their accounts are incomplete or ineffective (Swaby et al., 2022). Seventy percent is the upper end of a range rather than a constant, and the review says so, but even the lower end of that range destroys the assumption that an interview will do.

This is not a new discovery so much as a newly measured one. Nisbett and Wilson established the general case in 1977: when people report on their own mental processes, those reports are frequently not based on introspective access at all but on a theory of what would plausibly have influenced them (Nisbett & Wilson, 1977). What has changed since is that the effect has been measured in real work rather than argued from the laboratory, and that a method exists for getting round it.

> **WHAT THIS MEANS IN PRACTICE**
>
> The interview with your exemplary performer will go well. They will be helpful, articulate and generous with their time, and they will describe a version of the job that is coherent and largely fictional, not deceitful, but reassembled. If your method for copying the best consists of asking them how they do it, you have built the whole thing on the least reliable instrument available.

How much *goes missing*

The field that has measured this most carefully is surgery, because there the cost of an incomplete description is immediate and countable.

Clark and colleagues put expert surgeons in front of a procedure they had performed hundreds of times and asked them to describe it for a trainee. Then the same procedure was reconstructed using cognitive task analysis (structured probing, observation and comparison across several experts) and the two accounts were set side by side. The reconstructed version contained steps and decision points that the surgeons' own descriptions had left out entirely (Clark et al., 2012). Not details of style: actions they perform every time and had stopped noticing.

The method used to recover them has since been reviewed for organizational research specifically, which matters here because most of the older evidence comes from medicine and the military. Brown, Power and Gore set out what cognitive task analysis is for and where it applies in organizations: it is a family of techniques for eliciting the cognition that sits behind skilled work in its own context, precisely because that cognition is not available to interview (Brown et al., 2024).

And the consequence is measurable rather than theoretical. Training built from a reconstructed description rather than a self-report produces better performance, in general instructional settings (Tofel-Grehl & Feldon, 2013) and in surgical training specifically (Edwards et al., 2021). Something real is being recovered, and it is exactly the part the expert could not say.

> **WHY THIS IS THE HEART OF THE QUESTION**
>
> Every organization that has tried to copy its best people has run into this without naming it. The star gives a talk, writes the guide, mentors the team, and the effect is disappointing, so the conclusion is drawn that the difference must be personal after all. It is not. The difference was real; the method of extracting it was the one method guaranteed to miss most of it.

What to do
instead of asking

The remedy is not a better interview. It is to stop relying on recall and start relying on the record, and there are four ways to get one.

METHOD	WHAT IT RECOVERS	WHAT IT COSTS
Watch one complete piece of work	The actual sequence, including the steps nobody mentions. Write down every action, every decision and every point where the work waits.	Half a day per performer, and the discipline not to interrupt.
Interrupt at the decision, not afterwards	"You just did something, what were you looking at?" Asked within seconds it retrieves the cue; asked in a debrief it retrieves a theory.	Nothing, but it takes practice not to ask why.
Compare across several exemplary performers	What all of them do is procedure. What only one does is personal style. A single interview cannot tell those apart.	Three or four observations rather than one.
Ask the standard performer where it went wrong	The failure points. People remember trouble far better than routine, so the weaker performer is often the better witness to the obstacle.	Nothing, and it is the most neglected source in the room.

Four ways to obtain a record rather than an account. The fourth is free, and in most analyses it produces the sharpest single finding.

The fourth deserves its own sentence, because it inverts the usual instinct. The exemplary performer has automated the obstacle away and can no longer see it. The standard performer is standing in front of it every day and can describe it precisely. Ask one what they do and the other what stops them, and the two answers assemble the finding between them.

Reading
the difference

Now do the arithmetic. Put what the exemplary performer does and has next to what the standard performer does and has, and subtract. What remains is the whole value of this question, and the subtraction is more literal than it sounds.

Work down four columns rather than writing a description. **What they produce**: the outcome and its number, for each. **What they do**: the observed sequence, from the record you built in section 04, step by step. **What they have**: every tool, screen, permission, list, contact and piece of information either of them touched. **What they know**: the rules and cues you saw them apply, which is the shortest column and the last one to fill, because most of what looks like knowledge turns out on inspection to be one of the other three.

THE EXEMPLARY PERFORMER
Decides a small-business loan file in 8.1 days

- Sends one consolidated request for information
- Knows which two fields hold the file up
- Has the credit desk's direct line

THE STANDARD PERFORMER
Decides the same files in 17.2 days, the unit average

- Sends each question as it arises: 3.1 rounds
- Learns which fields matter by being sent back
- Uses the shared mailbox, answered next day

WHAT IS LEFT WHEN YOU SUBTRACT

Reads as a trait	*Is in fact a condition of the work*	*Can you give it to everybody?*
Initiative	**A screen that invites one question at a time**	**yes**
Experience	**A rule nobody wrote down**	**yes**
Good with people	**An escalation route that is not published**	**yes**
Judgement	**Nine years of seeing the same files**	**no: this part is real**

Three of the four are not personal at all. They are things the organization gave to one person and withheld from thirty-three others, and it can stop withholding them on Monday.

One job, two people, one subtraction. The middle column is what the analysis reports; the left column is what an appraisal would have recorded instead.

Doing the subtraction, *line by line*

The figure opposite is the finished result. Getting there is a mechanical procedure, and it is worth setting out because the mistakes people make in it are all the same three.

STEP	WHAT YOU DO	THE MISTAKE TO AVOID
1 · Line up the sequences	Put the two observed sequences side by side, in order, and mark every point where they diverge: a step one takes and the other does not, or takes differently.	Comparing summaries instead of sequences. Two summaries always agree; two sequences rarely do.
2 · Ask what made each divergence possible	At every divergence: what did the first one have at that moment that the second did not? A screen, a number, a name, a rule, a permission, an amount of time.	Accepting "she just knows" as an answer. Something made it knowable. Find what.
3 · Name each item as a thing, not a quality	Write it as a noun the organization could buy, publish or configure. "The credit desk's direct number", not "good relationships"	Leaving it as an adjective. An adjective cannot be assigned, costed or delivered.
4 · Mark what is left over	Whatever will not convert into a noun is the residue: genuine accumulated judgement. Write it down as such and stop trying to distribute it.	Pretending the residue is zero, which makes the whole finding look like advocacy.

Four steps. Steps two and three are where the analysis is actually done; the first is bookkeeping and the fourth is honesty.

Step three is the one that decides whether anything happens afterwards. The same observation can be written two ways, and only one of them can be acted on. "She anticipates what the credit desk will ask" is an observation about a person and it terminates there. "The two fields the credit desk most often queries are not flagged anywhere in the application screen" is the same observation written as a property of the work, and it has a cost, an owner and a completion date attached to it before anyone has finished reading the sentence.

> **THE TEST FOR A WELL-WRITTEN FINDING**
>
> Read it aloud and ask: could somebody who has never met either of these two people act on this? If the answer needs a description of a person's character, the finding is not finished. If it names a thing, a place and a change, it is.

Four qualities that look like character *and are conditions of the work*

The same four differences turn up so often that they are worth naming. Each of them is normally written down as a personal quality, and each of them is in fact a statement about the environment.

WHAT YOU OBSERVE	WHAT IT IS USUALLY CALLED	WHAT THE FINDING ACTUALLY IS
They get round the system	Initiative, ownership, being system-smart.	A procedure that obstructs the work. Every workaround is a defect report written by somebody who could not wait for the fix.
They know what others do not	Curiosity, being well-informed	An information channel that does not reach everyone. What they know is not secret; it is simply not distributed.
They have better tools	Resourcefulness	A tool that exists and has not been given out: a license, a template, a report, a permission.
They have a better network	Being good with people.	The official route does not work, so the effective one is private. Somebody has a direct line and nobody published it.

Four observations, four findings, and not one of them is about the person. Reframed this way each becomes an action with a cost and an owner.

The reframing in the third column is not a rhetorical preference. It is what turns the finding into an assignment. “She takes ownership” has no owner, no cost and no completion date; “the query screen sends one question at a time” has all three, and it can be assigned on Friday. An explanation in terms of disposition closes the inquiry, because there is nothing further to establish and nothing to do except admire one person and hope for the others.

Some of the difference does belong to the person: nine years of pattern recognition, an aptitude, a relationship built over a decade. That part is real and it is not distributable. But it is the residue left *after* the subtraction, not the explanation offered before it, and in practice it is smaller than anybody expects.

Every workaround is a defect report **written by somebody who could not wait for the fix.**

From one person
to the average of a unit

Here is where most attempts at this fail, and they fail in the same way: the organization tries to copy the behavior. Behavior is the one thing that cannot be handed over, because it only happens when the conditions permit it.

Three standard moves and why each disappoints. Making the exemplary performer a trainer converts a producer into a teacher, leaves the conditions untouched, and, as section 04 established, asks them to teach the part of the job they can least describe. Writing up a best-practice document produces a description of behavior, which the reader cannot enact for the same reason. Pairing people up transfers whatever the exemplar can articulate, one person at a time, at the cost of the exemplar's own output. All three leave the advantage where it was.

Gilbert's own purpose is the corrective. He did not compute the ratio to celebrate the exemplar; he computed it to size what the organization was losing across everybody else, and the whole of his method points at the middle of the distribution. The question is therefore never "how do we make more people like her?" It is: **what does she have that the other thirty-three do not, and which of those conditions is the organization free to hand out?**

DISTRIBUTE THIS	NOT THIS	WHY IT WORKS
The tool, the template, the license, the report	A description of how they use it.	A tool is shared by definition. Give it out and the whole unit has it on Monday.
The rule they discovered, written into the procedure or the screen	An instruction to be more thorough.	The screen invites the behavior instead of the analyst having to teach it.
The route: who to ring, what to send, in what order	"Build your network"	A published route replaces a private one, and it survives the person leaving.
The information, on a standing schedule	"Stay informed"	What one person went and found becomes what everybody receives.

The left column is shared by construction, which is why it lifts a unit rather than an individual. The right column has to be lifted one person at a time and usually is not. The measured direction agrees: across domains, changing what somebody can reach is among the strongest behavioral levers, and informing alone among the weakest (Albarracín et al., 2024).

Four ways to hand it over, *in order of how well they hold*

Not every condition can be distributed the same way, and the four routes differ sharply in how much they depend on anybody remembering to do something. Prefer the ones that do not.

ROUTE	WHAT IT LOOKS LIKE	HOW WELL IT HOLDS
Build it into the work	The screen changes. The form has the field. The report arrives on Monday whether anyone asks or not.	**Best.** Nobody has to remember, so it survives turnover, holidays and a busy quarter.
Put it where the decision is made	A one-page checklist in the cab, next to the terminal, in the folder that opens with the case.	**Good.** It depends on being visible at the right moment, which is a design problem rather than a motivation problem.
Publish the route	Who to ring, what to send, in what order, on the page people already open.	**Fair.** It works while it is current, and it decays silently. Give it an owner and a review date.
Teach it	A briefing, a session, a walkthrough of the recovered procedure.	**Weakest, and sometimes necessary.** Where the condition genuinely is knowledge, this is the fix, but check the first three routes first.

Four routes, in descending order of durability. The instinct is to start at the bottom because it is the one the organization knows how to buy.

The ordering is not a preference; it is what the evidence about behavior at work keeps producing. Number 03 sets it out in full: what most reliably changes what people do is a change in what they can reach at the moment of the decision, and instruction is the weakest of the available levers. This question is where that argument stops being general and acquires a name, a screen and a phone number.

> SEQUENCE, NOT MENU
>
> Do the first route for everything that can be built in, then the second for what is left, and only then consider the third and fourth. Organizations reliably invert this: they train first because training is procurable, and discover afterwards that the trained people return to a job that still sends one question at a time.

Two more conditions *have to change*

Distributing the conditions is necessary and it is not sufficient. Two further moves turn this from an anecdote into a result, and both of them are about what you measure afterwards.

Reset the desired level to what the exemplary performer actually achieves. Not to perfection: to the number somebody in the building is already producing. This is the most useful thing this question hands to the rest of the analysis, and it is worth being explicit about why.

A desired level set by ambition invites one argument and always the same one: that it cannot be done here, with these customers, on this system. A desired level set by observation cannot be argued that way, because it has already been done here, with these customers, on this system, by a colleague whose name is known. The target stops being a demand and becomes a fact about the organization's own capability. Number 06 takes how desired performance is established in general; this is the version of it that survives contact with a skeptical room.

Two refinements keep it honest. Set it at what the exemplar reliably achieves rather than at their best month, because a best month is not a level. And where the exemplar's advantage turns out to be partly undistributable, set the target at the distributable part only: in the case below, the six days that the consolidated request explains rather than the full nine.

Measure the average of the unit afterwards, not the exemplar. This sounds obvious and is almost never done, because the natural thing to report is the success story rather than the mean of thirty-four people.

The reason it matters is that the two numbers can move independently, and which of them moves tells you what actually happened. If the mean moves and the exemplar does not, a condition was distributed and everybody else caught up, which is the result you were after. If the exemplar improves and the mean does not, you have given attention to one person. If both move, something changed for everybody including her, which is usually a systems change and worth understanding separately.

Behind this sits a quieter point about how performance is distributed at all. O'Boyle and Aguinis examined a large body of performance data and found the familiar bell curve to be a poorer description than usually assumed, with a heavier tail of very high producers than a normal distribution allows (O'Boyle & Aguinis, 2012). For this question the implication is practical rather than statistical: if a small number of people account for a disproportionate share of output, then the mean is dragged by the middle and the middle is exactly where a distributed condition does its work. Reporting the exemplar tells you about the tail. Reporting the mean tells you whether the organization changed.

The lending officer
who sent one email

A regional bank, 34 small-business lending officers, 4,200 applications a year. The figures are illustrative and reconcile from the rates stated with them.

The business need was decision time. The bank promised a credit decision within ten working days and averaged 17.2, and applications that ran past twelve days were dropping out at 22 percent against 9 percent for the rest. The proposal on the table was a two-day course on credit-file quality for all 34 officers.

Before the course was approved, somebody asked the third question. On the bank's own dashboard one officer was running at 8.1 days, had been for six quarters, and nobody had mentioned it: the dashboard reported the branch average, and her branch had two officers.

THE COMPARISON

Her	8.1 days. One consolidated request for missing information per file.
The other 33	17.2 days. An average of 3.1 separate requests per file, each costing about 2.9 days of waiting.
What she did differently	Read the whole file first and sent every question at once. The application screen presents one query field and one send button, so it invites the opposite.
How she found out	By accident, in her first month, when a customer complained about being asked three times. She had never mentioned it because it did not feel like a method.
What the subtraction explains	2.1 avoidable rounds at 2.9 days is **6.1 days** of the 9.1-day difference.
What it does not explain	The remaining **3.0 days**. Nobody established where they went, and the report said so.

One officer, one habit, and two-thirds of a nine-day difference accounted for. The other third stayed open, which is the honest way to report it.

What was distributed, *and what it returned*

Nothing was taught. Three conditions were changed, and none of them was a description of her behavior.

The screen acquired a second button, *hold and consolidate*, which parks a query until the officer has finished reading and then sends the accumulated list as one message. The credit desk published a checklist of the two fields that account for most returns. And the desk's direct number, which she had and the others did not, went on the intranet page next to the shared mailbox.

Three changes, all of them shared by construction. Within two quarters the unit average moved from 17.2 days to 10.4. She herself did not improve at all, which is exactly what should happen: she was already doing it.

WHAT IT RETURNED

Applications past twelve days	From 61 percent to 14 percent of the 4,200 a year.
Drop-outs avoided	At 22 percent past twelve days against 9 percent inside it, **roughly 260 applications a year** that previously walked away.
Worth	At a 74 percent approval rate and $2,900 of first-year contribution a funded loan, **$551,000 (€485k) a year**. The method for that arithmetic is number 13.
What it cost	Two days of development on the screen, one page on the intranet, and a phone number.

The two-day course for 34 officers was never bought. It would have addressed the sixth condition, and the finding was in the first and second.

THE UNCOMFORTABLE PART

She had been doing this for six quarters in an organization that measured her, appraised her and promoted from her branch, and nobody had asked. Not because anyone was careless, but because the dashboard reported branch averages, and an average of two people hides one of them completely. The invisibility was arithmetic, not indifference.

What this *does not claim*

This question is the one most often done badly, and the failures are predictable. Four of them.

- **The comparison is not an experiment.** Nobody randomized anything. The exemplary performer may have an easier territory, a longer tenure, a better manager or a quieter branch. Establish that the groups are comparable before you attribute anything to the difference, or the whole finding is confounded and will be dismantled by the first skeptic in the room.
- **Selecting on a metric selects on what the metric captures.** If the measure is volume you will find whoever is best at volume, including at the expense of everything the measure does not see. Check the exemplar against a second measure before treating them as one.
- **Formalizing a practice can destroy it.** Some exemplary performance depends on discretion: judging when the rule should not apply. Writing that into a procedure removes the judgement and often the benefit with it. Distribute the tool and the information freely; be careful with the rule.
- **Part of the difference really is the person,** and when it is, the finding is about selection and staffing rather than about learning. Saying so plainly is more useful than pretending everything is environmental, and it is the only honest way to keep the claim that most of it is not.

SUMMARY

Somebody already achieves it. What they have is mostly not character: it is a tool, a route, a rule or a piece of information that the organization handed to one person and withheld from the rest. Find them on a measure rather than by nomination, watch rather than ask, subtract rather than admire, and distribute the conditions rather than the behavior. Then measure the average of the unit, because that is the only number that can tell you whether any of it worked.

RUN IT YOURSELF

The comparison, *in six lines*

Twenty minutes, on the problem in front of you. Answer these six and you have question three. Leave one blank and you have found the first thing to establish.

1 · The measure that separates exemplary from standard

Which number, at what value for each? If you cannot state it, you are about to run a popularity contest.

2 · Who has met it for more than two quarters

Names, from the data first and the manager second. And are they comparable: same territory, same customers, same tenure?

3 · Your comparison group

Which standard performers will you observe alongside them? Without these there is no finding.

4 · What you saw, not what you were told

One complete piece of work, watched end to end, for each. Every decision, and every place the work waits.

5 · The subtraction

What does the first have that the second does not? For each item: a tool, a rule, a route, information, or genuinely the person?

6 · The average you will measure afterwards

Which unit, which number, over what period. If only the exemplar improves, nothing was distributed.

WHERE THIS PAPER SITS

Two problems, *one method, sixteen papers*

	BUSINESS PROBLEM	TRAINING REQUEST
Owned by	The manager or director who owns the result.	L&D and the HR business partner.
What it prevents	Spending on the wrong intervention: a system, a reorganization, a hire.	Spending on a program that cannot reach the cause.
What it returns	The fix that moves the number, chosen on evidence.	The budget and the working hours that were not spent.
Method used	The same three analyses and five questions, which is why one capability serves both, and why the diagnosis no longer has to be bought in case by case.	

One method, two returns, and a standing capability instead of a standing consultancy line.

Two models carry the method. Gilbert's six conditions ask what kind of cause; Rummler and Brache's three levels ask where it sits. Both are sound and this series uses both, and it leads with Gilbert, because his unit is the accomplishment, which is what an organization already counts, and because his six conditions are the categories an organization can actually change.

Where the method *comes from*

Two further pieces close the set and carry no number, because each points at all the numbered papers. Before the guide sits Arets L&D Business Models, the destination piece for the people who receive the requests: four models a learning function can run, and the route from HR-driven to business-driven. Last sits Solutions that cannot work, the field guide, ordered by the sentence that lands on your desk: seventeen forms a performance problem arrives in, each with an address in one of the numbered papers. Three orders in all: the destination you choose, the question you are on, and the form that landed on your desk.

HOW THIS SERIES CAME ABOUT

Human performance technology travels the way a craft does: from hand to hand. Vivian Heijnen and Jos Arets learned the analysis from Dennis Mankin, in San Diego, more than twenty years ago, and have kept in contact with him ever since. Both remain grateful to him: for the rigor of the method as he taught it, and for the leadership he has given performance consulting over a long career.

What followed was twenty years of use. Vivian took the method into organizations in the Netherlands and well beyond them, and what accumulates over that many analyses is largely tacit: the cut nobody thought to make, the question that opens a defensive room, the finding that turns out to be an artifact of the data. Making that knowledge explicit, case by case, is what this series is for. Jos wrote it down as it went, in two books and more than forty articles and blog posts, most of them in Dutch, and one seven-part English series, *Is AI the bicycle of the mind?*, which number 15 grows out of. The books: *Liever (g)een training* with Boudewijn Overduin, *Kostbaar misverstand* with Vivian Heijnen, which this series takes a good deal of its thinking from.

None of that is reprinted here. It has been reworked: brought up to date, tested against the research that has appeared since, and dropped where the evidence no longer carries it. Theory that has never met a client is a position; practice that cannot account for itself is a story. This series is what remains when each is held to the other. Performance Consultant AI carries the same discipline to the desk where the decision is actually made.

About the cases. Every worked case in this series comes from our own practice. Names are changed and details altered so that no organization is recognizable, and the arithmetic reconciles. That is said once, here, rather than under each case: client work is confidential, and a case that named anybody would be a breach rather than a proof.

THE SERIES

Sixteen papers, *six parts*

Number 01 is the map and number 02 is the foundation. From number 05 on, every number is depth on exactly one step of the method. The filled block is the paper you are holding; the two unnumbered pieces close the set.

I · WHAT HPT IS

The discipline, where it came from, and why the reflex it replaces survives.

- 01 The performance improvement process
- 02 Human performance technology
- 03 Environment before capability
- 04 The system and the performer

II · THE BUSINESS ANALYSIS

Question 1.

- 05 Business needs and performance needs

III · THE PERFORMANCE ANALYSIS

Questions 2 and 3.

- 06 Gap analysis
- **07 The exemplary performer**

IV · THE CAUSE ANALYSIS

Questions 4 and 5.

- 08 Cause analysis: the three levels
- 09 Cause analysis: the six conditions

V · THE DECISION

From cause to solution, and what counts as proof.

- 10 Intervention selection
- 11 Learning metrics and business metrics

VI · THE FUNCTION

The request, the money, the implementation, the assistant and the value.

- 12 The solution request
- 13 The cost of non-performance
- 14 Implementation and maintenance
- 15 Performance augmentation
- 16 The Value Flip Calculator

Arets L&D Business Models · the destination piece, after 16
No number: the four positions a function can hold, and the move the sixteen papers make simple.

Solutions that cannot work · the field guide, at the end
No number: it points at all sixteen. The same material, ordered by the sentence that lands on your desk.

Business problem: a result is moving the wrong way · Training request: a solution has already been chosen · Both doors: after the decision

REFERENCES

Where every claim *in this paper comes from*

Albarracín, D., Fayaz-Farkhad, B., & Granados Samayoa, J. A. (2024). Determinants of behaviour and their efficacy as targets of behavioural change interventions. *Nature Reviews Psychology, 3*(6), 377–392. https://doi.org/10.1038/s44159-024-00305-0

Brown, O., Power, N., & Gore, J. (2024). Cognitive task analysis: Eliciting expert cognition in context. *Organizational Research Methods, 28*(3), 375–404. https://doi.org/10.1177/10944281241271216

Clark, R. E., Pugh, C. M., Yates, K. A., Inaba, K., Green, D. J., & Sullivan, M. E. (2012). The use of cognitive task analysis to improve instructional descriptions of procedures. *Journal of Surgical Research, 173*(1), e37–e42. https://doi.org/10.1016/j.jss.2011.09.003

Debatin, T., Hopp, M. D. S., Vialle, W., & Ziegler, A. (2021). The meta-analyses of deliberate practice underestimate the effect size because they neglect the core characteristic of individualization — an analysis and empirical evidence. *Current Psychology, 42*(13), 10815–10825. https://doi.org/10.1007/s12144-021-02326-x

Edwards, T. C., Coombs, A. W., Szyszka, B., Logishetty, K., & Cobb, J. P. (2021). Cognitive task analysis-based training in surgery: A meta-analysis. *BJS Open, 5*(6), zrab122. https://doi.org/10.1093/bjsopen/zrab122

Elliott, P. H., & Folsom, A. C. (2013). *Exemplary performance: Driving business results by benchmarking your star performers.* Jossey-Bass. ISBN 978-1-118-20420-7

Gilbert, T. F. (1978). *Human competence: Engineering worthy performance.* McGraw-Hill.

Where every claim *comes from · continued*

Hambrick, D. Z., Macnamara, B. N., & Oswald, F. L. (2020). Is the deliberate practice view defensible? A review of evidence and discussion of issues. *Frontiers in Psychology, 11*, 1134. https://doi.org/10.3389/fpsyg.2020.01134

Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. *Psychological Science, 25*(8), 1608–1618. https://doi.org/10.1177/0956797614535810

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

O'Boyle, E., & Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. *Personnel Psychology, 65*(1), 79–119. https://doi.org/10.1111/j.1744-6570.2011.01239.x

Rummler, G. A., & Brache, A. P. (2013). *Improving performance: How to manage the white space on the organization chart* (3rd ed.). Jossey-Bass. (Original work published 1990)

Swaby, L., Shu, P., Hind, D., & Sutherland, K. (2022). The use of cognitive task analysis in clinical and health services research — a systematic review. *Pilot and Feasibility Studies, 8*(1), 57. https://doi.org/10.1186/s40814-022-01002-6

Tofel-Grehl, C., & Feldon, D. F. (2013). Cognitive task analysis–based training: A meta-analysis of studies. *Journal of Cognitive Engineering and Decision Making, 7*(3), 293–304. https://doi.org/10.1177/1555343412474821

The three books named earlier in this paper are listed in full in number 02, so that they do not recur in every reference list.