



— A GROWTH ENGINEERING HANDBOOK

Neurogogy

Built for how you're wired

Neurogogy is the fusion of neuroscience and pedagogy:

designing training around what the brain actually requires rather than around what feels like it is working.

Those two things come apart more often than the industry admits. The techniques with the strongest evidence behind them are often the ones learners rate worst.

This handbook sets out what the evidence supports, what it means for the way you build, and which numbers we should all stop repeating.

SIX PARTS

130+ SOURCES

2026 EDITION



The six pillars

01	How the brain learns	11
	Encoding, consolidation, retrieval and forgetting. The constraints you cannot design around.	

02	The state the brain learns in	23
	Emotion, motivation, sleep, stress and attention. The conditions that decide whether any of it lands.	

03	What works	39
	Seven techniques with evidence behind them, what each is worth, and why almost nobody uses them.	

04	How learning is structured	54
	Objectives, scaffolding, social learning and transfer. The architecture around the techniques.	

05	Designing brain-friendly learning	69
	Microlearning, multimedia, gamification, story and habit. Five formats, and the mechanism each one serves.	

06	What to avoid	84
	The ideas that survived their own disproof, and how to recognise the next dubious claim.	



A working document, not a literature review

Every claim in it is traceable. Where a figure is contested, we say so. Where a popular statistic turns out to be folklore, we name it and retire it, including the ones our own industry has been repeating for forty years.

Each part synthesises the research and then points you to the full article on [The Lab](#)⁷, our research library, where the studies, the caveats and the counter-evidence are set out at length. Read the part for the argument. Follow the links when you need the detail. Every number carries a reference to the claims register at the back, which gives the full citation, the sample it came from and a link to the source.



The gap

Almost every conversation about training effectiveness reaches the same number sooner or later: only 10% of training transfers to the job.

It is not a research finding. It comes from a 1982 article in *Training and Development Journal* by David Georgenson^[1]. It does not even appear there as his own conclusion. It appears in the opening paragraph as a line spoken by an unnamed training director: "I would estimate that only 10 percent of content which is presented in the classroom is reflected in behavioral change on the job."

No study. No method. No data. An illustrative aside in a trade magazine, which the profession then spent four decades citing as evidence. In 2011, Ford, Yelon and Billington^[2] gave it the name it deserves: the 10% delusion.

THE NUMBER EVERYONE REPEATS

Only 10% of training transfers to the job. Cited for four decades as evidence of systemic failure.

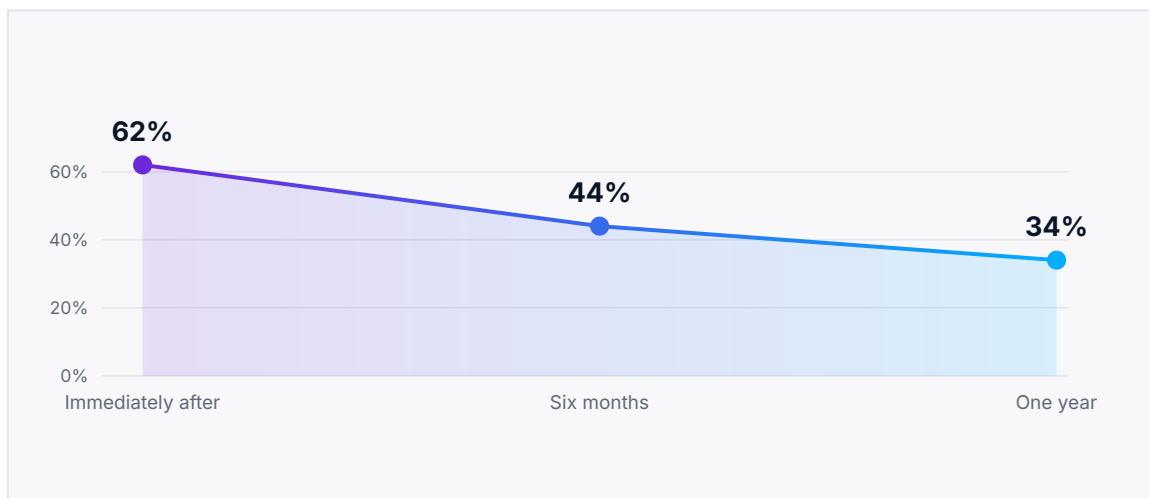
WHAT IT ACTUALLY IS

A line of dialogue attributed to a hypothetical training director in a 1982 trade magazine. No study, no method, no data.

So what does the evidence show? When Saks and Belcourt^[3] surveyed 150 training and development professionals, respondents estimated that 62% of employees apply what they learned immediately after training, 44% are still applying it six months later, and 34% after a year.

Figure 1. Training reaches the job. Then it drains away.

Proportion of employees still applying training, as estimated by 150 training and development professionals · Saks & Belcourt^[3]



These are practitioner estimates rather than observed behaviour, and worth reading that way. They describe something different from catastrophic failure, and more useful.

That distinction matters, because the two diagnoses lead to opposite prescriptions. If training fails at the point of delivery, you fix the content. If training works and then decays, you fix everything that happens after the content: the spacing, the retrieval, the reinforcement, the manager who either creates room to practise or does not. One of those is a courseware problem. The other is a design problem, which is where the research keeps pointing.

This handbook is about closing that gap. Not by adding more content, or better video, or a slicker interface, but by designing against how the brain actually encodes, stores, retrieves and loses information. The science has been sitting in the journals for decades. Bliss and Lømo demonstrated the cellular mechanism of memory strengthening in 1973.^[4] Ebbinghaus charted forgetting in the 1880s. Sweller formalised cognitive load in 1988.^[5] Very little of it has reached the average compliance module.

A second gap is less comfortable to name. A large amount of what our industry calls brain science is not science. Learners have styles that should be matched. Retention follows a neat pyramid of percentages. Ebbinghaus proved 90% of what you learn is forgotten within a week. Only 10% of training ever reaches the job. Every one of those claims is either unsupported or actively contradicted by the evidence, and every one of them still shapes real training budgets.

OUR POSITION

Design for the brain first, and learning will follow. That means accepting constraints you would rather not have, including a working memory that holds about four things, a forgetting process that begins the moment learning ends, and an attention system that no amount of enthusiasm will override.

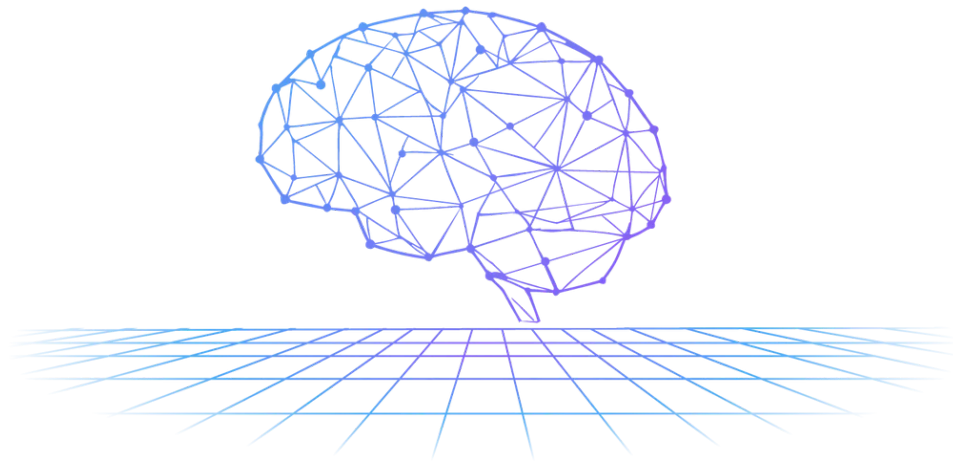
DEFINITION

What neurogogy is

Pedagogy puts the teacher in charge. Andragogy, as Malcolm Knowles^[6] framed it, recognises that adults direct their own learning. Heutagogy goes further, handing learners the choice of what to learn and why.

Neurogogy asks a different question. Not who leads, but what the brain requires. It is the fusion of neuroscience and pedagogy: evidence about how the brain encodes, consolidates and retrieves information, applied directly to how training gets designed, delivered and reinforced.

It does not replace the frameworks above. It underpins them. You can run a beautifully learner-led programme that still ignores cognitive load, still never revisits anything, and still produces nothing durable. Autonomy is a motivational condition. It is not a memory mechanism.



The brain you are designing for. Not a metaphor, and not infinitely adaptable: a physical system with a fixed working capacity, a default rate of decay, and a set of conditions it runs better under.

Neurogogy rests on two commitments



Mechanism over preference

The question is never what learners enjoy, or what they say helps them. It is what demonstrably produces retention and transfer. These frequently point in opposite directions, and Part 3 explains why: the conditions that feel most productive during learning are often the ones that produce the weakest durable memory.



Design as the delivery system

Neuroscience is only worth anything here if it changes what gets built. Every part ends with what the evidence means for the structure of a programme: how long, how often, in what order, reinforced how, measured by what.



The standard we applied

Four rules, applied to every number in this handbook. They are worth stating because they are what makes the difference between a white paper and a brochure with footnotes.



Read the primary source, not the summary

Every figure here was checked against the paper it came from. Not a blog post, not a conference slide, not another vendor's white paper. Five of the sources are books with no online version we can link you to. Each of those entries is marked print only.



State the sample next to the finding

A result from sixteen people and a result from ninety-seven thousand are not the same kind of thing, and hiding the difference in a footnote is a choice. Where a study is small, you will see how small it is at the point the number is used.



Translate the statistics

An effect size of 0.83 means nothing to most readers, and a decimal presented without a scale is a way of sounding precise without being clear. Every effect size in this handbook is given with its plain-language equivalent, and proportions are distinguished from percentage points.



When the number shrinks, use the smaller one

Effects are routinely largest in the laboratory that discovered them and smaller once a field pools its results. Where a single striking result and a pooled one disagree, we have set expectations by the pooled figure, even when the bigger number was the one we would rather have quoted.

How to read the numbers

Six terms do most of the work in this handbook. If they are already familiar, skip the box.

Effect size

One number for how much difference something made, on a scale that lets you compare a maths study with a memory study. Written d or g , and with a bar over it ($\bar{g}$) when it is an average of several. Roughly: 0.2 is small, 0.5 is moderate, 0.8 is large. A related measure, the correlation, is written r or ρ and runs from -1 to 1 .

Where one appears in this handbook, we give the plain-language version beside it.

Standard deviation

The spread of a set of scores. An effect of "one standard deviation" moves the average person from the middle of the group to roughly the top 16%. It is the unit effect sizes are counted in.

Two standard deviations is the famous tutoring claim in Part 4. It does not survive.

Meta-analysis

A study of studies. Instead of running a new experiment, the researchers pool the results of every experiment already run and calculate an average. Pooled figures are almost always smaller than the single study that made a technique famous, and almost always more reliable.

Where the two disagree, we plan against the pooled one.

Confidence interval

The range the true answer is probably in. A narrow interval means the studies agree. A wide one means the direction may be clear but the size is not. An interval that crosses zero means the effect might not exist at all.

We report the interval wherever it changes what you should conclude.

Publication bias

Studies that find something get published. Studies that find nothing often do not. So the published literature overstates almost everything, and statistical corrections exist to estimate by how much.

Correcting for it is what takes growth mindset from a finding to a non-finding.

Preregistered

The researchers published their hypothesis and analysis plan before collecting data, so they could not quietly change the question once they saw the results. It is the strongest form of evidence in this handbook.

Preregistered replications are what killed ego depletion in Part 2. Where you see a *Bayesian analysis*, that is a second way of asking the same question: not just whether an effect failed to appear, but how much more likely the data is if there is no effect at all.

What that cost us

Applying those rules meant retiring numbers we had used ourselves. These are the ones that did not survive.

THE CLAIM	WHAT WE FOUND
Only 10% of training transfers to the job	An unnamed training director's aside in a 1982 trade article ^[1]
One-to-one tutoring is worth two standard deviations	0.37 across 96 randomised trials, once tested at scale ^[7]
People remember 10% of what they read and 90% of what they do	No study anywhere. The citation chain runs out around 1913 ^[8]
It takes 66 days to form a habit	A median from 82 volunteers, with a range of 18 to 254 days ^[9]
Development splits 70:20:10	A 1988 survey asking executives to recall what had

	developed them ^[10]
Ebbinghaus proved 90% of learning is forgotten within a week	His own data shows 21.1% savings still present at 31 days ^[11]
Attention spans are shorter than a goldfish's	No study of either species. Believed by half of UK adults surveyed ^[12]
Growth mindset interventions raise attainment	$d = 0.05$ across 63 studies, and non-significant once corrected for publication bias ^[13]

A note on what this means for our own marketing. Several of the numbers above have appeared in Growth Engineering material. Retiring them costs us some useful copy. We think a handbook that quietly kept them would not be worth reading, and that an industry which checked its own citations more often would be a better one to sell into.

How the Brain Learns

Four constraints you cannot design around, and the folklore that has grown up around each of them.

THE ARGUMENT OF THIS PART

Memory is a process rather than a place. Working memory is far smaller than you think. The brain physically reorganises in response to practice. And forgetting is the default state, not a malfunction.



CONSTRAINT 01 OF 04

Memory is a process, not a place

In September 1953, a 27-year-old man named Henry Molaison underwent bilateral medial temporal lobe resection to control severe epilepsy. Scoville and Milner's 1957 report^[14] described what the surgery cost him.

Surgeons removed the front two-thirds of the hippocampus on both sides of H.M.'s brain to stop his seizures. He and nine other patients who had had similar operations were then given standard intelligence and memory tests.

INTELLIGENCE

IQ 112, up from 104

OLDER MEMORIES

Largely undisturbed

HOLDING IN MIND

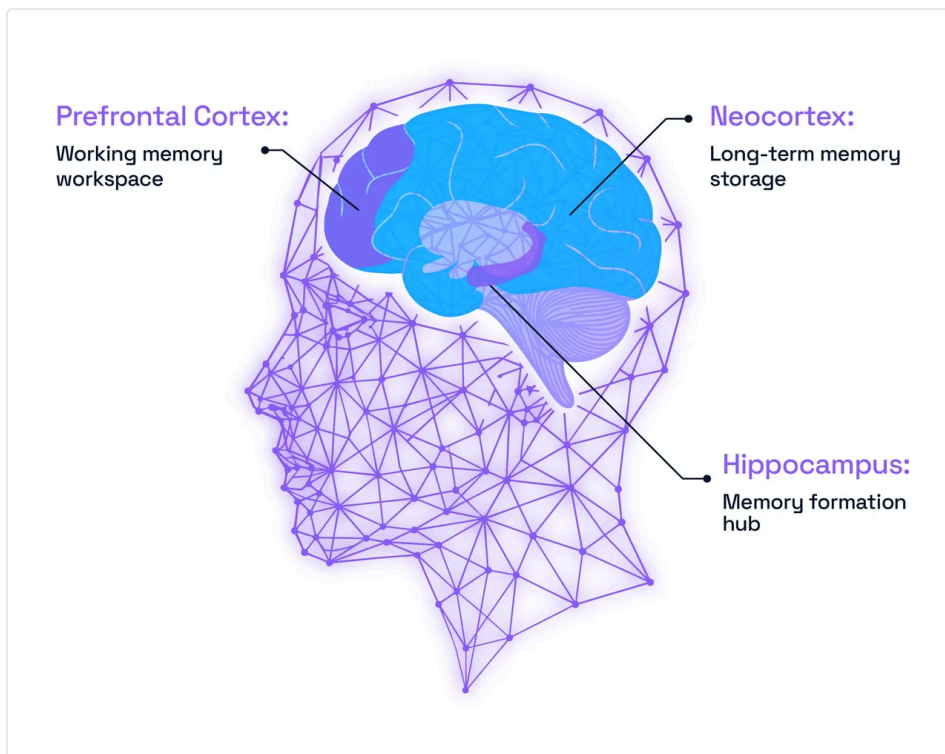
Minutes, if not interrupted

NEW MEMORIES

None, for the rest of his life

- Intelligence intact, memory gone. Three of the ten patients tested showed the severe defect, and its severity tracked how far back the surgery had gone, which is what ties the loss to the hippocampus rather than to surgery in general. The gap between those four rows is the single most instructive fact in learning science: memory is neither one system nor stored in one place.

Molaison's case established the architecture. Holding something in mind for a moment, storing it for years, and retrieving it later are separate operations, carried out by different structures, and they fail independently. His capacity to acquire new motor skills, demonstrated in later testing, showed that even long-term memory splits into systems that can survive without one another.



Three structures, three jobs. The prefrontal cortex holds what you are working on now. The hippocampus binds it into something durable. The neocortex is where it ends up if consolidation succeeds. Damage to the middle one is what removed Molaison's ability to form new memories while leaving the other two intact.

The mechanism beneath all of this is physical. In 1973, Bliss and Lømo^[4] stimulated the perforant path in anaesthetised rabbits and found that brief bursts of high-frequency stimulation strengthened the response of dentate granule cells for anywhere from 30 minutes to 10 hours. Long-term potentiation, as it became known, is the closest thing we have to memory made visible: repeated activation of a pathway makes that pathway easier to activate again.

🕒 DESIGN CONSEQUENCE

A single exposure builds almost nothing worth keeping. Encoding, consolidation and retrieval are three jobs, and a one-off training event only attempts the first.

[READ ON THE LAB](#)

The Neuroscience of Memory





The bottleneck

George Miller's 1956 paper^[15] is the most cited and most misread work in the field. The title gave us "the magical number seven, plus or minus two", and the phrase entered folklore. Read the paper and you find Miller openly sceptical of his own number, describing the recurrence of seven as "only a pernicious, Pythagorean coincidence". His substantive point was subtler and more useful: the limit applies to chunks, not to information, so recoding material into larger meaningful units increases how much you can hold.

WHAT THE INDUSTRY DESIGNS FOR

7 ± 2

Miller's headline number, which Miller himself called a "pernicious, Pythagorean coincidence".

WHAT THE EVIDENCE SUPPORTS

4 (3–5)

Cowan's central capacity limit^[16], once rehearsal, chunking and long-term support are stripped out.

That constraint is what cognitive load theory is built on. John Sweller's insight, formalised from 1988, is that instruction which exceeds working memory capacity does not merely slow learning down. It prevents it.

Two kinds of load are worth separating, because you deal with them in opposite ways.

● Intrinsic load

The difficulty built into the material itself: how many pieces have to be held in mind at once to make sense of it. Pricing a single product is light. Pricing one where discount, region and contract length all interact is heavy. You can sequence it, chunk it and build the prerequisites first. You cannot wish it away.

● **Extraneous load**

The difficulty the design adds on top, which does nothing for learning: the split-screen diagram that makes learners hold one element while hunting for the other, the narration reading the on-screen text aloud, the decorative animation. Unlike intrinsic load, all of this is yours to remove, and removing it is usually the cheapest improvement available.

One correction is due here, and it applies to a great deal of published L&D writing including our own.

○ **THE THREE-PART MODEL**

Intrinsic, extraneous and germane load, with germane presented as a third dial you can turn up to increase learning.

◎ **SWELLER'S CURRENT POSITION**

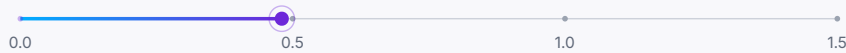
Two things to manage. Germane load "redistributes working memory resources from extraneous activities to activities directly relevant to learning", with "a redistributive function... rather than imposing a load in its own right".^[17]

The most direct way to cut extraneous load is to remove the search. Set a novice a problem and most of their working memory goes on hunting for a solution path while trying to hold the problem itself in mind. Give them a worked solution to study instead and that hunt disappears, leaving capacity for the method. It is the cleanest test of the theory available, because the content is identical and only the load changes.

Barbieri and colleagues^[18] pooled 55 studies across 43 articles and 181 effect sizes, and found worked examples improved mathematics performance at a medium average effect.

WORKED EXAMPLES, MATHEMATICS PERFORMANCE

g =
0.48



Medium. The average learner given worked examples finished ahead of roughly 68% of those who were not.

Ginns' meta-analysis of 50 studies^[19] found substantial gains from integrating spatially or temporally separated information, with the effect concentrated in complex material and close to nothing for simple material.

DESIGN CONSEQUENCE

Cutting extraneous load works, and it works most where the content is hardest. Which is precisely where most training is thinnest.

READ ON THE LAB

What Is Cognitive Load Theory?



CONSTRAINT 03 OF 04

The brain reorganises

WHAT IT IS

Neuroplasticity is the brain physically changing in response to what you repeatedly do. Connections that fire together strengthen and grow new points of contact, the fibres carrying them get better insulated so signals travel faster, and connections that go unused are pruned away. It is the mechanism that makes any training work at all.

The machinery does not switch off with age. What changes with age is what it takes to engage it, and how much of the result you should expect to be able to see.

What training actually produces in adults

The best-powered test of this is not a brain scan. The ACTIVE trial randomised 2,832 adults aged 65 to 94 to training in memory, in reasoning or in processing speed, or to nothing at all. Each programme ran for ten sessions. Everyone was then followed for a decade. ^{[20][21]}

● ACTIVE trial 2002–2014 · Randomised, 10-year follow-up [21]		
Adults in their sixties, seventies and eighties were randomly assigned to one of three training programmes or to no training. Each programme ran for ten sessions across five to six weeks. Everyone was retested at intervals for the next ten years, on the ability they had trained and on several they had not.		
WHO 2,832 adults, aged 65 to 94	HOW MUCH TRAINING 10 sessions, 5–6 weeks	STILL AHEAD AFTER 10 YEARS Speed and reasoning, not memory
● At ten years the trained groups were ahead of the untrained control by 0.66 for speed and 0.23 for reasoning, both statistically significant. Memory training came in at 0.06 and did not. To translate the largest of those: a decade after ten sessions, the average speed-trained person was still ahead of roughly three quarters of the people who had no training at all.		

Ten sessions, and two of the three trained abilities were still measurably better ten years later. That is a stronger durability result than most corporate programmes could claim at ten weeks.

The same trial sets the limit. The gains stayed inside the ability that was trained. Speed training made people faster and left their reasoning where it was, and no arm improved performance on observed everyday tasks at five years or at ten. ^[22] Adults learn well. They do not generalise for free.

What triggers the change

Lövdén and colleagues separate two things the industry runs together. ^[23] *Flexibility* is the brain performing better within the capacity it already has. *Plasticity* is the harder change, where the structure itself adapts to do something it previously could not. Plasticity is triggered by a mismatch between what the brain can currently supply and what the situation

demands. A task comfortably inside current capacity creates no adaptive pressure. One far outside it creates none either. And the mismatch has to last, because plasticity is, in their word, sluggish.

That is a design brief rather than a caveat. Find the gap, and hold people in it.

Why the brain scans are less impressive than the behaviour

Zatorre, Fields and Johansen-Berg reviewed what a scanner is picking up when a trained brain looks different from an untrained one.^[24] Synapses are added and pruned, support cells multiply, axons gain insulation that makes established pathways faster, and blood supply increases where the work is.

Two timings are worth carrying into a plan. Grey-matter change has been detected after as little as seven days of training. White-matter change develops across roughly six weeks.

The pictures are real. They are also small. Trainee London taxi drivers who passed the Knowledge showed hippocampal growth across three years, where those who failed and a control group showed none. The same drivers ended up measurably worse than controls at recalling a complex figure, so capacity was reallocated rather than created.^[25]

Across 33 imaging studies the typical structural effect runs to 2 to 5%. Regions often expand early in learning and partly shrink back while performance carries on improving, and the change fades once practice stops.^{[26][27]}

And the largest test of the lot found nothing at all. Judd and Kievit used the 1972 rise in the UK school leaving age as a natural experiment, comparing brain scans of roughly 5,100 people who were made to stay on an extra year against those who were not.^[28] Across 117 structural measures the answer was no detectable difference, and the analysis was registered before anyone looked at the data.

That is not the deflating result it first appears. A year of school plainly teaches people things. What it does not do is leave a mark large enough for a brain scan to find. The finding is about the resolution of the instrument, not about whether adults learn, which is exactly why the behavioural evidence above carries more weight than any scan.

One claim here has changed in our favour. Adult humans do grow new neurons in the hippocampus. Two independent studies reading the genetic activity of individual cells have now found the progenitors, after a decade of real dispute.^{[29][30]} The rates are low, they vary enormously between people, and none of this is what the training scans are showing, so it rescues none of the marketing built on it.

The outcome of a training programme is not a state you reach. It is a state you maintain.

 DESIGN CONSEQUENCE

Adults keep the machinery, so three things follow. Set the difficulty at the gap between what somebody can do now and what the job demands. Keep them in that gap long enough for it to register, because a single hard afternoon changes nothing. And train the exact capability you want back, because the gains do not spread.

READ ON THE LAB

Brain Plasticity



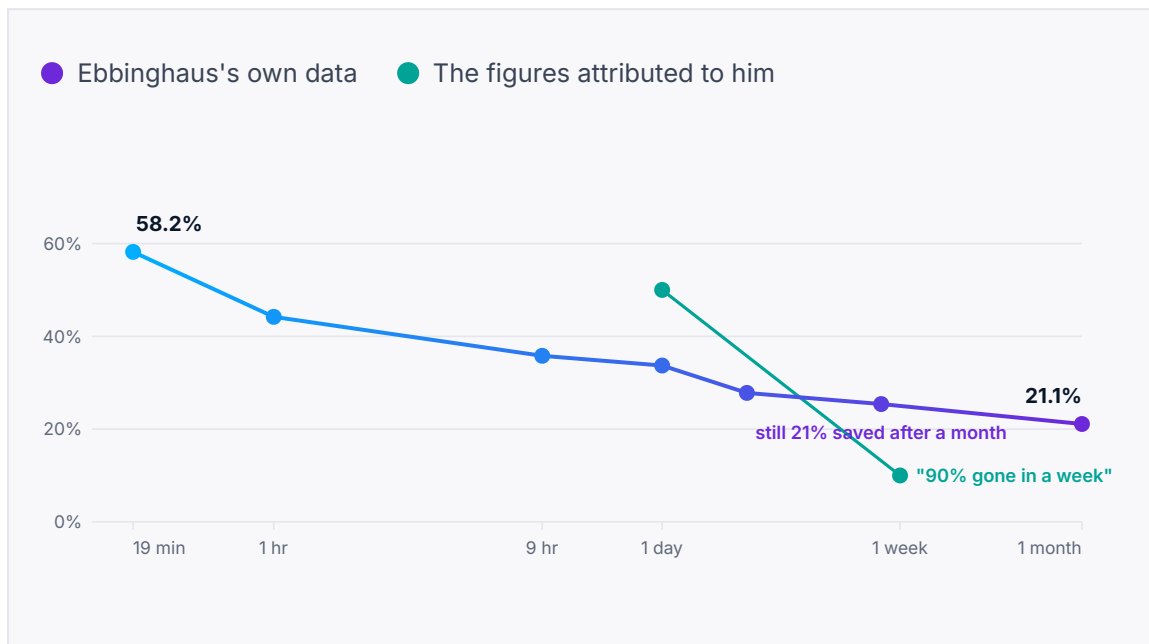
CONSTRAINT 04 OF 04

Forgetting, and the curve you were sold

Hermann Ebbinghaus spent the 1880s learning lists of thirteen nonsense syllables and testing how much effort he saved when relearning them later. His data, reproduced by Murre and Dros in their 2015 replication^[11], look nothing like the version our industry repeats.

Figure 4. The curve Ebbinghaus actually measured, against the one you have been shown

Savings on relearning, plotted on a logarithmic time axis · data as reproduced by Murre & Dros, 2015



His one-day figure is 33.7%, so the loss at a day is nearer two thirds than a half. And forgetting does not continue to 90% within a week: it flattens, and is still at 21% savings after a month. The popular version runs in the wrong direction on both counts.

Three caveats belong on every use of this curve

- **One subject**

The subject was Ebbinghaus himself. Murre and Dros replicated the shape with a single subject over 70 hours of testing.

- **Nonsense syllables**

The material was chosen specifically to strip out meaning. Job-relevant material decays far more slowly, because it connects to what learners already know.

- **Savings, not recall**

Savings measures relearning efficiency, not the proportion of items a learner could produce. The two are routinely confused.

So what survives? The shape survives, and the shape is enough. Retention falls steeply at first, then flattens. Each reinforcement resets the slope.

- **DESIGN CONSEQUENCE**

Design against the shape of forgetting and stop quoting its numbers. Reinforcement schedules are the intervention. The percentages are folklore.

[READ ON THE LAB](#)

[What Is The Forgetting Curve?](#)



PART 1

What this changes

Four design consequences follow, and none of them are about content quality.

- **Build for four chunks, not seven**

Sequence and chunk intrinsic complexity, and strip extraneous load ruthlessly, starting with the hardest material rather than the easiest.

- **Treat one exposure as encoding only**

Consolidation and retrieval are separate jobs that require separate events at separate times.

- **Pitch at the gap, then maintain it**

No gap between current ability and real demand, no structural change. And what practice builds, the absence of practice takes back.

- **Design against the shape of forgetting**

Reinforcement schedules are the intervention. The percentages attributed to Ebbinghaus are folklore. They are not his.

The State the Brain Learns In

Emotion, motivation, sleep, stress and attention. Five conditions that decide what a programme achieves, and most of them sit outside the course.

THE ARGUMENT OF THIS PART

A learner's capacity to encode, consolidate and retrieve is not fixed. It moves with how they feel, how much they slept, how threatened they are and what else is competing for their attention. None of that is soft. All of it is measurable, and most of it sits outside the course.

Part 1 dealt with the machinery. This part deals with its operating conditions. It is also the part of the field with the most folklore in it. Several of the statistics L&D uses to argue for exactly the right conclusions turn out to have no research behind them. We have named those where they arise. A fabricated number is the first thing an informed reader goes after, and it takes the good argument down with it.

Emotion is the filing system

The brain does not store what is important. It stores what is tagged, and emotion does the tagging.

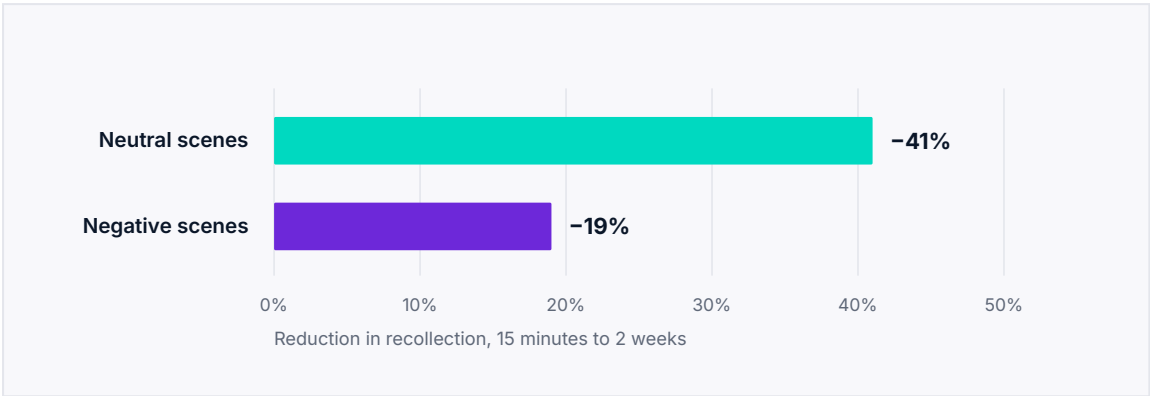
The mechanism is modulation rather than storage. Emotional arousal triggers adrenal stress hormones, which drive noradrenergic activation in the amygdala, which in turn strengthens consolidation in the hippocampus and cortex. Cahill, Prins, Weber and McGaugh demonstrated this in humans in 1994 by blocking it. ^[31]

Cahill, Prins, Weber & McGaugh 1994 · Nature [31]		
Volunteers took either propranolol, a drug that blocks the adrenaline system, or a placebo, then watched a slide story that was either emotionally upsetting or matched but neutral. Memory for the slides was tested afterwards.		
INTERVENTION	EMOTIONAL STORY	NEUTRAL STORY
Propranolol before the story	Memory advantage abolished	Unaffected
Block the beta-adrenergic system and the emotional memory advantage disappears, while ordinary memory carries on. The amygdala does not hold the memory. It decides how hard the rest of the brain works to keep it.		

The effect shows up as a difference in decay rather than a difference in learning. Anderson and colleagues tested recollection across delays running from fifteen minutes to two weeks. ^[32]

Figure 5. What two weeks does to a memory

Reduction in recollection between 15 minutes and 2 weeks, negative versus neutral scenes



Anderson, Yamaguchi, Grabski & Lacka (2006), Experiment 1: 48 students, 16 in each of three delay conditions (15 minutes, 1 week, 2 weeks). The recollection advantage for negative scenes was significant only at the longest delay, which is the point: this is a difference in what survives, not in what goes in. ^[32]

Two further details are worth carrying. The effect appeared for recollection and not for vaguer feelings of familiarity. And it did not appear for fearful faces, so emotional intensity on its own is not the lever.

Proximity matters too. Sharot, Martorella, Delgado and Phelps scanned New Yorkers three years after 9/11 and split them by how vivid their recall was. ^[33] The split turned out to track distance: the vivid group had been an average of 2.1 miles from the towers that morning, the rest 4.5 miles.

● Sharot, Martorella, Delgado & Phelps 2007 · PNAS [33]		
Three years after 9/11, 24 people who had been in Manhattan that day recalled 60 personal memories inside an MRI scanner. Twenty-two of them fell clearly into one of the two groups below. Each cue asked either for a memory of 11 September or for one from the summer before it, which is the comparison.		
VIVID RECALL GROUP 83% of 12, mean 2.1 miles away	LESS VIVID GROUP 40% of 10, mean 4.5 miles away	THE EFFECT ITSELF More left-amygdala activity for 9/11 than for that summer
● The percentages are the proportion of people in each group showing the pattern, not the size of the response. The difference between the groups was significant (p = .025). Same event, same city, same day: personal stakes rather than newsworthiness engaged the mechanism.		

Now the uncomfortable part

The curve the field reaches for to describe this, the inverted U of arousal against performance, is not what its source says.

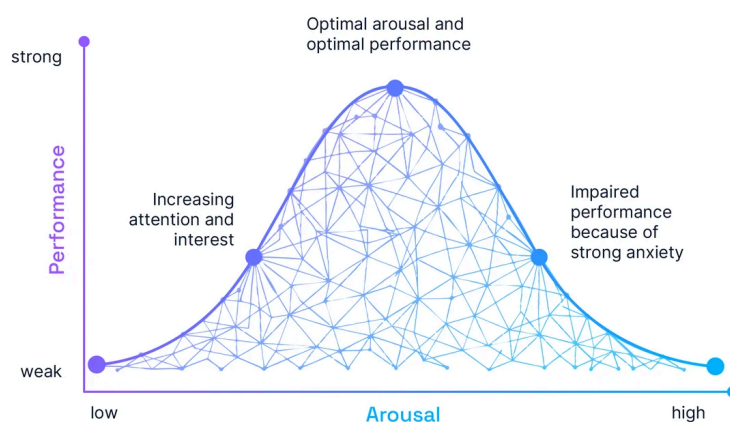
🕒 THE VERSION IN CIRCULATION

The Yerkes-Dodson curve: performance rises with arousal to an optimum, then falls. Therefore a bit of pressure improves learning.

🕒 WHAT YERKES AND DODSON DID

Forty dancing mice learning a white-black discrimination under electric shock. ^[34] Strong stimulation sped up easy learning and impaired difficult learning. No arousal, no humans, and no curve.

Yerkes-Dodson Law Bell Curve



The diagram, as it is drawn in the literature and in our own back catalogue. It is a clean, memorable shape. It is not what Yerkes and Dodson measured. There is no arousal axis in their paper, no human participants, and no curve of this kind. It was drawn decades later by other people, and it has been redrawn ever since.

Practitioners should not seek to increase performance through the manipulation of employee stress levels.

CORBETT (2015), REVIEWING THE LAW'S USE IN MANAGEMENT ^[35]

There is a real inverted U in this territory. It is chemical rather than motivational. Salehi, Cordero and Sandi noted in 2010 that despite universal belief in the curve, nobody had demonstrated it under constant experimental conditions, so they did. ^[36] Rats trained in a water maze at a moderate stress level made fewer errors than rats trained under either milder or harsher conditions, tracking their corticosterone.

Which is a statement about a stress hormone in a water maze, not about deadlines in an office. None of this makes emotion the problem. Emotion is the tagging mechanism this whole section is built on, and worth designing for. What the curve does not support is the separate idea that adding pressure will sharpen performance.

DESIGN CONSEQUENCE

Emotion belongs in learning because it determines what gets kept. It does not belong there as pressure. The correction makes the case sharper, not weaker.

READ ON THE LAB

Emotional Engagement in Learning



CONDITION 02 OF 05

Motivation, and the thing nobody in L&D says out loud

Deci's 1971 experiments produced the finding that launched fifty years of argument. ^[37] Participants paid per puzzle solved spent less of a later free-choice period on the puzzles than they had before. Verbal praise produced no such decline. The individual studies were small and the original results were not decisive, but the pattern held up under synthesis.

A meta-analysis pooling 128 experiments in which people were paid, or not paid, to do something they already found interesting. The measure is what they did once the reward stopped and nobody was watching.

STUDIES POOLED

128

ALL TANGIBLE
REWARDS

$d = -0.34$

POSITIVE FEEDBACK

$d = +0.33$

PAID FOR ENGAGING

$d = -0.40$

PAID FOR
COMPLETING

$d = -0.36$

PAID FOR
PERFORMING WELL

$d = -0.28$

- Outcome measured is free-choice persistence: whether people kept doing the task once nobody was paying or watching. The closer the reward is tied to simply doing the thing, the worse the effect.

Which is where most L&D writing stops, because it supports a comfortable conclusion: intrinsic good, extrinsic bad, build for meaning.

The larger and more recent evidence does not support that conclusion. Cerasoli, Nicklin and Ford pooled 183 samples and more than 212,000 people, and found the two work on different things. [39]

Figure 7. Incentives buy volume. Meaning buys care.

Standardised weights from meta-analytic regression, both predictors entered together · higher means the motivation explains more



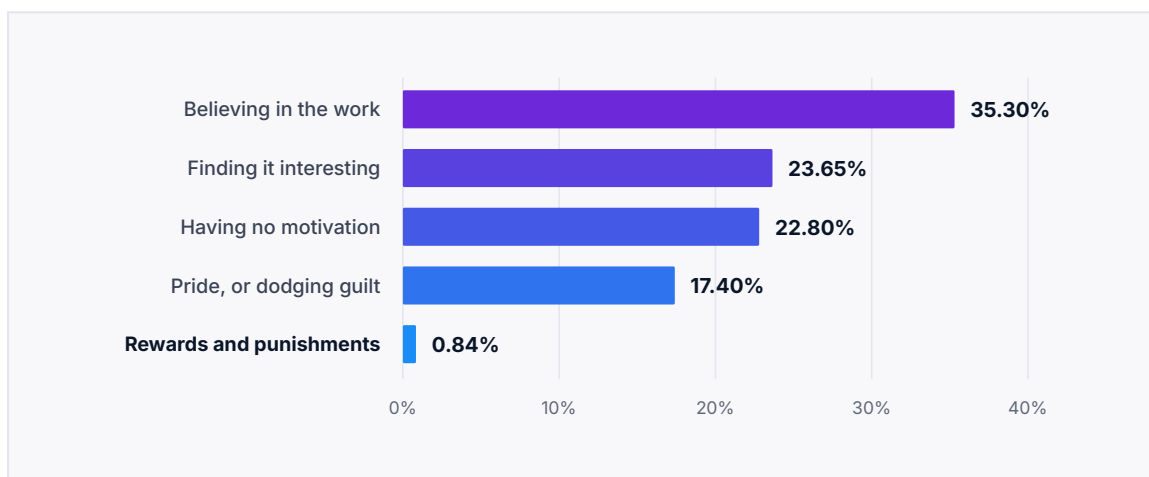
Cerasoli, Nicklin & Ford (2014), Table 4. The paper argues explicitly against the intrinsic-beats-extrinsic framing: the two predict different things. [39]

So the honest position is not that rewards corrode motivation. It is that tangible, expected, controlling rewards for work people already find interesting reduce their willingness to keep doing it unpaid, and that incentives buy volume while meaning buys care. If you are driving completion of mandatory compliance content, incentives are the right instrument. If you are trying to change how someone exercises judgement, they are the wrong one, and no amount of them will substitute.

The size of that gap is worth stating plainly. Van den Broeck and colleagues pooled 124 workplace samples and asked how the five types of motivation divide up the variance they jointly explain in job performance.^[40] Together the five account for a quarter of it.

Figure 8. Five motivations, ranked by what they explain about performance

Share of the variance the five together explain in job performance ($R^2 = .25$), 124 samples



Van den Broeck, Howard, Van Vaerenbergh, Leroy & Gagné (2021), rescaled relative weights for performance. These are shares of explained variance, not of all variance in performance.^[40]

Read that ranking twice. For performance specifically, believing in the work beats enjoying it, and being paid for it explains almost nothing. The same pattern holds in education. Pooling 344 samples and 223,209 learners, Howard and colleagues found intrinsic motivation related to student success and wellbeing, and personal value particularly strongly related to persistence, while motivation driven by a desire to obtain rewards or avoid punishment was associated with neither performance nor persistence, and was associated with decreased wellbeing.^[41]

Underneath all of this sits a mechanism that is routinely described backwards. Dopamine is not a pleasure signal. It tracks prediction error, the gap between what was expected and what arrived, and it drives the

willingness to expend effort. Wanting and liking are separate systems. That distinction is why a reward that arrives predictably stops motivating while an unexpected one still does, and why the feeling of progress does more work than the prize at the end.

Boredom is weaker than it feels

The engagement industry treats boredom as the enemy. Boredom is real and consistent. It is also far smaller than the industry assumes.

🕒 Shui & Zhu 2026 · Three-level meta-analysis [42]	
A meta-analysis of 147 independent studies relating how bored students said they were to how well they went on to perform academically.	
STUDIES 147 independent	STUDENTS 131,446
CORRELATION $r = -.190$	VARIANCE (R^2) ~3.6%
• The variance figure is r^2 , our own arithmetic on the reported correlation. Real, consistent, and not the catastrophe the engagement industry describes.	

Qi and colleagues, pooling 21 studies and 240 effect sizes, found negative emotions associated with worse online learning performance at $r = -.303$, and positive emotions with better performance at $r = .478$.^[43] The positive side of the ledger is the larger one, which argues for building interest rather than merely removing tedium.

🕒 DESIGN CONSEQUENCE

Match the instrument to the outcome. Incentives move completion. Meaning moves judgement. And the bigger prize is on the positive side: building interest beats stripping out tedium.

READ ON THE LAB

Intrinsic Motivation vs Extrinsic Motivation



Dopamine and Learning



Sleep is part of the programme, whether you plan for it or not

Consolidation is not something that happens after learning. It is part of learning, and most of it happens while the learner is unconscious.

Diekelmann and Born's review sets out the mechanism.^[44] During slow-wave sleep, slow oscillations, spindles and ripples coordinate the reactivation and redistribution of hippocampus-dependent memories towards the neocortex. REM sleep supports their synaptic stabilisation. Sleep does not protect memory. It processes it.

Three findings make this operational.

● **Yoo, Hu, Gujar, Jolesz & Walker** 2007 · Nature Neuroscience ^[45]

Twenty-eight adults viewed 150 photographs inside a scanner. Half had been kept awake for roughly 35 hours beforehand. Both groups then slept normally for two nights before the recognition test, so this is not simply tiredness at the test.

DESIGN

14 sleep-deprived · 14 rested

DEPRIVATION

One night, before learning

RECOGNITION, 2 DAYS LATER

19% worse

- Sleep before learning determines how much gets in. Measured as a discrimination index rather than as items recalled.

Sixty-six adults practised a number puzzle that had an undisclosed shortcut buried in it, then had eight hours of sleep, eight hours awake overnight, or eight hours awake in the day, before being retested.

TASK

**Hidden shortcut
in a number
sequence**

AFTER SLEEP

**Insight more than
twice as likely**

TIME OF DAY

Ruled out

- Sleep did not enhance insight in people who had not done the initial training. This is restructuring of something already encoded, not a general boost.

Adults were held to four, six or eight hours in bed a night for two straight weeks under laboratory conditions, with attention and reaction time measured daily and compared against total sleep deprivation.

DESIGN

4, 6 or 8 hours in bed

DURATION

14 consecutive nights

SIX HOURS OR LESS

**Equivalent to up to 2 nights
awake**

SLEEPINESS

Did not separate 6h from 4h

- The authors' own conclusion is the one that matters for programme design: subjects were largely unaware of their own increasing deficits.

🕒 **DESIGN CONSEQUENCE**

Put sleep between the input and the point where it has to be used. A workshop that teaches in the morning and applies in the afternoon gives the material no night at all. Learners will still rate it well. They are rating it at the moment their fluency peaks, which is the moment before the forgetting starts.

[READ ON THE LAB](#)

The Neuroscience of Sleep and Learning



Stress: the real mechanism, not the wellbeing-poster version

The scale of this is not in dispute. The International Labour Organization's April 2026 global report puts the losses from psychosocial risks at work at 1.37% of global GDP each year. ^[48] What is in dispute is what to do about it, and most of the received wisdom points at the wrong lever.

Start with what holds

🕒 de Quervain, Roozendaal, Nitsch, McGaugh & Hock

2000 · Nature Neuroscience

^[49]

Thirty-six adults learned a list of 60 words, then took a single 25 mg dose of cortisone or a placebo, timed either before learning, after learning or an hour before the recall test 24 hours later.

INTERVENTION

Cortisone at stress-level dose

EFFECT

Retrieval impaired

MATERIAL

Previously learned word list

- Acute cortisol elevation impairs getting things out, not putting them in.

🕒 Oei et al. 2007 · Brain Imaging and Behavior

^[50]

Twenty-one young men took hydrocortisone or a placebo and were scanned while retrieving material they had learned earlier, so the cost could be located as well as measured.

PARTICIPANTS

21 young men

DOSE

20 mg hydrocortisone

ON RETRIEVAL

Hippocampus and prefrontal cortex both down

- The same finding from the other direction. Imaging shows where the cost lands.

Sustained exposure does more. Newcomer and colleagues gave 51 healthy adults placebo or one of two cortisol doses for four days, the higher dose approximating the exposure seen during major stress.^[51]

🕒 Newcomer et al. 1999 · Archives of General Psychiatry [51]	
Healthy adults took cortisol at one of two doses, or a placebo, for four days, with memory tested during dosing and again after the drug had cleared.	
SAMPLE 51 healthy adults	HIGHER DOSE Paragraph recall down
UNTOUCHED Non-verbal memory, attention, executive function	AFTER WASHOUT Gone
● The word the authors use is <i>reversible</i> . This is a state, not damage, and states can be designed for.	

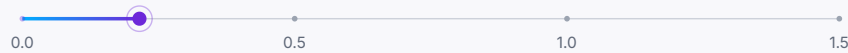
Then the correction

It matters because it changes what you can do about it.

🕒 THE WELLBEING-POSTER VERSION
Stress raises cortisol, cortisol harms memory, therefore lower cortisol and memory improves.
🕒 WHAT THE META-ANALYSIS FOUND
Across 113 studies and 6,216 participants, stress reliably raised cortisol, but the magnitude of the cortisol response was not related to the effect of stress on memory. ^[52]

STRESS JUST BEFORE OR DURING RETRIEVAL, 102
EFFECTS

g =
-0.215



Small, and consistent: the confidence interval runs from -0.346 to -0.085 , and the studies agree closely with each other. Stress at retrieval reliably costs you something, but not much.

"Reduce cortisol" is therefore not the lever the wellbeing literature implies. Something about the stressed state impairs retrieval, and cortisol concentration is not a reliable index of it.

What is actionable is the trigger list

Sonia Lupien's Centre for Studies on Human Stress sets out the four ingredients that reliably provoke a stress response.^[53] The acronym, N.U.T.S., is the centre's. The components come from Mason's work and from Dickerson and Kemeny's meta-analysis of social-evaluative threat.^[54]

N



Novelty

Something the learner has not encountered before. An unfamiliar interface counts.

U



Unpredictability

No way of knowing what is coming. Unclear expectations, unannounced assessment.

T



Threat to the ego

Competence called into question. Assessment that can embarrass someone in front of peers.

S



Sense of control

Little or none over the situation. No choice over pace, order or route.

DESIGN CONSEQUENCE

Every one of the four is something a learning designer controls directly. That is a four-item audit you can run before you add anything, worth more than any resilience module.

READ ON THE LAB

Taming Cortisol: The Neuroscience of Stress-Free Learning Design



CONDITION 05 OF 05

Attention: a real constraint, wrapped in bad statistics

Start with what holds. Three findings are robust, and enough on their own to justify redesigning most corporate training.

- **Working memory runs to about four chunks.** Established in Part 1, and the ceiling every other constraint sits under.
- **Switching between tasks costs time,** and the cost grows with rule complexity.
- **Interruption is expensive to recover from.** The recovery, not the interruption, is where the time goes.

Now the clear-out, because this topic carries more folklore than any other in L&D. Every row below is a number in current circulation, and none of them should be.

Table 1. Four attention statistics to retire

THE CLAIM	WHAT IT ACTUALLY IS	STATUS
-----------	---------------------	--------

THE CLAIM	WHAT IT ACTUALLY IS	STATUS
Willpower is a fuel tank	Ego depletion failed a 23-laboratory preregistered replication of 2,141 people at $d = 0.04$, with a confidence interval spanning zero. ^[55] A second, larger multi-site test across 36 labs and 3,531 people returned $d = 0.06$, and its Bayesian analysis found the data four times more likely under the null. ^[56]	Failed replication
Task switching costs 40% of productive time	A remark attributed to David Meyer on an APA explainer page, which reports that he "has said" switching can cost as much as 40%. No study cited. The 2001 paper it is attached to measured switching costs in fractions of a second. ^[57]	Not a finding
Media multitaskers have worse cognitive control	Two powered replications found a significant effect in 5 of 14 tests, only two surviving a conservative Bayesian analysis. The accompanying meta-analysis turned non-significant once publication bias was corrected. ^[58]	Largely unreplicated
2.5% of people are "supertaskers"	Five people out of 200 tested in a driving simulator, identified by a threshold on difference scores. Never independently replicated. ^[59]	Single study

🎯 DESIGN CONSEQUENCE

None of that weakens the design case. It strengthens it, because the case never needed the numbers. Attention is limited, protecting it is a design decision, and the honest version is more defensible than the dramatic one.

[READ ON THE LAB](#)

[The Neuroscience of Focus](#)



[The Multitasking Myth](#)



PART 2

What this changes

Four consequences, none of which are about content.

- **Design for the state, not just the session**

Sleep, stress and attention account for a large share of what a programme achieves, and all three sit outside the course. If you control none of them, you are optimising the smaller half of the problem.

- **Use emotion to tag, not to pressure**

Relevance, stakes and narrative earn their place because they determine what gets consolidated. Pressure does not, and the evidence usually cited to justify it is a study of mice.

- **Match the motivator to the outcome**

Incentives buy volume. Meaning buys care. Of everything the five motivation types explain about job performance, rewards and punishments account for under 1%, so if judgement is what you need, the reward budget is not where to find it.

- **Audit for the four stress triggers before you add anything**

Novelty, unpredictability, threat to ego, low control. Removing those is cheaper than anything you could add, and the one thing here you can act on this week.

What Works

Seven techniques with evidence behind them, what each one is actually worth, and the reason almost nobody uses them.

THE ARGUMENT OF THIS PART

The conditions that make learning feel productive and the conditions that make it durable are frequently opposites. Every technique below costs the learner something in the moment and pays them back later. Every technique that feels smooth costs them later instead. Robert Bjork called that category **desirable difficulties**. It is the thread running through all seven.

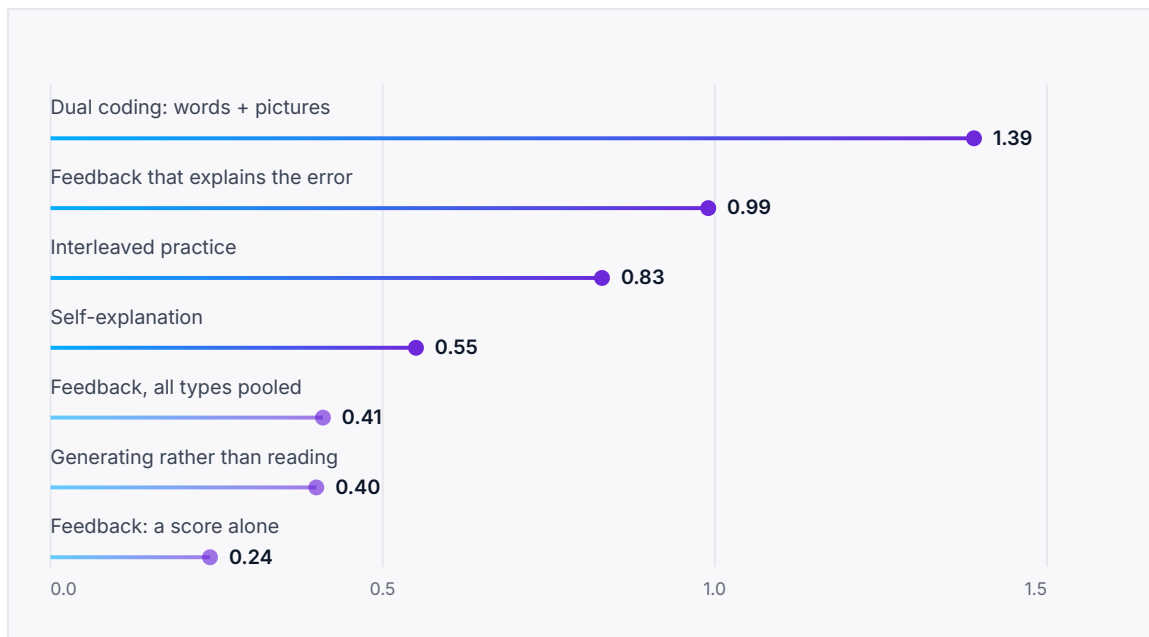
Parts 1 and 2 dealt with the machinery and its operating conditions. This part is the shortlist.

In 2013, John Dunlosky and colleagues^[60] rated ten widely used study techniques against the evidence behind them. Two reached high utility: practice testing and distributed practice, which this part calls retrieval and spacing. Three more reached moderate: elaborative interrogation, self-explanation and interleaved practice. Rereading and highlighting, the two techniques that dominate classrooms and corporate learning alike, came bottom.

That verdict sets the order of this part. None of this is new. Most of it has been sitting in the literature for thirty years or more. The gap is not knowledge.

Figure 9. What each technique in this part is worth

Effect sizes as reported by the studies cited in each section



These come from different studies with different outcome measures, populations and designs, so read them as orders of magnitude rather than a ranking. An effect of 0.40 lifts an average learner to roughly the 66th percentile of the comparison group, 0.83 to roughly the 80th, and 1.39 to roughly the 92nd.



TECHNIQUE 01 OF 07

Retrieval: the difference between testing to measure and testing to learn

WHAT IT IS

Trying to pull something out of memory rather than putting it in again. A quiz, a flashcard, or closing the book and writing down what you can remember.

Assessment asks what a learner knows. Retrieval practice uses the asking to build what they know. Same mechanic, different purpose. Retrieval practice is worth considerably more.

Karpicke and Roediger demonstrated the size of it in 2008.^[61] Students learned forty Swahili-English word pairs. Those who kept being tested on the pairs recalled 80% a week later. Those who kept restudying them recalled 36%. Same material, same total study time, and more than double the retention.

🕒 **Karpicke & Roediger** 2008 · Science

Students learned 40 Swahili–English word pairs. One group kept being tested on them, the other kept restudying them, with total study time held equal. Both were tested a week later.

MATERIAL

40 word pairs

TESTED

1 week later

REPEATED TESTING

80% recalled

REPEATED STUDY

36% recalled

- Total study time was held constant across conditions, which is what makes the comparison worth having.

Almost nobody does this. Surveying 177 undergraduates about how they study, Karpicke, Butler and Roediger^[62] found 84% listed rereading among their strategies and 11% mentioned practising retrieval. Asked to name the single strategy they used most, 55% said rereading. Practising recall was named by 1%. Two students out of 177.

🕒 **WHAT LEARNERS DO**

Reread until it feels familiar. 84% report doing it, and 55% name it as their main strategy.

🕒 **WHAT WORKS**

Put the material away and try to produce it. 11% report doing it, and 1% name it as their main strategy.

The reason is not ignorance, it is fluency. Roediger and Karpicke^[63] found that learners who restudied felt more confident about what they would remember, and then performed dramatically worse than learners who had been tested. Restudying produces the sensation of knowing. Retrieval produces the knowing, and feels worse doing it.

🕒 **DESIGN CONSEQUENCE**

Every quiz in your programme is doing two jobs: measuring what stuck, and making more of it stick. Most organisations bank the first and throw away the second, because the quiz sits at the end of the module rather than through it.



Spacing, and the schedule you don't need to buy

WHAT IT IS

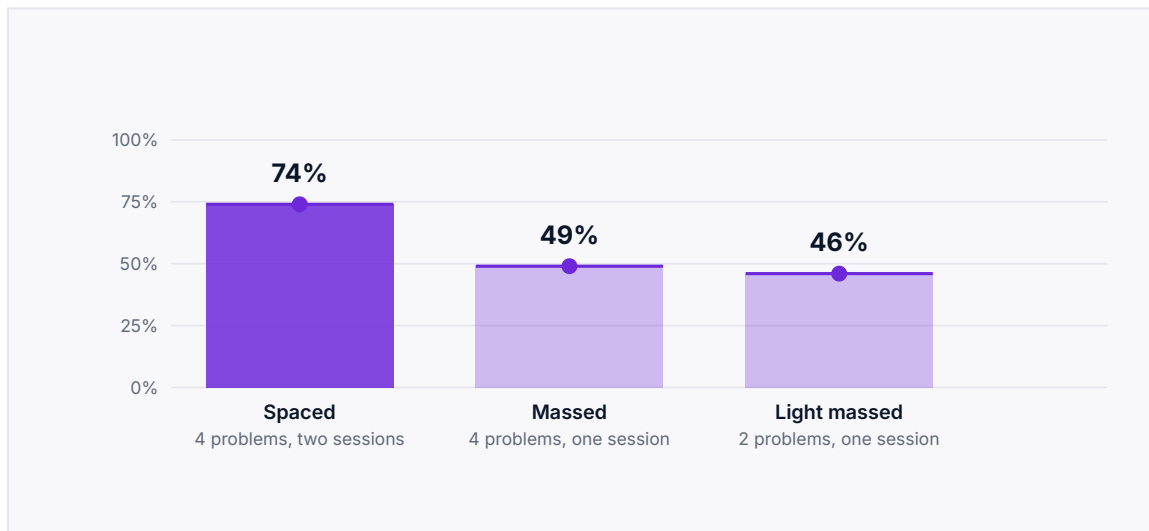
Separating repeated encounters with the same material in time, instead of massing them into one session. The material does not change. Only the calendar does.

Spacing is the other technique that cleared Dunlosky's bar, and the cheapest intervention in this handbook. It needs no new content, no new platform and no new budget. It needs a calendar.

Rohrer and Taylor ran the cleanest demonstration of why in 2007.^[64] Three groups of undergraduates practised the same kind of maths problem, then sat a test a week later.

Figure 10. The gain is in the gap, not the volume

Test accuracy one week after practice · Rohrer & Taylor, 2007



The third group is the one that matters. Doubling the amount of massed practice bought three percentage points. Spreading the same four problems across two sessions bought twenty-five.

How long the gap should be depends on how long you need the material to last. Cepeda and colleagues^[65] taught facts to more than 1,350 people, reviewed them after gaps of up to three and a half months, and tested up

to a year later.

Lengthening the gap raised performance and then lowered it again, so there is an optimum rather than a "longer is better" rule. The optimum tracks the horizon you are designing for, but it does not scale with it: the further out you need the material to survive, the smaller a fraction of that horizon the best gap becomes.

Turned into dates, which is the form you can actually schedule against:

- **Needs to last a week**
Review roughly one to three days after the first session.
- **Needs to last a month**
Review roughly a week later.
- **Needs to last a year**
Review roughly three to five weeks later.

Those dates are our own arithmetic on the proportions Cepeda reports, which run from about 20 to 40% of a one-week delay down to about 5 to 10% of a one-year delay. Treat them as the right order of magnitude, not as a prescription.

Then the part the market would rather not hear. Karpicke and Bauernschmidt^[66] gave learners three repeated tests on the same items and varied both the total spacing and its pattern. More total spacing produced a 200% improvement in long-term retention over tests with no gap at all. But expanding intervals, the schedule almost every spaced repetition product is built on, performed no better than equal or even shrinking ones.

The relative schedule of repeated tests had no discernible impact.

KARPICKE & BAUERNSCHMIDT, 2011

DESIGN CONSEQUENCE

Total spacing is the active ingredient. Algorithms earn their keep by forcing spacing to happen at all, not because their curve is optimal.

[READ ON THE LAB](#)

Spaced Repetition



TECHNIQUE 03 OF 07

Interleaving, and the condition that decides whether it works

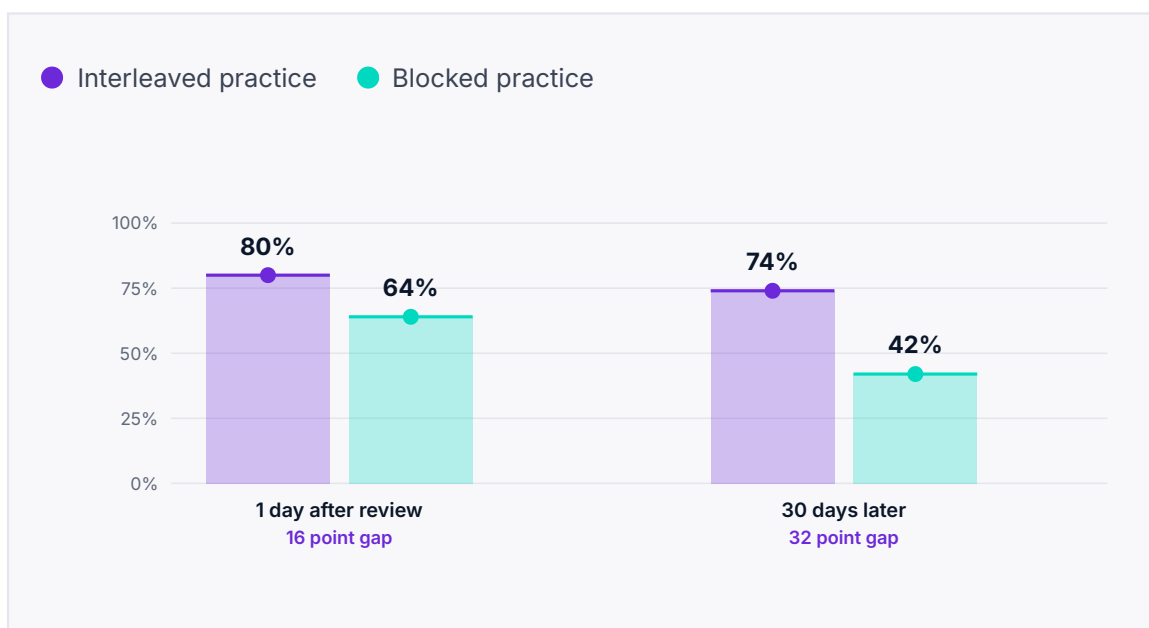
WHAT IT IS

Mixing different problem types within a practice session, so a learner has to work out which method a problem calls for before they can apply it.

Most training is blocked. One topic, practised until it feels solid, then the next. Interleaving mixes the practice instead, so learners have to work out which approach a problem calls for before they can apply it. That second step is the whole point, and blocked practice removes it.

Figure 11. The gap doubles as the delay grows

Percentage of test questions answered correctly · 126 seventh-graders, three-month practice period · Rohrer, Dedrick & Stershic^[67]



A sixteen point gap at one day becomes a thirty-two point gap at thirty. Widening over time is the signature of a durability effect rather than a performance one.

The strongest classroom evidence in this handbook comes from the same team five years later.

🕒 **Rohrer, Dedrick, Hartwig & Cheung** 2020 · Randomised trial [\[68\]](#)

Fifty-four middle-school maths classes were randomised to interleaved or blocked practice of the same material, taught by their own teachers over the school year, then given an unannounced test a month after practice ended.

SAMPLE

787 students

SCOPE

54 classes · 15 teachers

RESULT

61% vs 38%

EFFECT SIZE

d = 0.83

- Unannounced test one month after practice ended. Positive for every one of the 15 teachers. Classes were randomised rather than individual students, which is normal for classroom research and worth stating plainly.

HOW BIG IS 0.83?

d = 0.83



Large by any standard. The average interleaved student finished ahead of roughly 80% of the blocked group.

Set against that, the wider literature is more modest. Brunmair and Richter^[69] pooled 59 studies and found an overall interleaving effect of 0.42, and for mathematics specifically 0.34. For word learning, blocking actually won. One large trial and a whole literature rarely agree, and where they disagree the literature is the safer number to plan against.

There is also a condition, and the paper puts it in the title. Brunmair and Richter called their meta-analysis "Similarity matters". That is the finding in two words. Interleaving pays when the things being mixed are genuinely confusable, because the benefit comes from learning to tell them apart. Mixing unrelated topics does much less, and shuffling a programme for its own sake does nothing.

🕒 DESIGN CONSEQUENCE

Block the first exposure, then interleave everything after it. Interleaving before a learner has any foundation is not a desirable difficulty. It is just difficulty.

[READ ON THE LAB](#)

[Interleaving](#)



TECHNIQUE 04 OF 07

Desirable difficulties: why the thing that feels worst works best

WHAT IT IS

Robert Bjork's term for conditions that make learning slower and harder while it is happening, and stronger once it is done. The three techniques above are all instances of it.

Retrieval, spacing and interleaving share a property. Each one makes practice harder and makes performance during practice worse. Each one also produces more of what is left a month later. Bjork's point was that these are not two separate findings: the difficulty is doing the work.

The clearest single demonstration is not about any of the three. It is about being wrong on purpose. Kornell and colleagues^[70] gave learners word pairs they had no way of guessing. One group saw the cue alone, failed to produce the answer, and was then shown it. The other group simply studied the pair. Both groups had the same thirteen seconds per item.

Learners either guessed at a word pair and failed before being shown the answer, or studied the pair directly. Time per item was held equal, and any trial where somebody guessed correctly was thrown out, so what is being measured is the value of a failed attempt.

FAILED FIRST, THEN SHOWN

67% recalled

STUDIED THE ANSWER

55% recalled

AFTER 38 HOURS

47% against 35%

- Twelve percentage points for time spent failing, and the gap is still twelve points a day and a half later. One honest limit: in the same paper, the same design run with general-knowledge trivia questions at equal time produced 32% against 32%. So failing first pays when the attempt teaches the learner something about the answer. Guessing at a word pair narrows the field. Guessing at a fact you have never met does not, and there the effect disappears.

The reach for an answer prepares the brain for the right one, even when the reach fails.

Three conditions decide whether a difficulty is desirable or merely difficult

🔵 **Learners need a foundation**

Difficulty works by forcing connections to existing knowledge. With no existing knowledge, learners generate misconceptions and then cement them.

🔵 **Feedback is not optional**

An uncorrected guess hardens into a false belief. The productive sequence is always attempt, then answer.

🔵 **The difficulty has to be meetable**

A learner who can reach the answer with effort is learning. A learner who cannot is being overloaded, and nothing is encoded at all.

🕒 DESIGN CONSEQUENCE

Stop treating a struggling learner as evidence of a design fault. Check the three conditions first: do they have a foundation, will they get the answer afterwards, and is the task within reach? If all three hold, the struggle is the mechanism working, and smoothing it out is what would cost you.

READ ON THE LAB

Desirable Difficulties



TECHNIQUE 05 OF 07

Dual coding is not learning styles

WHAT IT IS

Presenting the same idea in words and in pictures, so it is encoded through two channels rather than one. Not to be confused with the myth it is usually mistaken for.

This one needs its ground cleared before it can be used, because it is routinely confused with the most persistent myth in the field.

🚫 LEARNING STYLES

Learners have a preferred visual, auditory or kinaesthetic channel, and matching material to it improves learning. No credible supporting evidence. Believed by 89.1% of educators across 37 studies and 18 countries.

🕒 DUAL CODING

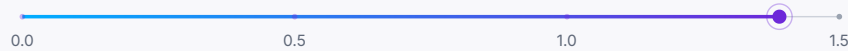
Every learner has two channels, verbal and visual. Material processed through both is held better than material processed through either alone.

Allan Paivio proposed the theory in 1971.^[71] The concreteness effect is the everyday evidence for it: words that evoke an image are recalled better than abstract ones.

Richard Mayer took it into instructional design.^[72] Across eleven controlled experiments in his own laboratory, learners given words and pictures together beat learners given words alone on every single comparison.

WORDS AND PICTURES VS WORDS ALONE

$d = 1.39$



The largest effect in this part. Eleven experiments, all in Mayer's own laboratory, and positive in every single comparison.

The counterintuitive corollary is more useful than the principle. Mayer also found that narration paired with graphics *and* on-screen text produced worse encoding than narration and graphics alone. Two channels beat one. Three inputs across two channels is worse than two, because the verbal channel is now processing the same content twice.

DESIGN CONSEQUENCE

The instruction is not "add visuals". It is: use each channel once, and never say in text what the narration is already saying.

READ ON THE LAB

Dual Coding



TECHNIQUE 06 OF 07

Generation: the brain remembers what it produces

WHAT IT IS

Making the learner produce the answer, the explanation or the example themselves, rather than giving them one that is already written.

Reading an answer and producing one are not the same cognitive event. Only producing builds a memory worth having.

🕒 Two meta-analyses on producing rather than reading

Both pool experiments in which one group was handed the material and the other had to produce it. Bertsch and colleagues cover generating a word or answer. Bisra and colleagues cover prompting learners to explain the material to themselves as they go.

BERTSCH ET AL.

86 studies · 445 effects^[73]

GENERATION EFFECT

$d = 0.40$

BISRA ET AL.

64 studies · 69 effects^[74]

SELF-EXPLANATION

$g = 0.55$

- Bisra and colleagues describe self-explanation as a potentially powerful intervention. In practical terms, 0.55 lifts an average learner into roughly the top third of the group.

Teaching is generation at full stretch, and the benefit arrives before the lesson does. Preparing to explain something forces a learner to organise scattered facts, anticipate questions and find the gaps in their own understanding, and the preparation alone does most of the work.

Which puts most current technology on the wrong side of the argument. A field experiment in Turkish schools^[75] gave students access to GPT-4 during maths practice. While they had the tool in front of them, performance rose sharply: 48% for a standard chatbot, 127% for a version built to tutor rather than answer.

Then the researchers removed it and ran an unassisted exam. The students who had used the standard chatbot scored 17% below students who had never had access at all. The practice gains were real. They belonged to the tool rather than to the learner. The tutoring version largely removed the penalty.

Generative AI Can Harm Learning.

BASTANI ET AL., PAPER TITLE

🕒 DESIGN CONSEQUENCE

The question is not whether your learners use AI. It is whether your AI does the thinking or makes them do it.



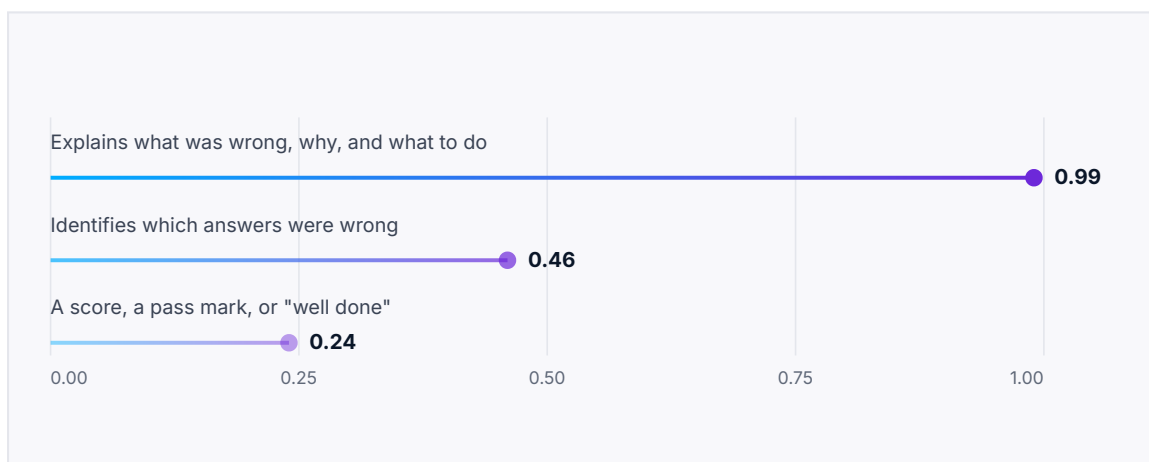
Feedback: more than a third of it makes performance worse

Feedback is the most trusted intervention in corporate learning, and the one most likely to do harm. In 1996, Kluger and DeNisi^[76] pooled 607 effect sizes from 131 studies covering 12,652 people. On average feedback helped, at 0.41. In more than 38% of cases it made performance worse. They noted in their opening line that these negative effects had been "largely ignored" for most of a century, and they were largely ignored for three decades afterwards.

What separates the two outcomes is not warmth, timing or delivery. It is information.

Figure 12. Feedback is worth what it tells the learner

Average effect on learning, by what the feedback tells the learner · 32 meta-analyses · 435 studies · 61,000+ participants · Wisniewski, Zierer & Hattie^[77]



The harm is not evenly spread either: 86% of the negative effects on motivation came from that bottom row. A completion percentage is bottom-row feedback. So is a badge.

The cheapest available fix is also the largest. Butler, Karpicke and Roediger^[78] tested learners on general knowledge facts and gave feedback on half the questions. Testing without feedback produced 41% recall on the final test. Testing with feedback produced 87%.

The most interesting part is what feedback did to the answers learners had got right but doubted: it doubled their retention. Showing the correct answer does not only fix errors. It tells a learner that the thing they hesitated over was right.

Meanwhile the thing teams usually argue about, whether feedback should be instant, makes almost no difference. Across 51 studies, Kandemir and colleagues^[79] found the gap between immediate and delayed feedback was 0.03, statistically indistinguishable from nothing. The authors note that few of those studies used delays longer than a day, so this settles the instant-versus-end-of-session argument rather than the question of feedback arriving a week later.

The reason organisations ship the weak version is not ignorance either. High-information feedback takes an expert twenty minutes a question and does not scale. Scores are free. That cost asymmetry, rather than any disagreement about the evidence, is why corporate feedback looks the way it does, and the part technology is now genuinely changing.

🎯 DESIGN CONSEQUENCE

Write the explanation once, at authoring time, and attach it to the wrong answer. It costs an author twenty minutes and every learner who picks that option gets it for nothing, for the life of the course. A score attached to nothing is the version that carries the risk of making performance worse.

[READ ON THE LAB](#)

Constructive Feedback



PART 3

What this changes

Four consequences, and none of them require new content.

- **Convert your assessments into interventions**

A quiz at the end of a module measures. The same questions distributed through and after it teach, and showing the correct answer afterwards is worth more than forty points of recall for one line of configuration.

- **Schedule the reinforcement before you build the content**

Spacing is the highest-return, lowest-cost change available, and the gap matters far more than the pattern. You do not need an adaptive algorithm to start. You need dates in a calendar.

- **Design for the delayed test, and then actually run one**

Every technique in this part looks worse than the alternative on an end-of-course quiz and better a month later. If completion and immediate scores are your only measures, your data will consistently recommend the weaker option.

- **Spend the feedback budget on explanation, not on delivery**

Timing makes almost no difference, warmth makes almost no difference, and information makes all of it. One well-written explanation of why a wrong answer is wrong outperforms any amount of polish on a score.

How Learning Is Structured

What order things come in, who else is in the room, and whether any of it reaches the job. Six borrowed models, and the evidence sitting underneath them.

THE ARGUMENT OF THIS PART

The models are borrowed frameworks. The evidence sits underneath them, in practices with their own literatures. Where the two come apart, we have followed the evidence.

Parts 1 to 3 dealt with the machinery, the conditions it runs in, and the techniques that work. This part deals with architecture. It is also the part of the field where our industry holds its strongest opinions on the thinnest evidence. Several of the models below are genuinely useful, and almost none of them say what they are commonly quoted as saying. Two of the most-cited numbers in corporate learning appear in this part, and both are wrong by a wide margin, in opposite directions.

Bloom's taxonomy is a vocabulary, not a staircase

WHAT IT IS

A six-way classification of the kinds of thinking a learning objective can ask for. In Anderson and Krathwohl's 2001 revision they are stated as verbs: *remember, understand, apply, analyse, evaluate, create*. Its job is to make an objective specific enough that somebody could mark it.

Benjamin Bloom and four collaborators published the taxonomy in 1956 to give educators a common language for describing what learners should be able to do.^[80] Anderson and Krathwohl revised it in 2001, turning the categories into verbs and moving creation to the top.^[81]

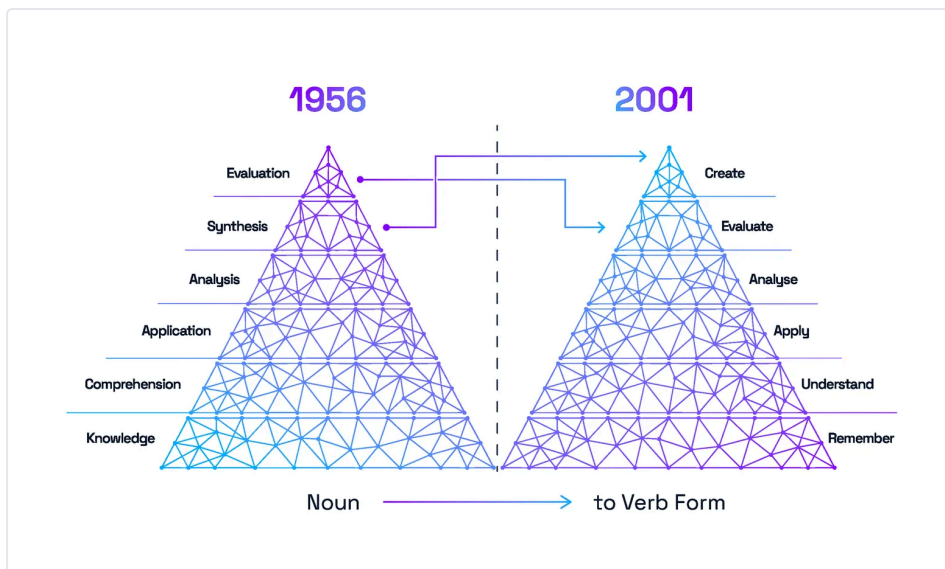
As a vocabulary it is excellent. It is why a learning objective can be written precisely enough to assess. "Understand the policy" is unassessable. "Apply the policy to an unfamiliar case" can be marked.

 THE PYRAMID

Master each level before the next becomes available. Recall, then understand, then apply, and so on up.

© WHAT THE LEVELS ARE

Kinds of cognitive work, not a sequence you climb. Learners analyse before they can recall reliably, and they create while still shaky on the fundamentals.



The 1956 taxonomy beside the 2001 revision. The categories became verbs, and creation moved to the top.

The useful discipline it imposes is on your own ambition. If a programme's objectives all sit at the bottom two levels, it is a knowledge transfer exercise, and no amount of scenario design will turn it into judgement development. That mismatch between stated objective and actual design is one of the most common faults in corporate learning, and Bloom's taxonomy is the cheapest tool for spotting it.

🕒 DESIGN CONSEQUENCE

Write the objectives first, in Bloom's verbs, and read them back. The verbs you have chosen tell you what you are actually building, which is not always what you told the business you were building.

READ ON THE LAB

Bloom's Taxonomy



The zone of proximal development, and what is actually in it

WHAT IT IS

The **zone of proximal development** is the gap between what a learner can already do on their own and what they can do with help. Below the gap the task is trivial. Above it they cannot get there at all. Inside it, help works.

Scaffolding is that help: temporary support that carries the part of the task the learner cannot yet manage. A worked example, a prompt, a partly completed template, a colleague who has done it before.

Both terms get attributed to Lev Vygotsky. Only the zone is his. He defined the zone briefly, in the last two years of his life.

The distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers.

VYGOTSKY, DEFINING THE ZONE OF PROXIMAL DEVELOPMENT ^[82]

He died in 1934, and the English text almost everyone quotes is *Mind in Society*, a compilation assembled and translated in 1978. Reviewing the primary texts, Chaiklin identifies three assumptions in the popular version that Vygotsky never made. ^[83]

01

Generality

That the zone applies to learning any kind of subject matter. Vygotsky was writing about child development, not skills training.

02

Assistance

That the active ingredient is a more competent other. The zone describes a state of readiness, not a delivery method.

03

Potential

That readiness is a fixed property of the learner, and the job is to locate it.

Scaffolding is not his at all. Wood, Bruner and Ross introduced it in 1976, forty years later.^[84] Gredler describes the zone itself as "a minor discussion" in Vygotsky's work about which "inaccurate information... attracted attention early on and became identified as a major aspect of his theory".^[85]

So the attribution is shaky. The practice is not.

🔗 **Belland, Walker, Kim & Lefler** 2017 · Review of Educational Research ^[86]

A meta-analysis of 144 experiments in which computer-based scaffolding was added to STEM teaching, spanning primary school through adult education.

EXPERIMENTS

144

OUTCOMES

333

COGNITIVE OUTCOMES

$\bar{g} = 0.46$

LARGEST EFFECT

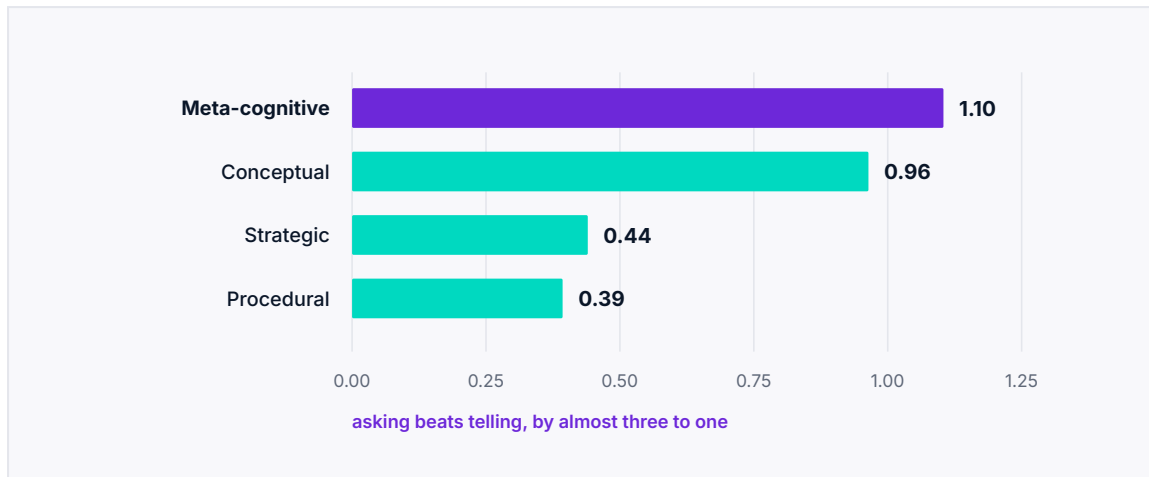
Adult learners

- Computer-based scaffolding in STEM subjects, primary through adult education. The average scaffolded learner finished around where the 68th-percentile unscaffolded learner did, and the effect was greatest among adults, which is unusual for a theory built on child development.

Doo and colleagues went narrower, pooling 64 effect sizes from 18 studies of online higher education covering 4,852 learners.^[87] The overall effect was 0.87. The breakdown inside it is the useful part.

Figure 14. What the scaffold asks for decides what it is worth

Effect on learning outcomes by scaffolding type · 64 effect sizes, 4,852 learners



Doo, Bonk & Heo (2020). Scaffolds that prompt learners to plan and monitor their own thinking are worth nearly three times a scaffold that tells them what to do next. Eighteen studies is a small base, so treat the ordering as firmer than the exact values.^[87]

The one thing everybody says about scaffolding

Fading, the gradual withdrawal of support as competence grows, is treated as definitional. It is also the claim with the least behind it.

⊗ WHAT THE FIELD ASSERTS

Scaffolding must fade. Support that persists creates dependency, so plan its removal from the start.

⊙ WHAT BELLAND'S TEAM FOUND

Fading appeared in only 16.5% of the 333 outcomes, and where it did appear it made no significant difference to the result.^[86]

Read that carefully, because it is not a licence to build permanent crutches. Fading is barely tested, the studies that do test it are mostly intelligent tutoring systems, and absence of evidence at this sample size is not evidence of absence. What it does mean is that anyone claiming fading is the active ingredient is ahead of the data, and that a design budget is better spent on what the scaffold asks the learner to do than on the schedule for taking it away.

Gating progression on demonstrated readiness, which is the zone's core claim, has firmer ground. In the literature this is called **mastery learning**: the learner stays on a unit, with corrective feedback and re-testing, until

they can demonstrate it, and only then moves on. Time varies. The standard does not.

🕒 Kulik, Kulik & Bangert-Drowns 1990 · Mastery learning [88]	
A meta-analysis of 108 controlled evaluations of mastery learning, where students only move on once they can demonstrate the current step.	
EVALUATIONS 108 controlled	EXAM PERFORMANCE 0.52 SD
LOWER APTITUDE 0.61	HIGHER APTITUDE 0.40
• The average student moved from the 50th to the 70th percentile. Lower-aptitude students gained more, but the authors are explicit that this difference was not statistically significant, so do not build a case on it.	

🕒 DESIGN CONSEQUENCE

The instruction is not "find the zone". It is: hold learners at a task until they can do it, and make the support ask them to think rather than tell them what to do.

[READ ON THE LAB](#)

The Zone of Proximal Development



● ● ● ○ ○ ○ STRUCTURE 03 OF 06

Two sigma, and the number that survived

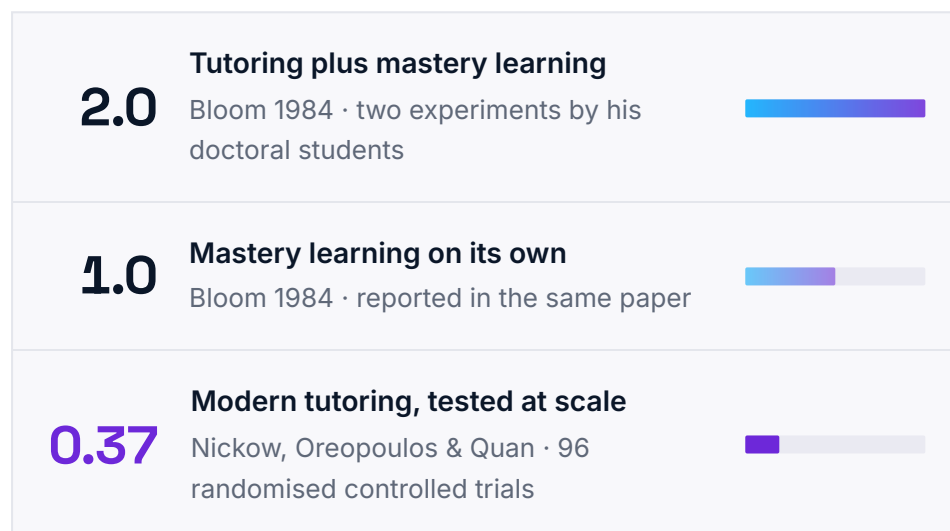
In 1984 Benjamin Bloom reported that students who received one-to-one tutoring combined with mastery learning performed two standard deviations above a conventionally taught control group.^[89]

The average tutored student was above 98% of the students in the control class.

BLOOM (1984)^[89]

About 90% of the tutored students reached the achievement level of the top 20% of the conventional group. The two figures describe the same result from different ends: shift a group by two standard deviations and roughly nine in ten of them clear what used to be the top fifth. It is the most-quoted number in the case for personalised learning, almost always with the mastery learning half dropped.

Bloom did not drop it. His tutoring condition was "followed periodically by formative tests, feedback-corrective procedures, and parallel formative tests", and the same paper puts mastery learning on its own at a full sigma. That is worth sitting with before the debunking starts: mastery learning, with no tutor attached, moved the average student a full standard deviation. Half the famous headline is a technique you can actually afford. Then the other half met a harder test.



Follow that sequence.^[7] Two sigma is two interventions, not one, in two experiments run by Bloom's doctoral students. Test the surviving half at scale across ninety-six randomised trials and you get 0.37. Nobody falsified anything. The number shrank each time it was asked a harder question, which is what usually happens.

Which leaves 0.37 as the honest figure, and 0.37 is still very good. It is comfortably better than most things an L&D team can buy, and the strongest evidence in this handbook for individual attention as a design principle.

DESIGN CONSEQUENCE

The problem was never that tutoring does not work. It is that one-to-one attention has never been affordable at organisational scale, and quoting 2.0 to justify a platform purchase invites the first informed reader to dismantle the argument.

READ ON THE LAB

The 2 Sigma Problem



STRUCTURE 04 OF 06

Other people are the delivery mechanism

WHAT IT IS

Social learning is the observation that people acquire behaviour by watching other people do it. The phrase comes from Albert Bandura, the psychologist whose work in the 1960s and 70s established that a behaviour can be picked up through observation alone, with no direct experience, instruction or reinforcement of the observer's own.^[90]

His more useful contribution for L&D was the distinction he drew afterwards: learning and performance are not the same event. Someone can observe a behaviour, understand it completely, and still not do it. Completion proves exposure and nothing else.

Behaviour modelling training is that theory turned into a method, and it has the strongest workplace evidence in this part.

A meta-analysis of 117 studies of behaviour modelling training: watching a skill demonstrated, practising it, and being given feedback on the attempt.

KNOWLEDGE AND SKILLS

~1 SD

JOB BEHAVIOUR

~0.25 SD

OVER TIME

Knowledge decayed, skills held

- A full standard deviation puts the average trained person ahead of roughly 84% of the untrained group. The gap between the two left-hand cells is the finding, and it cuts against the usual worry about training wearing off: declarative knowledge decayed, while effects on skills and job behaviour "remained stable or even increased".

Watching someone competent reliably teaches the skill. Whether it survives contact with the job depends on conditions the authors specify, and all five are things a programme owner controls.

- **Mixed models.** Learners saw negative as well as positive examples.
- **Trainee-generated scenarios.** Practice used situations the learners produced themselves.
- **Explicit goal setting.** Trainees were instructed to set goals.
- **Managers trained too.** Trainees' superiors went through it as well.
- **Consequences at work.** Rewards and sanctions were instituted in the work environment.

None of which happens in a climate where admitting ignorance is risky.

A meta-analysis pooling 136 samples on psychological safety, the shared belief that speaking up will not be held against you, and what it predicts at work.

SAMPLES POOLED

136

INDIVIDUALS

Over 22,000

LEARNING BEHAVIOUR

$\rho = .62$

FROM

15 samples · 4,648 people

- Among the strongest relationships in organisational research. People do not ask colleagues for help, admit gaps or try something unfamiliar in front of peers where doing so costs them.

🔗 DESIGN CONSEQUENCE

Social learning is not a feature you switch on. It is a climate you either have or do not, with a platform component. Audit the climate before you buy the platform.

[READ ON THE LAB](#)

Social Learning Theory



STRUCTURE 05 OF 06

Metacognition, and learners who cannot judge themselves

Metacognition is the highest-leverage structural addition available, and the cheapest.

A living evidence review pooling 355 studies of teaching pupils to plan, monitor and evaluate their own learning, reported as the additional months of progress a pupil makes in a year.

PROGRESS
+8 months

COST
Very low

STUDIES
355

POPULATION
School-age pupils

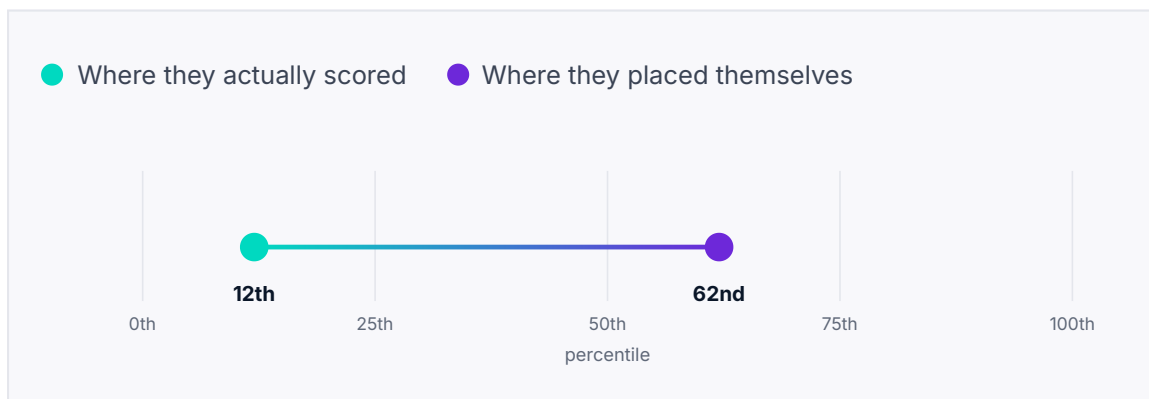
- Two caveats belong on this figure. The evidence base is early years, primary and secondary pupils, so treat it as directional for adults rather than transferable. And the EEF itself notes the topic lost a security padlock because a large share of the underlying studies were not independently evaluated.

The mechanism is straightforward. Learners who plan, monitor and evaluate their own learning select better strategies and notice their own gaps. This is why the scaffolds that prompted planning and self-monitoring outscored the ones that gave instructions. Same finding, approached from the other side.

The problem is that the whole approach runs on self-assessment, and self-assessment is systematically unreliable.

Figure 15. Where people place themselves, and where they land

Self-rated percentile against measured percentile



Kruger & Dunning (1999), logical reasoning study, bottom-quartile participants.^[94] The gap is 50 percentile points, and it runs in one direction.

It is not confined to the laboratory. Zenger surveyed engineers at two technology firms and found 32% at one and 42% at the other believed their skill placed them in the top 5% of performers at their company.^[95] Svenson asked people to rank themselves against the others in the room and found 93% of his American sample placing themselves above the *median* for driving skill.^[96]

🕒 DESIGN CONSEQUENCE

The fix is specific, and "more reflection" is not it. Self-assessment needs an external reference point attached to it: a quiz result, a peer review, a scored scenario, a manager's observation. Reflection without data returns the learner's existing self-image with more words around it.

One warning about what does not belong here. Growth mindset is routinely offered as the metacognitive intervention. It does not survive its own evidence base. Part 6 deals with it in full.

[READ ON THE LAB](#)
[Metacognition](#)



●●●●●●●●●● STRUCTURE 06 OF 06

Transfer is the only measure that matters

Everything in this part exists to serve one outcome, which is whether the capability shows up at work. This handbook opened on our industry's favourite number for this, so the short version will do here. It is fictional, and the real shape is more useful anyway.

🕒 "ONLY 10% OF TRAINING TRANSFERS"

Traces to a single 1982 article by David Georgenson, where it appears as an off-the-cuff estimate voiced by an unnamed training director. No study, no method, no data.^[1]

🕒 WHAT PRACTITIONERS REPORT

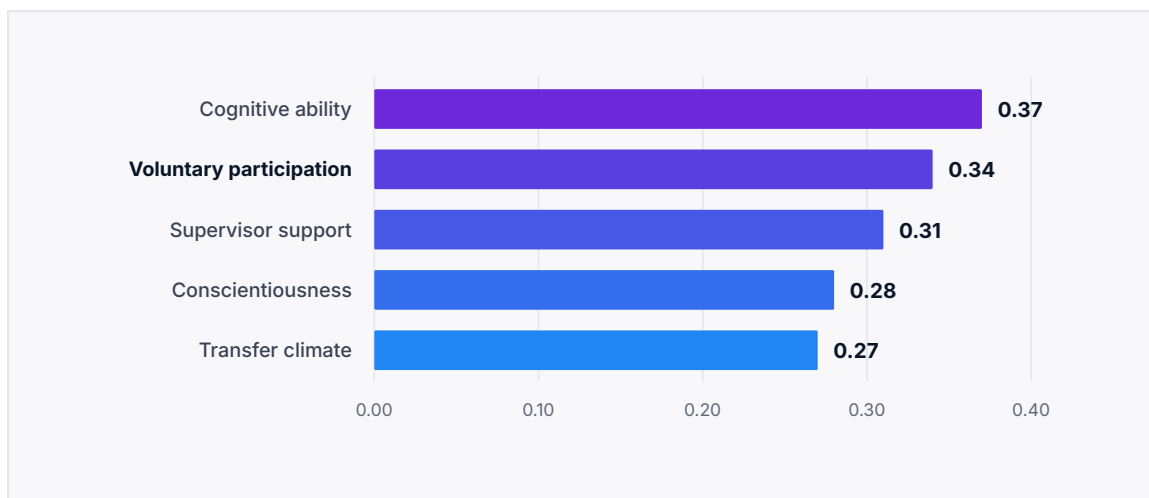
62% still being applied immediately after training, 44% after six months, 34% after a year, across 150 training and development professionals.^[3]

The Saks and Belcourt numbers are practitioner estimates rather than measurements, so hold them loosely. The shape is the point. Transfer does not fail at the door. It decays, on a timescale you can intervene on. Those are different problems with different fixes, and the fictional number points at the wrong one.

As for what drives it, the answer is inconvenient for anyone selling a single solution.

Figure 16. There is no lever, only a set of small additive ones

Corrected correlations with transfer · 89 studies, 325 correlations, 12,496 people



Blume, Ford, Baldwin & Huang (2010).^[97] Their own conclusion: "There is no clear superiority of individual variables over situational variables, or the reverse."

One warning belongs with those numbers, and it concerns a figure the chart deliberately leaves off. Pool every study and the work-environment relationship comes out at .54, which would make it far and away the biggest lever in the set. It more than halves, to .23, once you set aside the studies in which the same trainee rated both their own workplace and their own transfer.

Asking one person two questions on the same form and correlating the answers inflates the result, because somebody who feels warmly towards their employer tends to answer both questions warmly. The smaller figure is the honest one, and it sits in the same range as the transfer-climate estimate above.

DESIGN CONSEQUENCE

The largest single factor is something you select for rather than train. Which makes voluntary participation the most actionable finding in the set, because it sits at .34 and is a decision about how you enrol people rather than how you teach them.

[READ ON THE LAB](#)

[Learning Transfer](#)



PART 4

What this changes

Four consequences follow, and three of them are about what you stop doing.

● Use the models as vocabulary and the evidence as design

Bloom's taxonomy writes better objectives. Vygotsky's zone names a real phenomenon. Neither tells you how to build anything. Scaffolding, mastery gating and behaviour modelling do, and each has its own literature with its own numbers.

● Gate on readiness, not on the calendar

Holding a learner at a task until they can perform it is worth about half a standard deviation across 108 controlled evaluations. Cohort scheduling optimises for administrative convenience, and the learners who need the programme most are the ones who pay for it.

● Make the support ask a question, not give an answer

Scaffolds that prompted learners to plan and monitor their own thinking scored 1.10. Scaffolds that told them what to do next scored 0.39. That is the largest single design difference in this part, and it costs nothing to act on.

● Attach data to every act of self-assessment

Learners' judgement of their own competence is not weakly calibrated, it is confidently wrong in a predictable direction. Reflection prompts without an external reference point make people more articulate about an inaccurate self-image, not more accurate.

Designing Brain-Friendly Learning

Five formats dominate the market. Each has a research base, each is sold as though the format were the active ingredient, and none of them is.

THE ARGUMENT OF THIS PART

Every format here works by making something from Parts 1 to 3 affordable, and every one disappoints when it is bought as a substitute for that thing rather than a route to it. Name the mechanism, or you are buying a wrapper.

Parts 1 to 4 dealt with the machinery, the conditions it runs in, the techniques that work and the architecture around them. This part deals with the formats all of it actually arrives in. Which gives it a single test to apply to any format decision.

FORMAT	WHAT IT ACTUALLY DELIVERS	WHAT IT BECOMES WITHOUT THAT
Microlearning	Spacing and retrieval, made schedulable	A library of short videos
Multimedia	Cognitive load management, mostly by removal	Production value

FORMAT	WHAT IT ACTUALLY DELIVERS	WHAT IT BECOMES WITHOUT THAT
Gamification	Feedback and progress information	Decoration with a scoreboard
Storytelling	Encoding, via a dedicated decoder	Entertainment with a logo on it
Habit design	Return visits, which every other row depends on	A reminder nobody opens



FORMAT 01 OF 05

Microlearning is a schedule, not a length

The industry argument about microlearning is about duration, and duration is the least interesting thing about it.

WHAT IT IS

Learning delivered in short, self-contained units rather than in one long session. The industry argues about how short. The more useful question is what the shortness buys you.

Let's start with the honest case for brevity, because it is real and it has nothing to do with the brain. A five-minute unit fits into a working day. A sixty-minute course requires someone to defend an hour in the diary against everything else competing for it, and to defend it again for every repeat. Short units get started, get finished, and get come back to. That is a logistics argument rather than a cognitive one, and quite strong enough on its own.

What it is not is a research finding. The specific numbers the industry trades in have no study behind them.

A survey asking 144 talent development practitioners how long they thought microlearning should be. Nobody measured any learning.

MAXIMUM LENGTH

~13 minutes

IDEAL LENGTH

10 minutes

MOST EFFECTIVE

2 to 5 minutes
(59%)

- Three different numbers describing practitioner opinion, and not one of them is a finding. ATD's own glossary says the quiet part out loud: microlearning "should be as long as it needs to be to achieve the learning goal".

Ask instead what short units buy you, and the answer is the two techniques that cleared Dunlosky's bar in Part 3. Spacing only works if learners come back, and nobody returns to a sixty-minute course five times. Retrieval only works if the material is small enough to be asked about. Microlearning is the format that makes both of those schedulable.

● **DESIGN CONSEQUENCE**

It works when it is a schedule and it disappoints when it is a library of short videos. The unit of design is the return visit, not the runtime.

The evidence filed under the word "microlearning" is thin: a handful of small studies, mostly with students. The evidence for what microlearning actually delivers is not thin at all, and it comes from working professionals.

Twenty-three studies of practising and trainee health professionals, comparing the same digital training spread out over time against the same material delivered in one block. The outcomes include what clinicians went on to do at work, not only what they scored on a test afterwards.

KNOWLEDGE
AFTERWARDS

0.32 CI 0.13 to
0.51

BEHAVIOUR AT
WORK

0.67 CI 0.43 to
0.91

STUDIES AGREEING

Nearly all on
behaviour

- In plain terms, the average clinician trained on a spaced schedule outperformed roughly three quarters of those trained in one block on **behaviour**, and six in ten on **knowledge**. The behaviour figure is the one worth having. The studies barely disagree, and the authors rate it moderate certainty under GRADE. Spreading the same content out changed what people did at work, not only what they could recall.

The scale is there too. A 2026 review of spaced repetition in medical education pooled 21,415 learners and found an effect of 0.78 against conventional study methods, which puts the average spaced learner ahead of about 78% of those who studied the usual way.^[100] That is a larger body of evidence than anything published under the microlearning label, by roughly thirty to one.

One substitution to refuse, because a careful reader will catch it. Every result above compares spaced delivery against massed delivery. None of them isolates shortness as the active ingredient. What the evidence supports is distribution over time. Brevity is what makes distribution practical, which makes it a supporting argument rather than the finding itself.

One claim to leave behind on the way through

🕒 THE PITCH

Attention spans have collapsed to under a minute, so content must be tiny.

🕒 WHAT GLORIA MARK MEASURED

Time on a single screen before switching: about two and a half minutes in 2003, 47 seconds in her recent work. Others have since found 50 and 44 seconds, median around 40. ^[101] A real, replicated finding about how people move between windows.

It is not evidence that human attention capacity has shrunk, and Part 6 deals with the folklore version of that claim. Treat it as a constraint on the environment your content lands in, not a diagnosis of your learner.

READ ON THE LAB

Microlearning



FORMAT 02 OF 05

Mayer's principles are mostly a list of things to remove

Richard Mayer's cognitive theory of multimedia learning rests on three assumptions. The brain has separate but interacting channels for what it hears as words and what it sees. Each channel has limited capacity. And learning requires the learner to select, organise and integrate, which is effortful. From those, Mayer reports fifteen evidence-based principles drawn from more than two hundred experiments run by his own group. ^[102]

Read the fifteen together and a pattern appears. Four of them make it plain:

Coherence

Remove **extraneous material**. The interesting-but-irrelevant photograph is costing you.

Redundancy	Remove on-screen text that duplicates the narration . Both compete for the visual channel.
Signalling	Remove the learner's need to work out what matters .
Contiguity	Remove the distance between a label and the thing it labels, in space and in time.

The other eleven are in Mayer's own summary and most of them run the same way. Multimedia design guidance is largely subtraction, which is exactly why it is so rarely followed. Adding things is visible work. Removing them looks like having done less.

🔗 Noetel et al. 2022 · Review of Educational Research [103]	
An umbrella review pooling 29 meta-analyses covering 1,189 studies, testing whether Mayer's multimedia principles hold up when all the evidence is combined.	
REVIEWS POOLED 29	STUDIES 1,189
PARTICIPANTS 78,177	PRINCIPLES HELD 11 learning · 5 load
Verbatim: "The largest benefits were for captioning second-language videos, temporal/spatial contiguity, and signaling." Note which one comes first.	

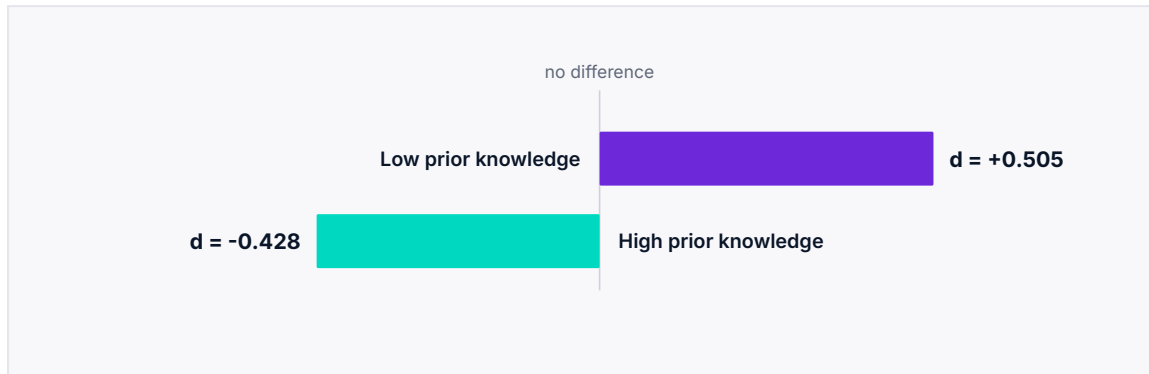
Captioning video for people working in a second language is among the highest-return decisions in the entire multimedia literature, and almost nobody in L&D frames it as a learning intervention rather than an accessibility tick-box.

One principle inverts

It matters more than the rest, because it explains why best-practice arguments go in circles.

Figure 17. The same scaffolding that lifts a beginner holds an expert back

Effect on learning of high-assistance instruction, by prior knowledge · 176 effect sizes, 60 experiments, 5,924 learners



Tetzlaff, Simonsmeier, Peters & Brod (2025).^[104] The bars point in opposite directions, and differ in length.

DESIGN CONSEQUENCE

The asymmetry is your tie-breaker. Helping a novice buys more than withholding help from an expert costs. Where the audience is mixed and segmentation is not available, err towards support.

READ ON THE LAB

Multimedia Learning Theory



FORMAT 03 OF 05

Gamification: the mechanics are not the mechanism

WHAT IT IS

Applying the mechanics of games (points, levels, leaderboards, streaks, badges) to something that is not a game. Nick Pelling coined the term in 2002 for game-like interfaces on cash machines and vending machines, calling it "the deliberately ugly word 'gamification'".^[105]

It reached learning about a decade later, carrying a set of claims that largely cannot be sourced. We tried to trace the best-known gamification statistics in corporate learning back to a primary study. Most lead to a vendor, a press release, or a number that contradicts itself elsewhere on the same website.

A meta-analysis of 22 experimental studies comparing gamified instruction with the same instruction without game elements.

EXPERIMENTAL STUDIES

22, from 2008 to 2023

ACADEMIC PERFORMANCE

$g = 0.782$

PERCENTILE SHIFT

Ahead of ~78%

- The authors' own word is "moderately positive". Their caution is the useful part: the effect varied by educational stage, by subject and by which game design elements were used. Gamification in itself is not the active ingredient.

That finding has an obvious reading and a better one. The obvious reading is that gamification works. The better one is that some game elements are doing the work while others are along for the ride, and the ones that reliably do the work are the ones carrying information.

🔗 **A progress bar** tells a learner where they are against where they need to be.

🔗 **A well-built leaderboard** tells them how their performance compares.

🔗 **A streak** tells them their behaviour has been consistent.

Each of those is feedback with a game mechanic attached, and Part 3 established that the information content of feedback, not its delivery, is what moves performance. A badge with nothing attached to it is the bottom-row feedback that made performance worse in Kluger and DeNisi's data.

● Microsoft, outsourced contact centres

Company account, not a trial

[107]

Microsoft's own published account of gamifying training across its outsourced contact centres. No control group, no independent measurement.

PRODUCTIVITY

+10%

ABSENTEEISM

**12%
improvement**

PAID SUPPORT
ATTACH RATE

Doubled

- Treat these as a company's account of its own programme rather than a controlled trial, because that is what they are.

Agents began trying to squeeze in a couple more calls to receive more points for the day.

MICROSOFT'S OWN DESCRIPTION OF THE MECHANISM^[107]

● DESIGN CONSEQUENCE

The mechanic changed behaviour reliably. Whether it changed the behaviour you wanted is a design question, and yours to answer rather than the mechanic's. Attach information to every mechanic, and check what the information is actually rewarding.

[READ ON THE LAB](#)

Gamification



FORMAT 04 OF 05

Story is the one format the brain has a dedicated decoder for

In 2010 a woman lay in an fMRI scanner at Princeton and told an unrehearsed, fifteen-minute account of something that happened to her as a freshman in high school. Uri Hasson's group recorded her brain activity, played the recording to eleven listeners in the same scanner, and found the listeners reproducing the speaker's patterns a few seconds

behind. In some regions the listeners ran ahead, anticipating what was coming. When the same story was played in Russian to non-Russian speakers, the alignment vanished. ^[108]

That last detail is what makes the study useful rather than merely striking. Neural coupling is not a response to sound, or to paying attention. It tracks comprehension, and it disappears the moment comprehension fails.

Which raises the obvious question for anyone running a training session: if brains fall into step during a story, does being in step predict who actually learns anything? Davidesco and colleagues built a small classroom in a lab to find out. Groups of four students were taught a short science lesson by a real teacher while everyone wore portable EEG, and were then tested immediately and again a week later. ^[109]

Synchrony did predict learning, but not between the people you would expect. Students who were in step *with each other* scored better on both tests. Synchrony with the teacher predicted only the delayed test, and only when the students' brain activity trailed the teacher's by around 300 milliseconds.

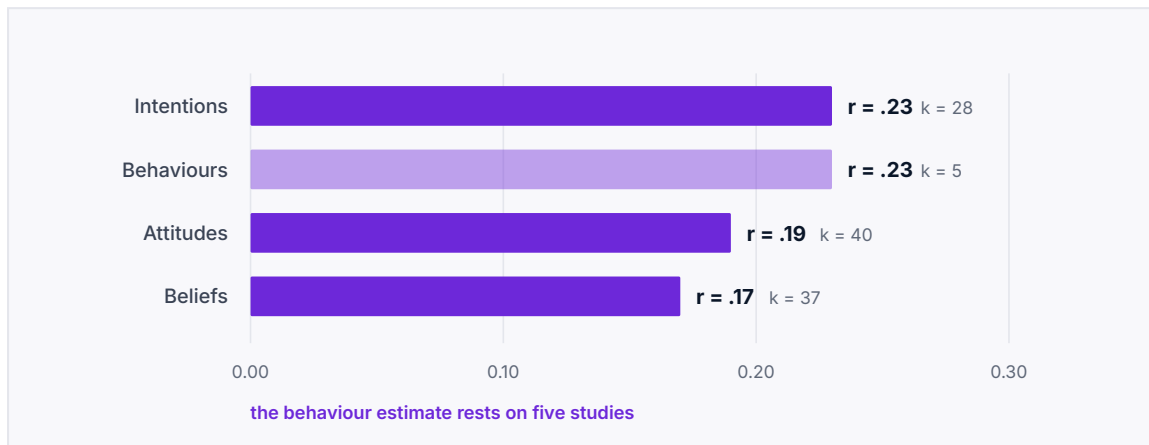
 Davidesco et al. 2023 · Psychological Science	
Nine groups of four students and a teacher sat through four seven-minute science lectures with EEG recording from everyone at once. Students were tested straight after each lecture and again a week later.	
DESIGN	SETTING
9 groups · 4 students + teacher	Simulated classroom
STUDENT TO STUDENT	STUDENT TO TEACHER
Immediate and delayed	Delayed only, ~300ms lag
• A classroom simulated in the lab rather than a real one, and 31 students in the final sample. The lag is the part worth carrying: what tracked delayed recall was not being in step with the teacher, but running a fraction of a second behind.	

The within-brain counterpart is narrative transportation, which Richard Gerrig named in 1993 ^[110] and Green and Brock made measurable in 2000: the state in which someone is absorbed enough in a story that the

room recedes.^[111] Two meta-analyses have asked whether it changes anything.

Figure 18. Small, consistent, and pointing the same way

Narrative persuasion · Braddock & Dillard, correlations with narrative exposure



Braddock & Dillard (2016).^[112] Note the behaviour estimate rests on five studies, which is why it is drawn faint. Van Laer and colleagues, pooling 132 effect sizes from 76 articles, found the same direction: the more transported a reader, the more their attitudes and intentions moved and the fewer critical thoughts they produced.^[113]

DESIGN CONSEQUENCE

A scenario is not a story because it has a name in it. It is a story when the learner is inside a situation with a stake in the outcome, and it earns its cognitive cost only when the thing the story encodes is the thing you need them to do.

Narrative that carries the objective is the most efficient format available.
Narrative that merely entertains is the subject of Part 6.

READ ON THE LAB

The Neuroscience of Storytelling



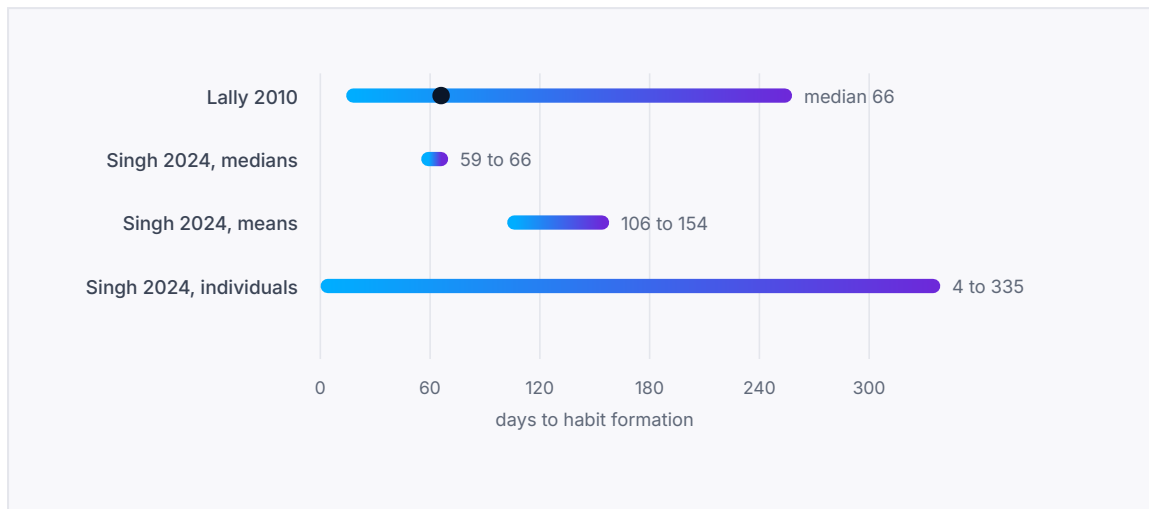
FORMAT 05 OF 05

Habits, and the number that does not exist

Every format above needs the learner to come back, which makes habit design the delivery layer for all of them. It is also where the field's appetite for a clean number is strongest and least satisfied.

Figure 19. How long it takes to form a habit

Days until the behaviour felt automatic · every honest answer is a range, and the ranges disagree with each other



Lally et al. (2010), 82 analysed volunteers.^[9] Singh et al. (2024) pooled 20 studies and 2,601 people, though only four of them reported a median or mean at all.^[114] The 66 days everyone quotes sits at the top end of the shortest bar on this chart, and is a median from a single 2010 study of 82 people.

Domain matters more than elapsed time. Buyalskaya and colleagues applied machine learning to two enormous panels of objective behaviour: over 12 million observations of gym attendance and over 40 million of hospital handwashing.^[115]

Contrary to the popular belief in a "magic number" of days to develop a habit, we find that it typically takes months to form the habit of going to the gym but weeks to develop the habit of handwashing in the hospital.

BUYALSKAYA ET AL. (2023)^[115]

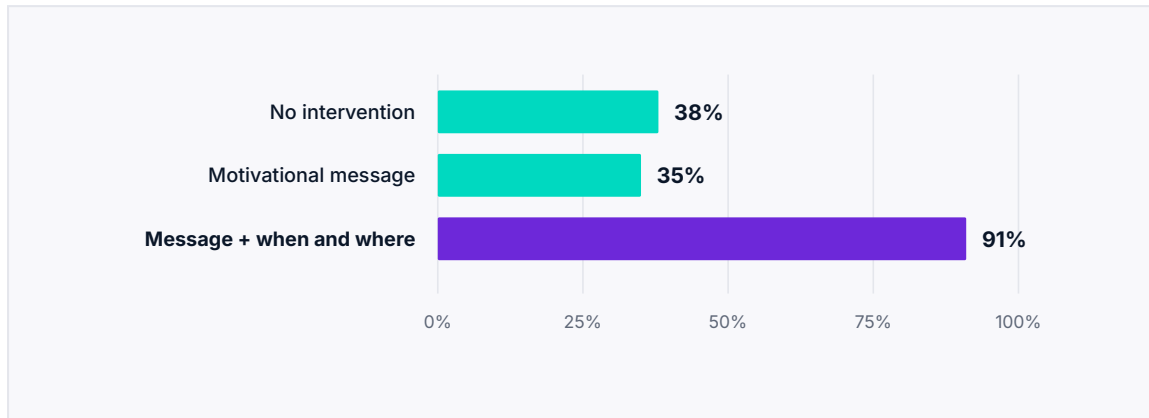
Same species, same mechanism, an order of magnitude between the two answers. The number you are looking for is a property of the behaviour, not of habit formation.

What does replicate is the structure. A cue triggers a routine that delivers a reward, and what you are engineering is the transfer of a behaviour from effortful prefrontal control to automatic execution. Two findings tell you where to spend.

Intention is close to worthless. Specification is powerful.

Figure 20. What a motivational message is worth on its own

Proportion exercising at least once a week



Milne, Orbell & Sheeran (2002).^[116] The motivational message on its own performed slightly worse than doing nothing. Adding a written statement of when and where took participation from 35% to 91%.

Removing the obstacle beats resisting it

Duckworth and colleagues randomly assigned students either to modify their situation or to regulate their own response.^[117] Across two field experiments, one with high school students and one with undergraduates, the students who removed the temptation met more of their academic goals. In the university study the advantage was partly explained by their reporting less temptation during the week, which is the point: they were not resisting better, they had less to resist.

Translated into a learning programme, the obstacles are rarely motivational. They are the things that stand between a person and the first thirty seconds of the task:

- **The sign-in**

A separate password, a VPN, a system nobody is already in. Put the learning where people already are, or make the link open the content rather than a login screen.

- **The search**

Six clicks to find the module somebody has been asked to complete. The reminder should land on the thing itself.

- **The unbooked time**

An expectation to fit it in around the job, which means it competes with the job and loses. A held slot in the calendar removes the daily decision.

- **The all-or-nothing unit**

A 45-minute module that cannot be paused. If the only available gap is ten minutes, nothing starts.

- **DESIGN CONSEQUENCE**

Willpower lost to rearranging the furniture. Ask for the specifics of when and where, and remove the obstacle rather than asking anyone to out-argue it.

[READ ON THE LAB](#)

[Learning Habits](#)



PART 5

What this changes

Four consequences, and each one is a decision you make before the build starts.

- **Name the mechanism before you choose the format**

Every format in this part is a delivery route for something established earlier in this handbook. Microlearning delivers spacing and retrieval. Gamification delivers feedback and progress information. Story delivers encoding. If a format decision cannot be traced to a mechanism, it is a preference dressed as a strategy.

- **Design multimedia by subtraction, and caption your video**

Most of Mayer's principles instruct you to remove something, and the highest-return finding in the largest synthesis available is captioning video for second-language learners. Both are cheap. Neither looks like work, which is why neither gets done.

- **Attach information to every game mechanic**

Points, badges and leaderboards move behaviour whether or not they carry meaning. The ones that improve performance tell the learner something about their performance. The ones that do not are the feedback condition that made people worse.

- **Engineer the cue, not the intention**

A motivational message moved 35% of people against a control group's 38%. The same message plus a written statement of when and where moved 91%. Ask for the specifics, and remove the obstacle rather than asking anyone to out-argue it.

What to Avoid

Almost none of these ideas is simply false. Each contains a true observation and a false inference, and the inference is the part the industry bought.

THE ARGUMENT OF THIS PART

The next bad idea will arrive in the same shape as the last one. A finding that holds, an extension that does not, and a number somewhere in the middle that nobody has traced.

The five parts before this one were about what to build. This one is about what to stop paying for. The temptation is to read it as a myths list. It is not.

THE OBSERVATION, WHICH HOLDS	THE INFERENCE, WHICH DOES NOT
Preferences are real. People will tell you how they like to learn.	Matching instruction to the stated preference improves learning.
People do remember more from doing something than from being told about it.	So the percentages attached to that ladder, 10% of what we read up to 90% of what we do, describe how much is retained.

Experience develops people. Most of it does happen at work.	The split is 70:20:10, and you should treat it as a target.
Beliefs about capability shape behaviour.	A mindset intervention will move attainment at scale.
Interest matters. People remember more from material that holds their attention than from material that bores them.	So make the material more interesting by adding vivid, memorable extras to it.

● ○ ○ ○ ○ ○ ○ IDEA 01 OF 06

Learning styles: the preference is real, the meshing is not

Part 3 established that learning styles have no credible evidence behind them. The more useful question is which claim, precisely, the evidence refutes, because the industry's usual defence is to retreat to a version nobody disputes.

Preferences exist. People will tell you, sincerely and consistently, that they prefer diagrams or discussion or doing. No researcher disputes this. What Pashler and colleagues tested is narrower, and they gave it a name.

The most common, but not the only, hypothesis about the instructional relevance of learning styles is the meshing hypothesis, according to which instruction is best provided in a format that matches the preferences of the learner.

PASHLER, MCDANIEL, ROHRER & BJORK (2008)^[118]

That is the claim that fails, and the only one anything has ever been built on. Their conclusion after looking for studies designed well enough to test it: "there is no adequate evidence base to justify incorporating learning-styles assessments into general educational practice." Willingham and

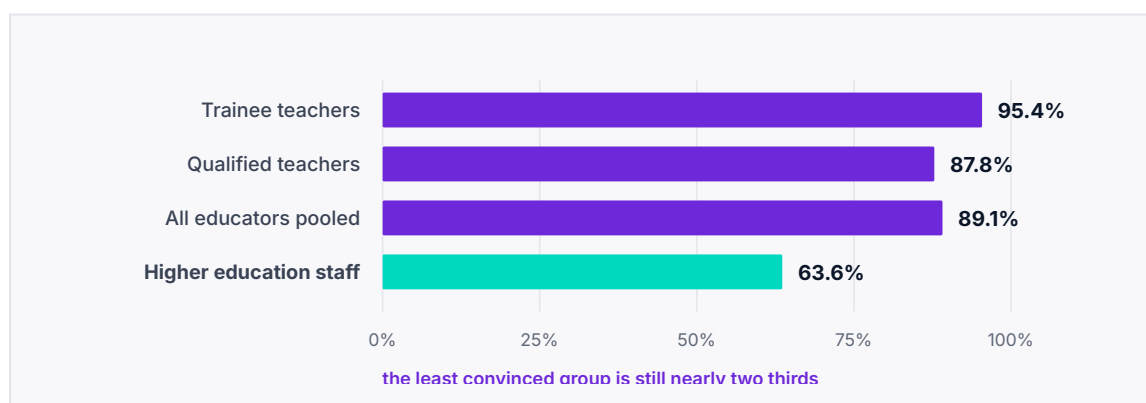
colleagues put it less diplomatically in 2015. Learning styles theories "have not panned out", and telling students so is a professional obligation. [119]

One further detail makes the position harder to argue with: the same review counts 71 separate learning styles schemes, and notes their authors did not claim the list was exhaustive. A field with 71 competing classifications and no agreed one is telling you something about its foundations.

● Krätzig & Arbuthnott 2006 · Journal of Educational Psychology [120]		
Sixty-five university students took three objective memory tests, one visual, one auditory and one done by touch, and their scores were compared with the learning style they said suited them.		
TESTED	AGAINST	CORRELATION
Visual, auditory, kinaesthetic memory	Stated style preference	None
● "Objective test performance did not correlate with learning style preference." When a second study asked participants how they had reached their self-assessment, the answers rested on general beliefs about themselves rather than any specific memory of learning something well one way and badly another. People are not reporting a measurement. They are reporting a self-image.		

Figure 21. Belief has barely moved

Educators agreeing that matching instruction to learning styles is effective · 37 studies, 15,405 educators, 18 countries



Newton & Salvi (2020). [121] Belief is highest among trainee teachers, which tells you where it is acquired. The higher education figure rests on only three of the 37 studies, and even there it is

nearly two thirds.

🕒 DESIGN CONSEQUENCE

Stop defending or attacking "learning styles" in general and name the meshing hypothesis instead. Preferences are worth knowing for engagement and consent. They are not a routing instruction.

[READ ON THE LAB](#)

[Learning Styles](#)



IDEA 02 OF 06

The Cone of Experience: percentages that were never measured

Edgar Dale published his Cone of Experience in 1946 as a visual metaphor for how abstract different kinds of learning experience are. It carried no numbers, no percentages and no claim about retention.

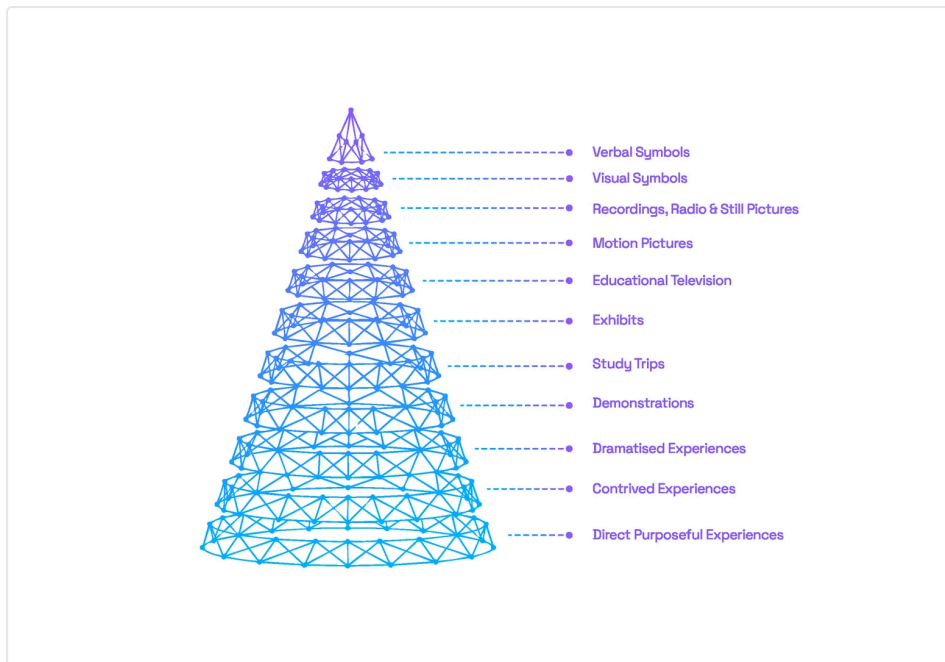
🕒 WHAT CIRCULATES

People remember 10% of what they read, 20% of what they hear, 90% of what they do. Every version disagrees with the others about the exact split, which is the first clue.

🕒 WHAT DALE DREW

Thalheimer, on the original: Dale "included no numbers in his cone" and "warned his readers not to take the cone too literally".
[\[122\]](#)

In 2014 a group of researchers devoted an entire special issue of *Educational Technology* to the corruption. Subramony and Molenda's introduction describes "the corrupted cone and its attendant 'data'" as "akin to a living organism, a virtual 21st century plague, that continues to spread and mutate all over the World Wide Web".^[8] Thalheimer traces the numbers backwards through decades of citation and finds them attached to a chain of secondary sources, with percentages of this shape appearing as early as 1914. At no point does the chain terminate in a study.



Dale's Cone as he drew it in 1946. Eleven bands, ordered by how concrete the experience is, and not a number anywhere.

The most instructive thing about the Cone is not that it is wrong. It is how the error was made. Dale drew a defensible observation, that experience varies in how concrete it is, and someone downstream converted a shape into arithmetic. Nobody fabricated data in a laboratory. They fabricated precision.

● DESIGN CONSEQUENCE

The fix is not to memorise that the percentages are false. It is to notice the shape of the claim. A tidy set of round numbers, arranged in a satisfying gradient and cited to nobody in particular, has almost certainly been made up. Real measurements are untidy, and they come with a sample size attached.

There is a real finding underneath, and it deserves to survive the debunking. Active retrieval beats passive review, and Part 3 gave you the evidence for it. That evidence has effect sizes and named studies. It does not need a pyramid.

READ ON THE LAB

The Cone of Experience



70:20:10: a recollection, not a measurement

The model says development splits 70% experience, 20% social, 10% formal. It is probably the single most quoted number in L&D strategy documents, and it has never been a research finding.

It traces to the Center for Creative Leadership's interview work with executives, published in 1988 by McCall, Lombardo and Morrison as *The Lessons of Experience*.^[10] The method was to ask successful managers to recall what had developed them. That is a survey of recollection, not a measurement of development, and recollection is exactly the faculty the rest of this handbook has shown to be unreliable about learning.

The 70% rule of informal learning needs to be set aside.

CLARDY (2018), AFTER EXAMINING FIVE LITERATURE TRADITIONS^[123]

Alan Clardy's review is the most thorough attempt to establish whether the ratio can be supported. He found the apparent convergence illusory, critiquing the traditions for "sloppy scholarship, inconsistent conceptualizations, and fundamental research protocol problems".

The awkward part for critics is that the underlying observation is sound. Most development does happen at work. Johnson, Blackman and Buick interviewed 145 managers across three phases and found four recurring misconceptions about the model.^[124]

- **That unstructured experience is automatic.** An overconfident assumption that experiential learning will by itself produce capability.
- **That social learning is narrow.** A thin reading of the 20 that misses its integrating role.
- **That formal training changes behaviour on its own.** The expectation that managers will act differently after a course, without active support.
- **That the three parts need no plan.** A lack of recognition that the relationship between them has to be designed.

Along with a finding worth keeping: in their words, "the social aspect of the framework is the 'glue' that integrates formal and experiential learning". Experience on its own is exposure. Experience plus someone to make sense of it is development.

So the ratio is not a target. Used as one, it produces the familiar failure: an organisation announces a 70:20:10 strategy, changes nothing about how work is designed, and relabels its existing courses as the 10.

● DESIGN CONSEQUENCE

Used as a diagnostic it is genuinely useful. Map where your development budget goes, then map where your people say they actually learned. If both land almost entirely in formal delivery, the model has told you something true about your organisation. That is a question, not a quota.

[READ ON THE LAB](#)

The 70:20:10 Model



IDEA 04 OF 06

Growth mindset: the effect that shrank

Part 4 noted in passing that mindset interventions do not carry the weight placed on them. This is the working.

Two meta-analyses in one paper: 273 studies relating mindset to attainment, and 43 studies testing whether mindset interventions change it.

CORRELATIONAL

273 studies

PARTICIPANTS

365,915

INTERVENTIONS

43 studies

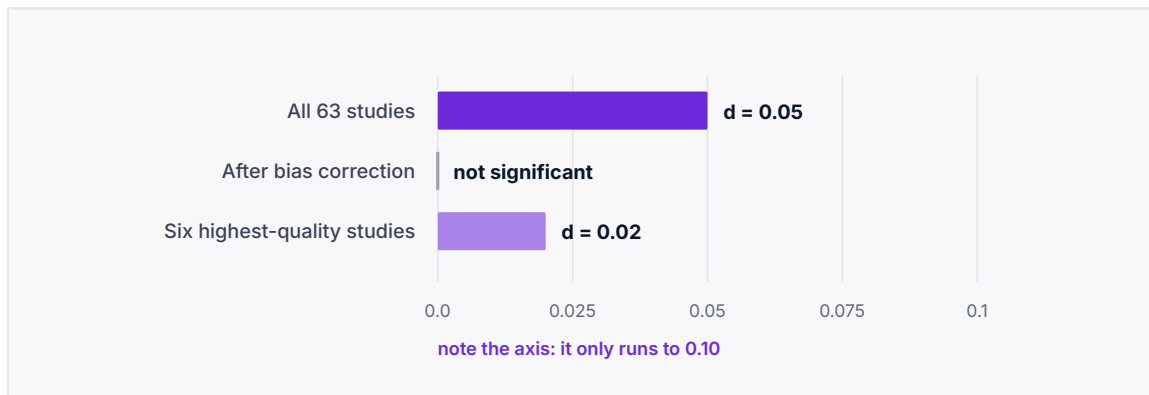
PARTICIPANTS

57,155

- The authors' own summary: "Overall effects were weak for both meta-analyses." Their one positive signal is that students with low socioeconomic status or who are academically at risk might benefit where average students did not.

Figure 23. The better the study, the smaller the effect

Effect size on academic achievement, by study quality · growth mindset interventions · Macnamara & Burgoyne



Macnamara & Burgoyne (2023), 63 studies and 97,672 participants.^[13] Their conclusion: apparent effects are "likely attributable to inadequate study design, reporting flaws, and bias".

Note what that sentence is actually saying. It is not that mindset does nothing. It is that the literature reporting mindset effects is shaped by which studies got published, and that the better the study, the smaller the effect. That is the signature of an effect that is mostly artefact, and a pattern you can look for anywhere.

Yeager and Dweck have engaged with the criticism rather than dismissing it. Their 2020 review argues that "large-scale studies, including preregistered replications and studies conducted by third parties... justify

confidence in growth mindset research", while conceding in the same paper that "mindset effects, however, are meaningfully heterogeneous across individuals and contexts".^[126]

🕒 **National Study of Learning Mindsets** Yeager et al., 2019 · Nature

[127]

A nationally representative randomised trial: 12,490 American ninth-graders across 65 schools got either a short online growth-mindset exercise or a matched control activity, with grades tracked afterwards.

STUDENTS

12,490 ninth-graders

SCHOOLS

65

LOWER-ACHIEVING

+0.11 SD on grades

CONDITION

Supportive peer norms

- The most rigorous trial in the literature, and the benefit held only where peer norms already supported taking on challenging work. The intervention does not create the culture. It needs one.

🕒 **INDEFENSIBLE**

Buy a mindset intervention, deploy it at scale, expect performance to move.

🕒 **DEFENSIBLE**

Beliefs about your own capability are real and worth attending to. They are shaped by what an organisation rewards, not by a workshop.

[READ ON THE LAB](#)

Growth Mindset



IDEA 05 OF 06

Seductive details: interest that competes with the objective

Part 5 distinguished narrative that carries the learning from narrative that merely entertains. This is the research on the second kind.

Garner, Gillingham and White named the effect in 1989.^[128] Seductive details are additions that are interesting to the learner and irrelevant to the objective: the dramatic photograph, the opening anecdote, the arresting statistic that has nothing to do with the task. Harp and Mayer's four experiments with 357 undergraduates found that learners who read a passage carrying such details "recalled fewer main ideas and generated fewer problem-solving transfer solutions" than those who read it without.^[129]

Figure 24. Large in the lab that found it, small across the field

Effect size for removing extraneous material · Mayer's own summary against the pooled literature



Mayer's coherence-principle table^[130] against Cheng, Wu, Wang & Wang (2026), 177 effect sizes from 50 studies.^[131] Both describe the same thing: what padding costs. The meta-analysis also identifies the route, which is extraneous cognitive load rather than any other kind.

Sit with the gap between those two numbers, because it is the most useful thing in this part. That gap is common in psychology and it usually means the smaller number is closer to the truth. Applied here, it says seductive details are a steady tax on comprehension rather than a catastrophe. Worth removing, because removal is free. Not worth stripping a module of everything that makes a human being want to open it.

123 learners took the same biology lesson. The lesson either carried interesting but irrelevant extra detail or it did not, and the lesson text was delivered either as on-screen writing or as narration over the diagrams. The question was whether the cost of the extra detail depends on how crowded the visual channel already is.

DIAGRAM PLUS ON-SCREEN TEXT

The extra detail damaged learning

DIAGRAM PLUS NARRATION

The same detail did no harm

- The same detail was harmful in one condition and harmless in the other, which means the problem is not the detail itself. It is the competition for a channel that is already full. Moving the words to audio freed the eyes for the diagram, and the extra detail then raised interest enough to pay for itself. The authors' mediation analysis confirms interest as the route.

🕒 **DESIGN CONSEQUENCE**

One limitation carries the practical weight. Almost all of this research uses captive participants who cannot leave. Your learners can. A module optimised to the last percentage point for encoding is also a module optimised for abandonment, and the retention rate of a course nobody finishes is zero.

READ ON THE LAB

Seductive Details



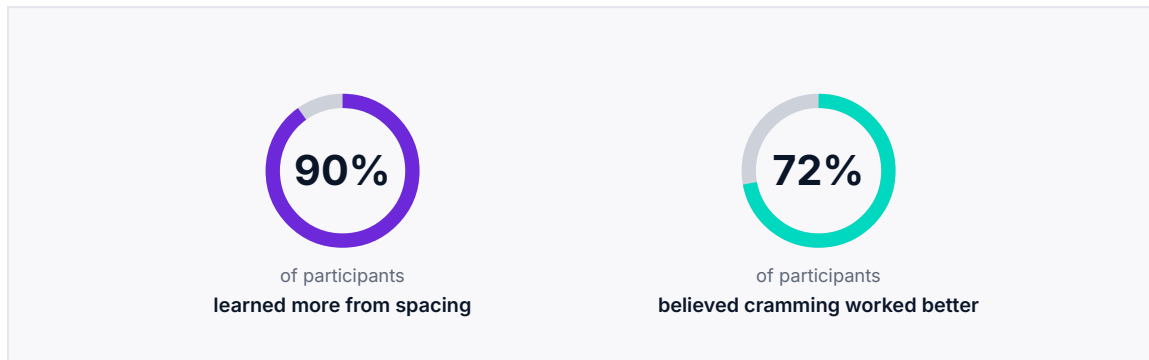
●●●●●● IDEA 06 OF 06

Why bad ideas survive contact with practitioners

The explanation is not that practitioners are careless. It is that learning is one of the few activities where doing it correctly feels worse than doing it badly.

Figure 25. Nine in ten learned more the effective way. Seven in ten reported the opposite.

Percentage of participants · Kornell, across experiments on spacing versus cramming



Kornell (2009).^[133] Verbatim: "spacing was more effective than massing for 90% of the participants, yet after the first study session, 72% of the participants believed that massing had been more effective than spacing." Introspection is not merely an imperfect guide to learning. It is an inverted one.

This is why the evaluation habits of the field entrench the problem. Alliger and colleagues found the correlation between reactions of any type and immediate learning sits at .08, and for affective reactions, whether people enjoyed it, at .02.^[134] Happy sheets measure enjoyment accurately and learning not at all, and every idea in this part scores well on that instrument.

The same meta-analysis contains the fix. Change the question, not the instrument.



One more that keeps coming back, and is not covered anywhere else in this handbook. Nielsen and colleagues analysed resting-state scans from 1,011 people aged 7 to 29 and found that lateralisation is real but local. Their data "are not consistent with a whole-brain phenotype of greater

'left-brained' or greater 'right-brained' network strength across individuals".^[135] Nobody is left-brained. Particular jobs are lateralised, and people are not.

The common thread is not stupidity. It is that all of them are pleasant, memorable, and confirmed by exactly the evidence a busy professional has to hand.

[READ ON THE LAB](#)

[Learning Myths](#)



PART 6

What this changes

Four consequences, and each one is a habit rather than a decision.

- **Separate the observation from the inference**

Almost every idea here is built on something true. Preferences exist, experience develops people, beliefs matter, interest matters. Ask what specific claim the evidence tests, then check whether that is the claim you are buying. It usually is not.

- **Distrust round numbers in a gradient**

10, 20, 70, 90. Fabricated precision has a look, and the look is a tidy series arranged in a satisfying order with no study at the end of it. Ask who measured it, on whom, and how many.

- **Assume the effect will shrink**

The pattern repeats across growth mindset, seductive details and most of this literature: large in the lab that found it, small once the field pools its results. When a headline effect size and a meta-analytic one disagree, plan against the meta-analysis.

- **Change the question on the happy sheet**

Reactions of any type correlate with learning at .08, and enjoyment at .02. But whether learners think the training will help them do their job correlates at .26. Learners who learn the most frequently report learning the least, because effective study feels harder.

What It All Amounts To

Six parts, twenty-two thousand words and more than a hundred and thirty sources went into this. Here is what this all amounts to.

WHAT THIS CLOSING SECTION IS FOR

Each part makes its own case. Some things are only visible from above: patterns that repeat across all six, and that will repeat in the next claim you are asked to believe.

CLOSING 01 OF 04

The argument in one page

● The machinery is small and leaky.

PART 1

Four chunks of working memory, and forgetting as the default. No amount of content design changes either one.

- **Most of what decides the outcome happens outside the course.** PARTS 2 & 4

Sleep, stress, cognitive ability, whether attendance was voluntary, whether it is safe to admit you do not know.

- **A few techniques do most of the work, and they all feel worse.** PART 3

Retrieval, spacing, interleaving, feedback that carries information, generation.

- **The famous models are vocabularies, not instructions.** PART 4

Bloom writes better objectives. Vygotsky names a real phenomenon. Neither tells you how to build.

- **Formats are delivery routes, never mechanisms.** PART 5

Microlearning delivers spacing and retrieval, or it is short video. Name the mechanism or you are buying a wrapper.

- **Almost every bad idea is a true observation with a false inference bolted on.** PART 6

Preferences are real, but matching to them does nothing. Experience develops people, but the ratio was never measured.



CLOSING 02 OF 04

Four things you can only see from above

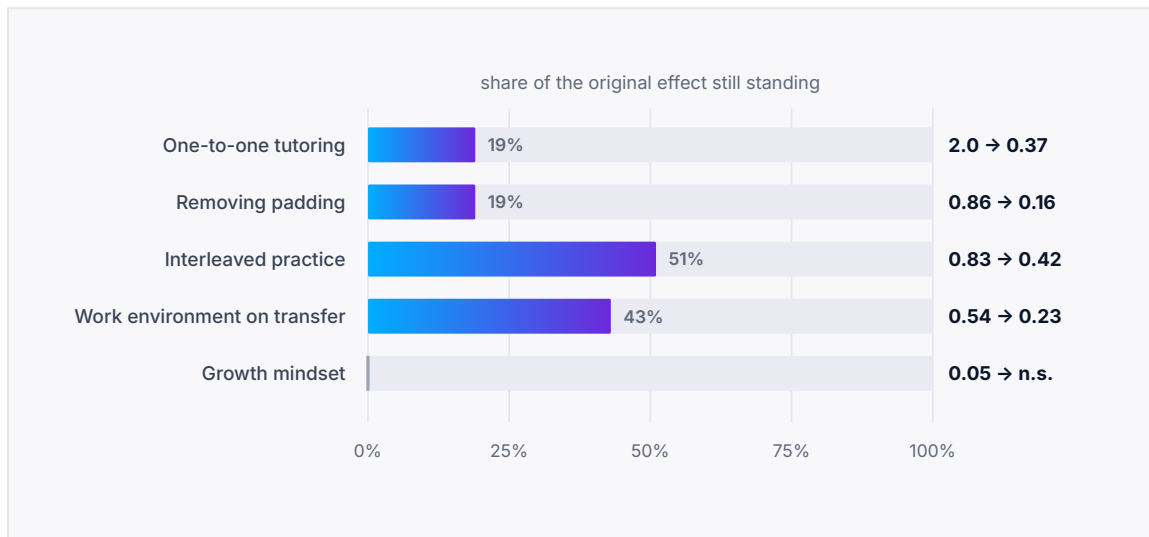
Four patterns repeat often enough to be worth naming, because the next dubious claim you meet will fit one of them.

One: the effect shrinks

The most reliable finding in this handbook is not about learning at all. It is about evidence.

Figure 26. How much of the headline survived a harder question

Five effects in this handbook, each measured twice



Percentages are our own arithmetic on the two published figures in each row, which are not always the same statistic. Every underlying number is registered in the part named. The mechanism differs each time. A bundle is separated, a single trial meets a literature, publication bias is corrected, a measurement artefact is stripped out. The direction never differs.

WHAT TO DO WITH THIS

When a headline effect size and a meta-analytic one disagree, plan against the meta-analysis. With only the headline, assume you are at the top of a curve nobody has tested yet.

Two: the thing that works feels worse

Nine in ten of Kornell's participants learned more from spacing. Seven in ten believed cramming had worked better. Not a calibration error but an inversion, and it explains why our instruments miss it: reactions correlate with learning at .08, enjoyment at .02, and the single-day intensive rates well precisely because fluency peaks when the feedback form appears.

One exception beats the rest of the evaluation apparatus combined. Ask whether learners think the training will help them do their job and the correlation is .26. Change the question, not the instrument.

Three: nearly every high-return action is a removal

This one surprised us, and it runs through every part.

- **Cognitive load.** Strip the extraneous, hardest material first. *Part 1*
- **Stress.** A four-item list of things to take away, not a resilience module to add. *Part 2*
- **Multimedia.** Most of Mayer's fifteen principles are subtractions. *Part 5*
- **Habits.** Duckworth's students succeeded by removing the temptation, not resisting it. *Part 5*
- **Seductive details.** Take out the padding. Removal is free. *Part 6*

Even the large exception proves the point: the highest-return finding in the largest multimedia synthesis available is captioning video for second-language learners, which means attaching a file that usually already exists.

Removal is cheap, fast and invisible. Nobody gets promoted for it, which is exactly why it stays available.

Four: the numbers are worse than the practices

L&D does more sensible things than its evidence base would suggest. Spaced reinforcement, practice, coaching, peer learning: the instincts are broadly right. The arithmetic is not. Nearly every headline figure in the field has no source, has one that says something else, or has one since contradicted. That is a reason to stop quoting the number, not to distrust the practice. The number is what an informed reader will attack.

What to do first

Ordered by what it costs you, cheapest first. Every one traces to a mechanism rather than a preference.



THIS WEEK · FOR NOTHING

Run the four-item stress audit

Novelty, unpredictability, threat to ego, low control. Take them out of your next programme.

PART 2

Turn captions on

The highest-return finding in the multimedia literature, and the file usually already exists.

PART 5

Delete the decoration

The stock photography, and the narration that reads the on-screen text aloud.

PARTS 1 & 5



THIS MONTH · FOR THE PRICE OF A SCHEDULING CONVERSATION

Put a night's sleep in the middle

Between input and application. Kill the two-day intensive that runs straight through.

PART 2

Swap one review for one retrieval

Same content, same slot in the calendar. The learners produce it instead of reading it.

PART 3

Change the post-course question

From "did you enjoy it?" at .02 to "will this help you do your job?" at .26.

PART 6



THIS QUARTER · FOR THE PRICE OF A DESIGN ARGUMENT

Make enrolment voluntary

Wherever you can. It is the largest actionable predictor of transfer in the literature.

PART 4

Gate on readiness, not the calendar

Hold learners at a task until they can do it. Worth about half a standard deviation.

PART 4

Rewrite your scaffolds as questions

Prompts that ask a learner to plan scored 1.10. Prompts that tell them what to do scored 0.39.

PART 4



Ongoing · for the price of a habit

Trace every number in your own materials to a primary source, and delete the ones that do not survive it. This handbook cost us more than forty claims. It was worth it.



CLOSING 04 OF 04

How to read the next claim

The specific myths in Part 6 will be replaced. The shape will not.

Five questions, in the order that disposes of a bad claim fastest. Most fail at the first two.

01

Who measured it, on whom, and how many?

A vendor page, a conference slide or another article is not the source. Keep going until you reach a method section, or stop using the number.

02 What exactly was the claim they tested?

Almost never the claim being sold to you. Preferences exist, but meshing was the hypothesis tested. Tutoring worked, but mastery learning was bundled with it.

03 Is there a pooled figure, and is it smaller?

It usually is. Use the pooled figure as your planning number.

04 Who were the participants, and can yours leave?

Undergraduates who cannot walk out are not your workforce. Almost the entire seductive-details literature rests on captive samples.

05 Does it round suspiciously well?

10, 20, 70, 90. Fabricated precision has a look: a tidy series in a satisfying order with no study at the end of it.

What we got wrong

It would be dishonest to end a handbook about unsourced numbers without saying how many of ours did not survive it. More than forty claims in our own drafts were corrected as we wrote. We had a scaffolding finding exactly backwards, load-bearing enough to have produced a design recommendation. We attributed an imaging result to the wrong paper, described a lab-simulated classroom as a live one, and deleted a tutoring effect size we could not trace.

Every one came from a reputable secondary source, several from our own published articles. That is the point. The distance between "widely repeated by serious people" and "true" is wider in this field than in most, and the only way to close it is to open the paper.

THE STANDARD WE ARE ASKING FOR

Not that you believe this handbook. That you can check it. Every figure in all six parts carries its citation and the sample behind it, which means every one is available to be argued with.



Built for how you're wired

Neurogogy is not a new science. It is the discipline of designing for the brain people actually have, rather than the one the industry finds convenient: small working memory, default forgetting, a state that shifts with sleep and stress, and instincts about its own learning that run almost exactly backwards.

None of the six parts asks you to take anything on trust. That was the whole design.

It is also how we build. L'Oréal Travel Retail runs its beauty advisor programme on exactly the mechanisms this handbook argues for: short units that make spacing and retrieval schedulable, a social layer that gives people a reason to come back, and game mechanics that carry information rather than decoration.

L'Oréal Travel Retail Company account, not a trial

My Beauty Club, a mobile learning programme built with Growth Engineering for beauty advisors across travel retail. Short product and sales units, clubs and a social feed, contests and leaderboards.

REACH

5,500+ beauty advisors

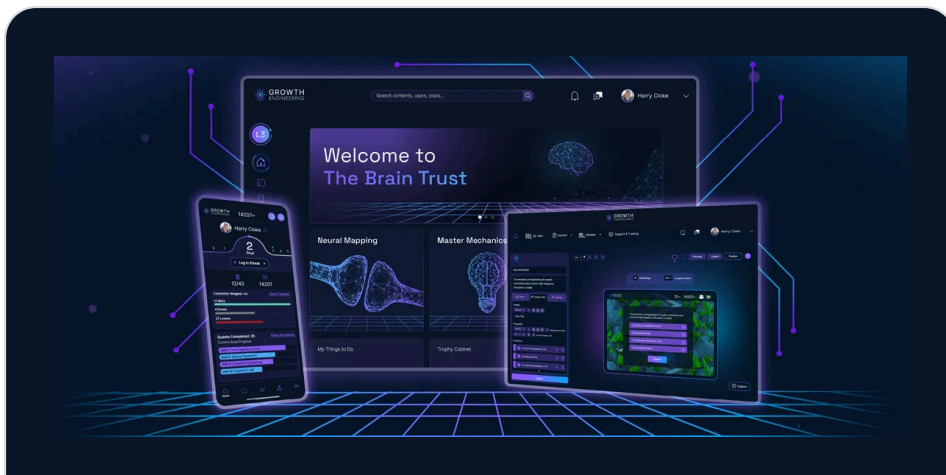
TERRITORIES

18 countries, six languages

SALES REVENUE

+20% in the Americas

- [Read the case study](#). Held to the same standard as everything else here: this is our own account of our own programme, with no control group and no independent measurement, and it belongs in a different category from the studies in the six parts. We include it because the mechanisms are the ones the evidence supports, not because the number proves anything on its own.



THE IMPACT SUITE

Knowing what works is the easy part

Every mechanism in this handbook ends up as a decision somebody makes on a Tuesday: how long, how often, in what order, reinforced how. The Impact Suite runs spacing, retrieval and social reinforcement as defaults rather than as things you remember to schedule. If you would rather talk it through than read another page about it, that works too.

[See how it works](#)

Every claim in this handbook

135 sources, numbered in the order they first appear. Each entry gives the full citation, the sample behind the figure, and a link to the source. Every one was read against the primary source.

1  Georgenson, D. L. (1982). The problem of transfer calls for partnership. *Training and Development Journal*, 36(10), 75–78.

Training and Development Journal, October 1982

Speaker unnamed

No study, method or data

assets.td.org/m/5943dbc6e4b0d42b/original/The-Problem-of...

2  Ford, J. K., Yelon, S. L., & Billington, A. Q. (2011). How much is transferred from training to the job? The 10% delusion as a catalyst for thinking about transfer. *Performance Improvement Quarterly*, 24(2), 7–24.

Perf Improv Q 24(2), 7–24

doi.org/10.1002/piq.20108

3  Saks, A. M., & Belcourt, M. (2006). An investigation of training activities and transfer of training in organizations. *Human Resource Management*, 45(4), 629–648.

150 training and development professionals

62% / 44% / 34%

Estimates on a 0–100% scale

doi.org/10.1002/hrm.20135

4  Bliss, T. V. P., & Lømo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of Physiology*, 232(2), 331–356.

Anaesthetised rabbits

Potentiation lasting 30 min to 10 hours

doi.org/10.1113/jphysiol.1973.sp010273

5 ↑ Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
Cognitive Science 12, 257–285
doi.org/10.1207/s15516709cog1202_4

6 ↑ Knowles, M. S. (1970). *The Modern Practice of Adult Education: Andragogy versus Pedagogy*. Association Press.
[eric.ed.gov · ED043812](https://eric.ed.gov/?id=ED043812)

7 ↑ Nickow, A., Oreopoulos, P., & Quan, V. (2020). The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence. *NBER Working Paper* 27476.
96 randomised controlled trials
0.37 SD
[nber.org · Working Paper 27476](https://nber.org/WorkingPaper27476)

8 ↑ Subramony, D., Molenda, M., Betrus, A., & Thalheimer, W. (2014). The Mythical Retention Chart and the Corruption of Dale's Cone of Experience. *Educational Technology*, 54(6). Special issue.
Citation chain traced to 1913
Dale's 1946 cone carried no numbers
www.worklearning.com/2015/01/05/mythical-retention-data-...

9 ↑ Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, 40(6), 998–1009.
96 volunteers · 82 analysed
Model fitted for **62**, good fit for **39**
Range **18–254 days**
onlinelibrary.wiley.com/doi/abs/10.1002/ejsp.674

10 ↑ McCall, M. W., Lombardo, M. M., & Morrison, A. M. (1988). *The Lessons of Experience: How Successful Executives Develop on the Job*. Lexington Books.
Interview study
Executives recalling their own development
archive.org/details/lessonsofexperie0000unse

11 ↑ Murre, J. M. J., & Dros, J. (2015). Replication and Analysis of Ebbinghaus' Forgetting Curve. *PLOS ONE*, 10(7), e0120644.
Single subject · 70 hours of testing
Savings at 31 days: **21.1%**
pmc.ncbi.nlm.nih.gov/articles/PMC4492928/

12 ↑ King's College London / Savanta ComRes. Public perceptions of attention and memory.

n = 2,093 UK adults

Savanta ComRes, 24–26 Sept 2021

50% believe it · 25% know it is false

www.kcl.ac.uk/news/are-attention-spans-really-collapsing...

13 ↑ Macnamara, B. N., & Burgoyne, A. P. (2023). Do Growth Mindset Interventions Impact Students' Academic Achievement? A Systematic Review and Meta-Analysis with Recommendations for Best Practices. *Psychological Bulletin*, 149(3-4), 133–173.

63 studies

n = 97,672 students

d = 0.05, and 0.02 in the six best studies

artscimedia.case.edu/wp-content/uploads/sites/141/2020/0...

14 ↑ Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry*, 20(1), 11–21.

Case report

IQ 112 post-operatively

doi.org/10.1136/jnnp.20.1.11

15 ↑ Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.

The origin of the seven-item claim

doi.org/10.1037/h0043158

16 ↑ Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.

Central capacity ~4 chunks

Population range 3–5

doi.org/10.1017/S0140525X01003922

17 ↑ Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, 31, 261–292.

Two load types, not three

doi.org/10.1007/s10648-019-09465-5

18 ↑ Barbieri, C. A., et al. (2023). A Meta-analysis of the Worked Examples Effect on Mathematics Performance. *Educational Psychology Review*, 35.

55 studies · 43 articles · 181 effect sizes

g = 0.48

doi.org/10.1007/s10648-023-09745-1

19 ↑
Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16(6), 511–525.

50 studies

Effect concentrated in complex material

doi.org/10.1016/j.learninstruc.2006.10.001

20 ↑
Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., Morris, J. N., Rebok, G. W., Smith, D. M., Tennstedt, S. L., Unverzagt, F. W., & Willis, S. L. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *JAMA*, 288(18), 2271–2281.

2,832 randomised, aged 65–94

4 arms, 10 sessions

Gains only in the trained ability

doi.org/10.1001/jama.288.18.2271

21 ↑
Rebok, G. W., Ball, K., Guey, L. T., Jones, R. N., Kim, H.-Y., King, J. W., Marsiske, M., Morris, J. N., Tennstedt, S. L., Unverzagt, F. W., & Willis, S. L. (2014). Ten-year effects of the ACTIVE cognitive training trial on cognition and everyday functioning in older adults. *Journal of the American Geriatrics Society*, 62(1), 16–24.

2,802 analysed · 44% retained at 10 yr

Speed **0.66** · Reasoning **0.23**

Memory 0.06, n.s.

doi.org/10.1111/jgs.12607

22 ↑
Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., Morris, J. N., Rebok, G. W., Unverzagt, F. W., Stoddard, A. M., & Wright, E. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *JAMA*, 296(23), 2805–2814.

1,877 at 5 years

Speed **0.76** · Reasoning **0.26** · Memory **0.23**

No effect on observed everyday tasks

doi.org/10.1001/jama.296.23.2805

23 ↑ Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, 136(4), 659–676.

Theoretical framework

Flexibility vs plasticity

Supply–demand mismatch

doi.org/10.1037/a0020080

24 ↑ Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature Neuroscience*, 15(4), 528–536.

Review of the imaging evidence

Grey matter from **7 days**

White matter over **6 weeks**

doi.org/10.1038/nn.3045

25 ↑ Woollett, K., & Maguire, E. A. (2011). Acquiring "the Knowledge" of London's Layout Drives Structural Brain Changes. *Current Biology*, 21(24), 2109–2114.

79 trainees · 31 controls at baseline

90 scanned twice

Mean interval **35.3 months**

doi.org/10.1016/j.cub.2011.11.018

26 ↑ Lövdén, M., Wenger, E., Mårtensson, J., Lindenberger, U., & Bäckman, L. (2013). Structural brain plasticity in adult learning and development. *Neuroscience & Biobehavioral Reviews*, 37(9), 2296–2310.

33 imaging studies reviewed

Typical effects **2–5%**

Expansion then partial renormalisation

doi.org/10.1016/j.neubiorev.2013.02.014

27 ↑ Draganski, B., et al. (2004). Changes in grey matter induced by training. *Nature*, 427, 311–312.

24 adults

3 scans over 6 months

Change reversed after practice stopped

doi.org/10.1038/427311a

28 ↑ Judd, N., & Kievit, R. (2025). No effect of additional education on long-term brain structure, a preregistered natural experiment in thousands of individuals. *eLife*, 13, RP101526.

UK Biobank · ~5,124 in optimised-bandwidth analysis

Regression discontinuity on the 1972 ROSLA reform

Null across **117** outcomes

doi.org/10.7554/eLife.101526.3

29 ↑ Dumitru, I., Paterlini, M., Zamboni, M., Ziegenhain, C., Giatrellis, S., Saghaleyni, R., Björklund, Å., Alkass, K., Tata, M., Druid, H., Sandberg, R., & Frisén, J. (2025). Identification of proliferating neural progenitors in the adult human hippocampus. *Science*.
Hippocampus, birth to age 78
Single-nucleus RNA sequencing
Progenitors localised to the dentate gyrus
doi.org/10.1126/science.adu9575

30 ↑ Disouky, A., Sanborn, M. A., Sabitha, K. R., et al. (2026). Human hippocampal neurogenesis in adulthood, ageing and Alzheimer's disease. *Nature*, 652, 1264–1273.
38 donors · **355,997** nuclei
Five cohorts including SuperAgers
Immature neurons reduced in Alzheimer's
doi.org/10.1038/s41586-026-10169-4

31 ↑ Cahill, L., Prins, B., Weber, M., & McGaugh, J. L. (1994). Beta-adrenergic activation and memory for emotional events. *Nature*, 371(6499), 702–704.
Propranolol vs placebo
Emotional story only
doi.org/10.1038/371702a0

32 ↑ Anderson, A. K., Yamaguchi, Y., Grabski, W., & Lacka, D. (2006). Emotional memories are not all created equal: Evidence for selective memory enhancement. *Learning & Memory*, 13(6), 711–718.
48 students · 16 per delay
Neutral **-41%**
Negative **-19%**
doi.org/10.1101/lm.388906

33 ↑ Sharot, T., Martorella, E. A., Delgado, M. R., & Phelps, E. A. (2007). How personal experience modulates the neural circuitry of memories of September 11. *PNAS*, 104(1), 389–394.
24 scanned, 3 years after
83% vs **40%** of each group showed the effect
Left amygdala
doi.org/10.1073/pnas.0609230103

34 ↑ Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5), 459–482.
40 dancing mice
White-black discrimination
Electric shock
psychclassics.yorku.ca/Yerkes/Law/

35 ↑ Corbett, M. (2015). From law to folklore: work stress and the Yerkes-Dodson Law. *Journal of Managerial Psychology*, 30(6), 741–752.
Review of the law's use in management
doi.org/10.1108/JMP-03-2013-0085

36 ↑ Salehi, B., Cordero, M. I., & Sandi, C. (2010). Learning under stress: the inverted-U-shape function revisited. *Learning & Memory*, 17(10), 522–530.
Rats · radial arm water maze
25°C / 19°C / 16°C
doi.org/10.1101/lm.1914110

37 ↑ Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18(1), 105–115.
The origin of the undermining effect
doi.org/10.1037/h0030644

38 ↑ Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125(6), 627–668.
128 studies
All tangible **d = -0.34**
Feedback **d = +0.33**
doi.org/10.1037/0033-2909.125.6.627

39 ↑ Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: A 40-year meta-analysis. *Psychological Bulletin*, 140(4), 980–1008.
k = 183 · N = 212,468
Quality .35 vs .06
Quantity .24 vs .33
doi.org/10.1037/a0035661

40 ↑ Van den Broeck, A., Howard, J. L., Van Vaerenbergh, Y., Leroy, H., & Gagné, M. (2021). Beyond intrinsic and extrinsic motivation: A meta-analysis on self-determination theory's multidimensional conceptualization of work motivation. *Organizational Psychology Review*, 11(3), 240–273.
124 samples $R^2 = .25$
External regulation **0.84%**
doi.org/10.1177/20413866211006173

41 ↑ Howard, J. L., Bureau, J., Guay, F., Chong, J. X. Y., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science*, 16(6), 1300–1323.
344 samples
223,209 participants
26 outcomes
doi.org/10.1177/1745691620966789

42 ↑ Shui, H., & Zhu, X. (2026). Unveiling the relationship between boredom and academic achievement: A three-level meta-analysis. *Social Psychology of Education*, 29.
147 studies · 109 publications
131,446 students $r = -.190$
doi.org/10.1007/s11218-026-10192-3

43 ↑ Qi, W., Liu, J., & Li, X. (2025). The impact of achievement emotions on learning performance in online learning context: a meta-analysis. *Frontiers in Psychology*, 16.
21 studies · 240 effect sizes
Positive $r = .478$
Negative $r = -.303$
doi.org/10.3389/fpsyg.2025.1599543

44 ↑ Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, 11(2), 114–126.
Review
SWS: system consolidation
REM: synaptic consolidation
doi.org/10.1038/nrn2762

45 ↑ Yoo, S.-S., Hu, P. T., Gujar, N., Jolesz, F. A., & Walker, M. P. (2007). A deficit in the ability to form new human memories without sleep. *Nature Neuroscience*, 10(3), 385–392.
14 deprived · 14 rested
Test at 2 days 19% deficit
doi.org/10.1038/nn1851

46 ↑ Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature*, 427(6972), 352–355.
Sleep vs nocturnal wake vs daytime wake
Insight >2× as likely
doi.org/10.1038/nature02223

47 ↑ Van Dongen, H. P. A., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, 26(2), 117–126.
n = 48
14 nights at 4h / 6h / 8h
doi.org/10.1093/sleep/26.2.117

48 ↑ International Labour Organization (2026). *The psychosocial working environment: Global developments and pathways for action*. Geneva, 22 April 2026.
1.37% of global GDP a year
ilo.org · 840,000 deaths a year linked to psychosocial risks at work

49 ↑ de Quervain, D. J.-F., Roozendaal, B., Nitsch, R. M., McGaugh, J. L., & Hock, C. (2000). Acute cortisone administration impairs retrieval of long-term declarative memory in humans. *Nature Neuroscience*, 3(4), 313–314.
Brief communication
Retrieval, not encoding
doi.org/10.1038/73873

50 ↑ Oei, N. Y. L., Elzinga, B. M., Wolf, O. T., de Ruiter, M. B., Damoiseaux, J. S., Kuijer, J. P., Veltman, D. J., Scheltens, P., & Rombouts, S. A. R. B. (2007). Glucocorticoids decrease hippocampal and prefrontal activation during declarative memory retrieval in young men. *Brain Imaging and Behavior*, 1(1–2), 31–41.
21 young men
20 mg hydrocortisone
Placebo-controlled crossover
doi.org/10.1007/s11682-007-9003-2

51 ↑ Newcomer, J. W., Selke, G., Melson, A. K., Hershey, T., Craft, S., Richards, K., & Alderson, A. L. (1999). Decreased memory performance in healthy humans induced by stress-level cortisol treatment. *Archives of General Psychiatry*, 56(6), 527–533.
n = 51
40 or 160 mg/day, 4 days
Reversible
doi.org/10.1001/archpsyc.56.6.527

52 ↑ Shields, G. S., Sazma, M. A., McCullough, A. M., & Yonelinas, A. P. (2017). The effects of acute stress on episodic memory: A meta-analysis and integrative review. *Psychological Bulletin*, 143(6), 636–675.
113 studies · 6,216 participants
Retrieval: $g = -0.215$
102 effects
doi.org/10.1037/bul0000100

53 ↑ Centre for Studies on Human Stress (Sonia Lupien's laboratory). *Sources of stress: the N.U.T.S. recipe*.
Novelty · Unpredictability · Threat to the ego · Sense of control
[humanstress.ca · Sources of stress](https://humanstress.ca/Sources_of_stress)

54 ↑ Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130(3), 355–391.
Social-evaluative threat
Uncontrollability
doi.org/10.1037/0033-2909.130.3.355

55 ↑ Hagger, M. S., Chatzisarantis, N. L. D., et al. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, 11(4), 546–573.
k = 23 labs N = 2,141
d = 0.04
doi.org/10.1177/1745691616652873

56 ↑ Vohs, K. D., Schmeichel, B. J., et al. (2021). A multisite preregistered paradigmatic test of the ego-depletion effect. *Psychological Science*, 32(10), 1566–1581.
k = 36 labs N = 3,531
d = 0.06
doi.org/10.1177/095679762189733

57 ↑ American Psychological Association. *Multitasking: Switching costs*. Research topic page.
The origin of the 40% claim
apa.org/topics/research/multitasking

58 ↑ Wiradhany, W., & Nieuwenstein, M. R. (2017). Cognitive control in media multitaskers: Two replication studies and a meta-analysis. *Attention, Perception, & Psychophysics*, 79(8), 2620–2641.
5 of 14 tests significant
2 survived Bayesian analysis
Meta went null
doi.org/10.3758/s13414-017-1408-4

59 ↑ Watson, J. M., & Strayer, D. L. (2010). Supertaskers: Profiles in extraordinary multitasking ability. *Psychonomic Bulletin & Review*, 17(4), 479–485.
200 participants
2.5% = 5 people
doi.org/10.3758/PBR.17.4.479

60 ↑ Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques. *Psychological Science in the Public Interest*, 14(1), 4–58.
Ten techniques reviewed
2 high · 3 moderate · 5 low
journals.sagepub.com/doi/abs/10.1177/1529100612453266

61 ↑ Karpicke, J. D., & Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science*, 319(5865), 966–968.
n = Swahili-English pairs, computerised flashcards
1 week delay
~80% vs 36%
www.purdue.edu/uns/x/2008a/080214KarpickeTest.html

62 ↑ Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, 17(4), 471–479.

n = 177 undergraduates

Washington University in St. Louis

Open-ended self-report

learninglab.psych.purdue.edu/downloads/2009/2009_Karpicke...

63 ↑ Roediger, H. L., & Karpicke, J. D. (2006). Test-Enhanced Learning. *Psychological Science*, 17(3), 249–255.

2 experiments

Prose passages

Final test at 5 min, 2 days or 1 week

pubmed.ncbi.nlm.nih.gov/16507066/

64 ↑ Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, 35, 481–498.

n = 60 analysed

University of South Florida

Test 1 week after practice

uweb.cas.usf.edu/~drohrer/pdfs/Rohrer&Taylor2007IS.p...

65 ↑ Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridge line of optimal retention. *Psychological Science*, 19(11), 1095–1102.

n = 1,354

Gaps up to 3.5 months

Tested up to 1 year

escholarship.org/uc/item/0kp5q19x

66 ↑ Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1250–1257.

3 repeated tests per item

Expanding / equal / contracting

eric.ed.gov/?id=EJ945342

67 ↑ Rohrer, D., Dedrick, R. F., & Stershic, S. (2015). Interleaved practice improves mathematics learning. *Journal of Educational Psychology*, 107(3), 900–908.

n = 126 seventh-graders

1 day: 80% vs 64%

30 days: 74% vs 42%

files.eric.ed.gov/fulltext/ED557355.pdf

68  Rohrer, D., Dedrick, R. F., Hartwig, M. K., & Cheung, C.-N. (2020). A randomized controlled trial of interleaved mathematics practice. *Journal of Educational Psychology*, 112(1), 40–52.

n = 787 students

54 classes · 15 teachers · 5 schools

d = 0.83, 95% CI [0.68, 0.97]

uweb.cas.usf.edu/~drohrer/pdfs/Rohrer_et_al_2020JEdPsy.ch...

69  Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, 145(11), 1029–1052.

59 studies · 238 effect sizes · 158 samples

Overall g = 0.42

Maths g = 0.34 · words g = -0.39

psycnet.apa.org/doi/10.1037/bul0000209

70  Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 989–998.

J Exp Psychol LMC 35, 989–998

psycnet.apa.org/doi/10.1037/a0015729

71  Paivio, A. (1971). *Imagery and Verbal Processes*. Holt, Rinehart & Winston.

Print only, no online version

Origin of dual coding theory

Book, no open text

72  Mayer, R. E. *Multimedia Learning*, chapter 4: The Coherence Principle. Cambridge University Press.

11 experiments

Median d = 1.39 for words + pictures

www.cambridge.org/core/books/abs/multimedia-learning/coh...

73  Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35(2), 201–210.

86 studies · 445 effect sizes

d = 0.40

doi.org/10.3758/BF03193441

74  Bisra, K., Liu, Q., Nesbit, J. C., Salimi, F., & Winne, P. H. (2018). Inducing Self-Explanation: A Meta-Analysis. *Educational Psychology Review*, 30, 703–725.

64 reports · 69 effect sizes

g = 0.55

doi.org/10.1007/s10648-018-9434-x

75  Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakci, Ö., & Mariman, R. (2024). Generative AI Can Harm Learning.
~1,000 high school students
~15% of the maths curriculum
+48% / +127% / -17%
scale.stanford.edu/publications/generative-ai-can-harm-learning/

78  Butler, A. C., Karpicke, J. D., & Roediger, H. L. Correct-answer feedback and final recall.
2 experiments · general knowledge facts
No test **24%** · test **41%** · test + feedback **87%**
Low-confidence correct: **40% → 85%**
doi.org/10.1037/0278-7393.34.4.918

76  Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance. *Psychological Bulletin*, 119(2), 254–284.
131 papers · 607 effect sizes
n = 12,652 participants · 23,663 observations
d = 0.41 · over 38% negative
psycnet.apa.org/doi/10.1037/0033-2909.119.2.254

79  Kandemir, M., Esposito, G., & Gurgand, L. (2026). A Meta-Analysis of the Impact of Feedback Timing on Learning Outcomes in Computer-Assisted Learning. *Educational Psychology Review*.
51 studies (1988–2024) · 160 effect sizes
g = 0.03, 95% CI [-0.08, 0.13], *p* = .61
link.springer.com/article/10.1007/s10648-026-10117-8

77  Wisniewski, B., Zierer, K., & Hattie, J. (2020). The Power of Feedback Revisited. *Frontiers in Psychology*, 10:3087.
435 studies · 994 effect sizes
n > 61,000
0.99 / 0.46 / 0.24 by information content
www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.003087
...

80  Bloom, B. S. (Ed.), Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. Longmans, Green.
Print only, no online version
Origin of the taxonomy
Five authors, not one
Book, no open text

81 ↑ Anderson, L. W., & Krathwohl, D. R. (Eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman.

Print only, no online version

The 2001 revision

Verbs, and creation at the top

Book, no open text

82 ↑ Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, p. 86.

Print only, no online version

Posthumous compilation

Vygotsky died 1934

Book, no open text

83 ↑ Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. In A. Kozulin et al. (Eds.), *Vygotsky's Educational Theory in Cultural Context* (pp. 39–64). Cambridge University Press.

Generality Assistance

Potential

[blogs.ubc.ca · Chaiklin, the zone of proximal development](https://blogs.ubc.ca/chaiklin-the-zone-of-proximal-development/)

84 ↑ Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100.

Origin of "scaffolding"

1976, not Vygotsky

doi.org/10.1111/j.1469-7610.1976.tb00381.x

85 ↑ Gredler, M. E. (2012). Understanding Vygotsky for the Classroom: Is It Too Late? *Educational Psychology Review*, 24, 113–131.

"a minor discussion"

doi.org/10.1007/s10648-011-9183-6

86 ↑ Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing Results From Empirical Research on Computer-Based Scaffolding in STEM Education: A Meta-Analysis. *Review of Educational Research*, 87(2), 309–344.

144 experiments · 333 outcomes

$\bar{g} = 0.46$

Fading in 16.5% of outcomes

doi.org/10.3102/0034654316670999

87 ↑ Doo, M. Y., Bonk, C., & Heo, H. (2020). A Meta-Analysis of Scaffolding Effects in Online Learning in Higher Education. *The International Review of Research in Open and Distributed Learning*, 21(3).
18 studies · 64 effect sizes
4,852 learners
Overall $g = 0.866$
doi.org/10.19173/irrodl.v21i3.4638

88 ↑ Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of Mastery Learning Programs: A Meta-Analysis. *Review of Educational Research*, 60(2), 265–299.
108 controlled evaluations
0.52 SD
50th → 70th percentile
uky.edu · Kulik, Kulik & Bangert-Drowns (1990)

89 ↑ Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher*, 13(6), 4–16.
Tutoring **2.0 σ**
Mastery learning **1.0 σ**
Two dissertation studies
gwern.net · Bloom (1984), full text

90 ↑ Bandura, A. (1977). *Social Learning Theory*. Prentice Hall. See also Bandura, Ross & Ross (1961), Transmission of aggression through imitation of aggressive models, *Journal of Abnormal and Social Psychology*, 63(3), 575–582.
Observational learning
Learning $\neq$ performance
doi.org/10.1037/h0045925
(Bandura, Ross & Ross, 1961)

91 ↑ Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. (2005). A Meta-Analytic Review of Behavior Modeling Training. *Journal of Applied Psychology*, 90(4), 692–709.
117 studies
Knowledge/skills ~ 1 SD
Job behaviour ~ 0.25 SD
doi.org/10.1037/0021-9010.90.4.692

92 ↑ Frazier, M. L., Fainshmidt, S., Klinger, R. L., Pezeshkan, A., & Vrancea, V. (2017). Psychological Safety: A Meta-Analytic Review and Extension. *Personnel Psychology*, 70(1), 113–165.
136 samples · 22,000+ individuals
Learning behaviour $\rho^2 = .62$
doi.org/10.1111/peps.12183

93 ↑ Education Endowment Foundation. *Metacognition and self-regulation*. Teaching and Learning Toolkit, review last updated May 2025.
+8 months 355 studies
Very low cost
[educationendowmentfoundation.org.uk · Metacognition and self-regulation](https://educationendowmentfoundation.org.uk/Metacognition-and-self-regulation)

94 ↑ Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134.
Study 2 · logical reasoning
Actual **12th** percentile
Estimated **62nd**
doi.org/10.1037/0022-3514.77.6.1121

95 ↑ Zenger, T. R. (1992). Why Do Employers Only Reward Extreme Performance? Examining the Relationships Among Performance, Pay, and Turnover. *Administrative Science Quarterly*, 37(2), 198–219. Reported in Dunning (2011).
Two companies Engineers
32% and **42%** claiming top 5%
[umich.edu · Dunning \(2011\)](https://umich.edu/Dunning%20(2011))

96 ↑ Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47(2), 143–148.
161 students · two samples
US **93%** above median for skill
[doi.org/10.1016/0001-6918\(81\)90005-6](https://doi.org/10.1016/0001-6918(81)90005-6)

97 ↑ 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.
89 studies · 325 correlations
12,496 people
Cognitive ability **.37**
doi.org/10.1177/0149206309352880

98 ↑ ATD Research. *Microlearning: Delivering Bite-Sized Knowledge*, and the ATD talent development glossary entry for microlearning.
144 practitioners
13 min max · 10 min ideal · 59% say 2–5
[td.org · Just How Micro Is Microlearning?](https://td.org/Just-How-Micro-Is-Microlearning?)

99 ↑ Martinengo, L., Ng, M. S. P., Ng, T. D. R., Ang, Y. I., Jabir, A. I., Kyaw, B. M., & Tudor Car, L. (2024). Spaced Digital Education for Health Professionals: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, 26, e57760.

23 studies · spaced vs massed

Knowledge **0.32** · behaviour **0.67**

Moderate certainty (GRADE)
doi.org/10.2196/57760

100 ↑ Maye, J. A., & Hurley, F. (2026). The Effectiveness of Spaced Repetition in Medical Education: A Systematic Review and Meta-Analysis. *The Clinical Teacher*, 23(2), e70353.

14 studies · **21,415** learners

SMD = 0.78, CI 0.56–0.99

PRISMA, MERSQI-assessed
doi.org/10.1111/tct.70353

101 ↑ Mark, G. (2023). *Attention Span*, and the underlying screen-switching studies from her group at UC Irvine.

2003: ~2.5 minutes

Recent: **47 seconds**

Others: 50s, 44s, median 40s

universityofcalifornia.edu · [Can't pay attention? You're not alone](#)

102 ↑ Mayer, R. E. (2024). The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review*, 36, Article 8.

15 principles

200+ experiments

doi.org/10.1007/s10648-023-09842-1

103 ↑ Noetel, M., Griffith, S., Delaney, O., Harris, N. R., Sanders, T., Parker, P., del Pozo Cruz, B., & Lonsdale, C. (2022). Multimedia Design for Learning: An Overview of Reviews With Meta-Analysis. *Review of Educational Research*, 92(3), 413–454.

29 reviews · 1,189 studies

78,177 participants

11 on learning · 5 on load

eric.ed.gov · [EJ1338120](#)

104 ↑ Tetzlaff, L., Simonsmeier, B., Peters, T., & Brod, G. (2025). A cornerstone of adaptivity: A meta-analysis of the expertise reversal effect. *Learning and Instruction*, 98, 102142.

176 effects · 60 experiments

5,924 learners

+0.505 novices · **-0.428**

experts

pedocs.de · [full text](#)

105 ↑ Pelling, N. (2011). The (short) prehistory of "gamification". *Funding Startups (& other impossibilities)*.

Coined late 2002

Cash machines, vending, in-flight video

nanodome.wordpress.com · [The \(short\) prehistory of "gamification"](#)

106 ↑ Zeng, J., Sun, D., & Looi, C.-K. (2024). Exploring the impact of gamification on students' academic performance: A comprehensive meta-analysis of studies from the year 2008 to 2023. *British Journal of Educational Technology*, 55(6), 2478–2502.

22 experimental studies

$g = 0.782$

doi.org/10.1111/bjet.13471

107 ↑ Microsoft New Zealand News Centre (2019). *Using Gamified Performance and Learning to Drive Call Center Agents*.

+10% productivity

12% absenteeism

improvement

Attach rate doubled

news.microsoft.com · [the original post](#)

108 ↑ Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *PNAS*, 107(32), 14425–14430.

1 speaker · 11 listeners

15-minute unrehearsed account

Russian control: coupling vanishes

pnas.org/doi/10.1073/pnas.1008662107

109 ↑ Davidesco, I., Laurent, E., Valk, H., West, T., Milne, C., Poeppel, D., & Dikker, S. (2023). The Temporal Dynamics of Brain-to-Brain Synchrony Between Students and Teachers Predict Learning Outcomes. *Psychological Science*, 34(5), 633–643.

9 groups · 4 students + teacher

31 in the final sample

Simulated classroom

doi.org/10.1177/09567976231163872

110 ↑ Gerrig, R. J. (1993). *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. Yale University Press.

Print only, no online version

Origin of "transportation"

Book, no open text

- 111 ↑ Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5), 701–721.

Made transportation measurable

doi.org/10.1037/0022-3514.79.5.701

- 112 ↑ Braddock, K., & Dillard, J. P. (2016). Meta-analytic evidence for the persuasive effect of narratives on beliefs, attitudes, intentions, and behaviors. *Communication Monographs*, 83(4), 446–467.

Beliefs $r = .17$ ($k=37$)

Attitudes $.19$ ($k=40$)

Intentions $.23$ ($k=28$)

Behaviours $.23$ ($k=5$)

doi.org/10.1080/03637751.2015.1128555

- 113 ↑ Van Laer, T., de Ruyter, K., Visconti, L. M., & Wetzels, M. (2014). The Extended Transportation-Imagery Model: A Meta-Analysis of the Antecedents and Consequences of Consumers' Narrative Transportation. *Journal of Consumer Research*, 40(5), 797–817.

132 effects · 76 articles

Attitudes $p = .44$

Critical thoughts $p = -.20$

doi.org/10.1086/673383

- 114 ↑ Singh, B., Murphy, A., Maher, C., & Smith, A. E. (2024). Time to Form a Habit: A Systematic Review and Meta-Analysis of Health Behaviour Habit Formation and Its Determinants. *Healthcare*, 12(23), 2488.

20 studies · 2,601 people

Medians 59–66 · means 106–154

Individuals 4–335 days

doi.org/10.3390/healthcare1232488

- 115 ↑ Buyalskaya, A., Ho, H., Milkman, K. L., Li, X., Duckworth, A. L., & Camerer, C. (2023). What can machine learning teach us about habit formation? Evidence from exercise and hygiene. *PNAS*, 120(17), e2216115120.

12m+ gym observations

40m+ handwashing observations

Months vs weeks

doi.org/10.1073/pnas.2216115120

- 116 ↑ Milne, S., Orbell, S., & Sheeran, P. (2002). Combining motivational and volitional interventions to promote exercise participation: Protection motivation theory and implementation intentions. *British Journal of Health Psychology*, 7(2), 163–184.

Control 38%

Motivation only 35%

Plus written plan 91%

doi.org/10.1348/135910702169420

- 117 ↑ Duckworth, A. L., White, R. E., Matteucci, A. J., Shearer, A., & Gross, J. J. (2016). A Stitch in Time: Strategic Self-Control in High School and College Students. *Journal of Educational Psychology*, 108(3), 329–341.

Two field experiments

High school (Study 2) · college (Study 3)

doi.org/10.1037/edu0000062

- 118 ↑ Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning Styles: Concepts and Evidence. *Psychological Science in the Public Interest*, 9(3), 105–119.

Commissioned review, not a study

71 schemes counted

[journals.sagepub.com · full text](https://journals.sagepub.com/fulltext)

- 119 ↑ Willingham, D. T., Hughes, E. M., & Dobolyi, D. G. (2015). The Scientific Status of Learning Styles Theories. *Teaching of Psychology*, 42(3), 266–271.

"have not panned out"

doi.org/10.1177/0098628315589505

- 120 ↑ Krätzig, G. K., & Arbuthnott, K. D. (2006). Perceptual learning style and learning proficiency: A test of the hypothesis. *Journal of Educational Psychology*, 98(1), 238–246.

Preference vs measured performance

No correlation

[eric.ed.gov · EJ734402](https://eric.ed.gov/EJ734402)

- 121 ↑ Newton, P. M., & Salvi, A. (2020). How Common Is Belief in the Learning Styles Neuromyth, and Does It Matter? A Pragmatic Systematic Review. *Frontiers in Education*, 5, 602451.

37 studies · 15,405 educators

18 countries **89.1%** overall

doi.org/10.3389/feduc.2020.602451

- 122 ↑ Dale, E. (1946). *Audio-Visual Methods in Teaching*. Dryden Press. Analysed in Thalheimer, W., *Mythical Retention Data & The Corrupted Cone*.

No numbers in the original

Percentages appear by 1914

[worklearning.com · Mythical Retention Data & The Corrupted Cone](https://worklearning.com/MythicalRetentionDataTheCorruptedCone)

- 123 ↑ Clardy, A. (2018). 70-20-10 and the Dominance of Informal Learning: A Fact in Search of Evidence. *Human Resource Development Review*, 17(2), 153–178.

Five literature traditions

Verdict: set it aside

doi.org/10.1177/1534484318759399

- 124 ↑ Johnson, S. J., Blackman, D. A., & Buick, F. (2018). The 70:20:10 framework and the transfer of learning. *Human Resource Development Quarterly*, 29(4), 383–402.

145 participants

5 + 18 + 122, three phases

Four misconceptions

doi.org/10.1002/hrdq.21330

125 ↑ Sisk, V. F., Burgoyne, A. P., Sun, J., Butler, J. L., & Macnamara, B. N. (2018). To What Extent and Under Which Circumstances Are Growth Mind-Sets Important to Academic Achievement? Two Meta-Analyses. *Psychological Science*, 29(4), 549–571.
273 studies · 365,915
43 studies · 57,155

doi.org/10.1177/0956797617739704

126 ↑ Yeager, D. S., & Dweck, C. S. (2020). What can be learned from growth mindset controversies? *American Psychologist*, 75(9), 1269–1284.

The authors' own response

doi.org/10.1037/amp0000794

127 ↑ Yeager, D. S., et al. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, 573, 364–369.
12,490 ninth-graders · 65 schools

+0.11 SD for lower achievers

doi.org/10.1038/s41586-019-1466-y

128 ↑ Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of "seductive details" on macroprocessing and microprocessing in adults and children. *Cognition and Instruction*, 6(1), 41–57.

Where the term originates

doi.org/10.1207/s1532690xcio601_2

129 ↑ Harp, S. F., & Mayer, R. E. (1998). How Seductive Details Do Their Damage: A Theory of Cognitive Interest in Science Learning. *Journal of Educational Psychology*, 90(3), 414–434.

4 experiments · 357 undergraduates

[eric.ed.gov · EJ576496](https://eric.ed.gov/EJ576496)

130 ↑ Mayer, R. E. *Research-Based Principles for Designing Multimedia Instruction*, coherence principle table.

18 of 19 tests

Median **d** = **0.86**

[unh.edu · Mayer's principles table](https://unh.edu/Mayer's%20principles%20table)

131 ↑ Cheng, C., Wu, Y., Wang, R., & Wang, Z. (2026). Seductive Details, Cognitive Load, and Learning Outcomes: A Multi-level Meta-analysis and MASEM. *Educational Psychology Review*, 38.

177 effects · 50 studies

g = **-0.16**

Via extraneous load

doi.org/10.1007/s10648-025-10099-z

132 ↑ Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior*, 44, 267–278.

N = 123 2 × 3 design

Biology lesson

doi.org/10.1016/j.chb.2014.10.061

133 ↑ Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, 23(9), 1297–1317.
Spacing better for **90%**
72% believed the opposite
[williams.edu · full text](#)

134 ↑ Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology*, 50(2), 341–358.
All reactions **.08**
Affective **.02** Utility **.26**
[doi.org/10.1111/j.1744-6570.1997.tb00911.x](#)

135 ↑ Nielsen, J. A., Zielinski, B. A., Ferguson, M. A., Lainhart, J. E., & Anderson, J. S. (2013). An Evaluation of the Left-Brain vs. Right-Brain Hypothesis with Resting State Functional Connectivity MRI. *PLOS ONE*, 8(8), e71275.
1,011 people · ages 7–29
Lateralisation is local
[doi.org/10.1371/journal.pone.0071275](#)
