OBESITY EXPERTISE

A first-principles curriculum in 196 books

BOOK 8 9th of 196 in the series

Teaching, mentorship and coalition building · Rung 1

Teaching as a craft with evidence behind it

HOW PEOPLE LEARN, HOW TO TEACH ONE SESSION, AND HOW TO LEAD THE PEOPLE YOU TEACH

Dr. Harsh Maheshwari

MD Community Medicine Resident, AIIMS Raipur

drharshmaheshwari.com · version 1.0

OBESITY EXPERTISE · BOOK 8

Teaching, mentorship and coalition building · Rung 1

Teaching as a craft with evidence behind it

Dr. Harsh Maheshwari

MD Community Medicine Resident, AIIMS Raipur

contact@drharshmaheshwari.com · drharshmaheshwari.com · ORCID 0000-0001-8848-1918

SERIES	Obesity Expertise: A first-principles curriculum in 196 books
BOOK	Book 8, Teaching, mentorship and coalition building · Rung 1 (9th of 196 in the series)
VERSION	1.0
LAST UPDATED	25 September 2026
AUTHOR	Dr. Harsh Maheshwari, MD Community Medicine Resident, AIIMS Raipur
CONTACT	contact@drharshmaheshwari.com
ORCID	https://orcid.org/0000-0001-8848-1918
NEWEST VERSION	drharshmaheshwari.com

© 2026 Harsh Maheshwari.

This book is shared free, for anyone to read, under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International licence (CC BY-NC-SA 4.0). You may copy, share and adapt it for non-commercial purposes, if you credit the author and share what you make under the same licence.

It quotes and adapts material from OpenStax textbooks (Rice University), which are licensed CC BY-NC-SA 4.0; that is why this book carries the same licence. It also quotes short passages from laws, reports and other works, each credited where it is used; those passages remain under their owners' terms. Fonts: Source Serif 4, Inter and JetBrains Mono, SIL Open Font License 1.1.

HOW THIS BOOK WAS MADE

Written for the author's own study, then shared. Drafted and checked with AI assistance (Claude, by Anthropic) under the author's direction. Every factual claim was checked against a source the checker had open, every quotation against its source text, and every worked answer recomputed. Mistakes can remain; the feedback page says how to report one.

NOT ADVICE

This book is for learning. It is not medical, legal or financial advice, and it does not replace a qualified professional who knows your situation.

About the author



Dr. Harsh Maheshwari

MD Community Medicine Resident, AIIMS Raipur

Harsh Maheshwari completed his MBBS at King George's Medical University (KGMU), Lucknow, and is a resident in MD Community Medicine at the All India Institute of Medical Sciences (AIIMS), Raipur. His MD thesis, on chrononutrition and obesity, was selected by the Indian Council of Medical Research (ICMR) for funding in 2026.

He learns best from first principles: rebuilding an idea from what came before it until the formula feels like the only reasonable answer. This series grew out of that habit.

His long-term aims are to reduce obesity in India and to help make India healthy.

Email	contact@drharshmaheshwari.com
Website	drharshmaheshwari.com
ORCID	0000-0001-8848-1918

Introduction

Why this book exists

Anyone who becomes good at a subject is soon asked to teach it, and later to lead the people who learn it. Both are usually done by imitation. This book treats teaching as a skill with evidence behind it: what the research on learning shows works, what it shows does not, and how to plan, run and measure one teaching session so that your learners can do something afterwards that they could not do before. It ends where leading people begins.

How to read it

Read Book 0 first, or at least the sections this book names as it goes. Then read this one in order. The book asks you to build one thing: a single session, taught to real learners, with a written objective and a measure of what they learned. Plan it as you read, and teach it when you reach the end.

Every section has the same shape. **Definition** states the idea exactly. **In plain terms** says it again in everyday words. **Illustration** works through one real case, usually with a source you can open yourself. **Where this picture breaks** says where the illustration stops being true. **Must-know points** are what should change what you do. Then come **exercises** and, in most sections, a graded **practice set**.

Do the exercises before you look anything up. The worked answers are in the appendix at the back, under the same numbers. The references for each Part follow that Part. The glossary lists every term of art in the plain words it was first given. The last pages list all 196 books in the series and show where this one sits.

If you are new to this

On a first reading, read each section's definition, plain-terms explanation, illustration and must-know points, and do the exercises. Worked derivations can be skimmed and revisited once the idea feels familiar. Keep a calculator beside you: every worked answer can be followed with pen, paper and a calculator. Anything beyond the stated prerequisites is explained where it first appears, or marked as beyond this book. The **Symbols used in this book** page lists every sign and Greek letter with how to read it and where it is first used.

How to send feedback

If something is wrong, unclear, or could be better, please write to **contact@drharshmaheshwari.com**. Put the book number, the version and the section in the subject line (for example: *Book 8, v1.0, section C5*). Say what the text says, what you think it should say, and, if you can, the source.

Corrections are made in new versions. The version number and the date of the last update are on the copyright page of every book, and the newest version of every book is at **drharshmaheshwari.com**.

WHERE THIS BOOK SITS

Book 8 of the series' 196 books (Book 0 and 195 rung books). Before it: Book 7, Question selection and research strategy · Rung 1. After it: Book 9, Scientific writing, visualisation and public communication · Rung 1. The full list is at the back.

Symbols used in this book

Every sign and Greek letter this book prints, how to read it aloud, what it means in this series, and the section where it first appears. Each is also explained where it is first used.

Symbol	Read as	Meaning	First in
SIGNS			
÷	divided by	Division	2 · Testing a teaching method: the comparison, the delayed test and the size of the difference
×	times	Multiplication	6 · Learning styles: a popular claim that fails its own test
–	minus	Subtraction, or a negative number	12 · Measuring what changed: pre-test, post-test, gain
±	plus or minus	A range either side of a value: 11.8 ± 0.6 means 11.2 to 12.4	12 · Measuring what changed: pre-test, post-test, gain
MARKS ON LETTERS			
x^2, 10^{-3}	superscript	Small raised characters are powers: x^2 = x × x, and 10^{-3} = 0.001	Before you start

Contents

Before you start 9

Rung 1 — Introductory 10

1 · Learning is what lasts, and how a session feels is not evidence of it 10
2 · Testing a teaching method: the comparison, the delayed test and the size of the difference 13
3 · Retrieval practice: pulling it out beats putting it in again 20
4 · Spacing: the same practice, spread out, and how far apart 26
5 · What does not work: rereading and highlighting 33
6 · Learning styles: a popular claim that fails its own test 38
7 · Explaining is how you find out whether you understand 47
8 · Working memory is small: cognitive load, and effort that helps 52
9 · Writing a learning objective that can be assessed 56
10 · A test that matches the objective 60
11 · Running one session around retrieval rather than transmission 65
12 · Measuring what changed: pre-test, post-test, gain 71
13 · Liking a session is not learning from it 79
14 · Where teaching sits in Indian medical education 83
15 · Leading people is a craft of its own 87
16 · Words for how a team works 92
References 97

Appendix · Worked answers 125

Glossary 148

The series: all 196 books 153

Before you start

This book uses the following Book 0 sections. Each is summarised where it is first needed; work through Book 0 itself if a summary is not enough.

- Book 0, A4 · Percentages, and the percentage of a percentage
- Book 0, A5 · Ratios, rates and proportions
- Book 0, B6 · Body-size units: kg, m, cm and kg/m^2
- Book 0, C2 · Rearranging an equation
- Book 0, D5 · Average and spread, and why the average is not the person
- Book 0, D6 · Sampling: how a part can tell you about a whole
- Book 0, D7 · Random error and systematic error as different things
- Book 0, F1 · Reading a table
- Book 0, F2 · Reading a figure, and the ways a figure misleads
- Book 0, F4 · Argument: premise, inference, conclusion
- Book 0, F5 · How government works, in the shape you need to read a regulation

Rung 1 — Introductory

1 · Learning is what lasts, and how a session feels is not evidence of it

DEFINITION. In this book, learning means a lasting change in what a person can do. It is measured by a test taken after a delay. The time between study and that test is the retention interval.

Performance is what a learner can do during a session or within minutes of it. Feeling of learning is the learner's own judgement of how much they learned, usually given as a rating.

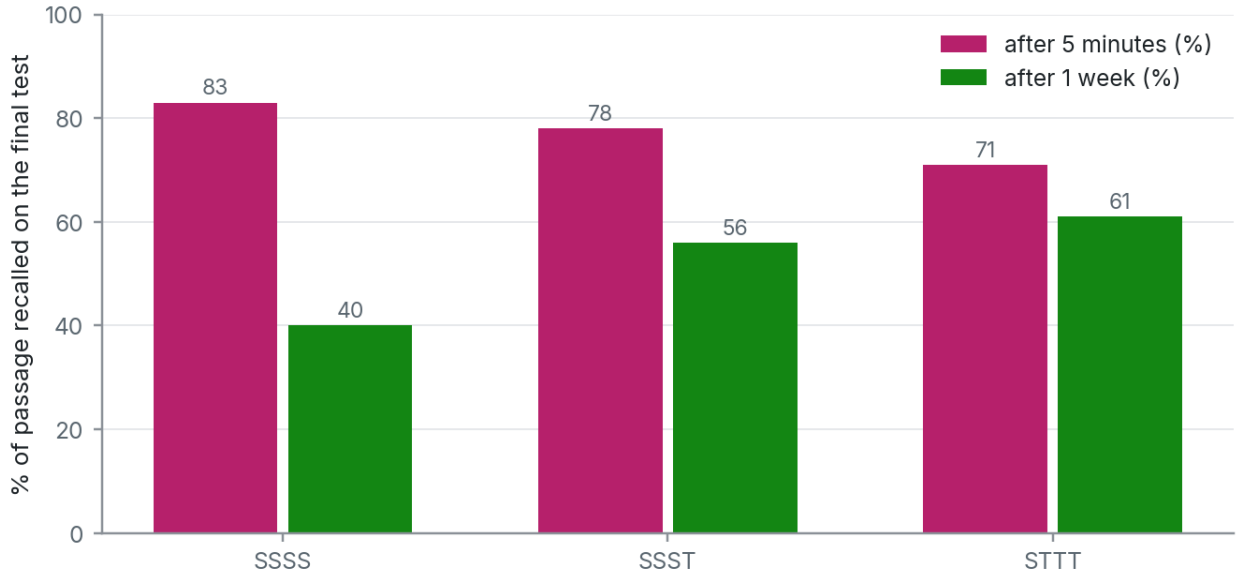
Transmission is teaching by telling: the teacher explains and the learners listen, take notes, and ask the odd question. [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]

IN PLAIN TERMS. A session can go well and teach much less than it seemed to.

Reading something twice makes it feel familiar, and the familiar feels learned. A smooth lecture feels clear, and the clear feels learned. Work that makes learners struggle feels worse, even when it teaches more.

So a quiz at the end of your session measures performance. A feedback form measures feeling. Neither measures learning. You find out about learning by testing again after a delay, a week or more, with new questions on the same objective.

"They enjoyed it and they answered well at the end, so they learned it." Both premises can be true. The inference still does not carry, because the conclusion is about a later day that neither premise measured.



Percentage of the passage recalled on the final test in Roediger and Karpicke's second experiment. After 5 minutes, the group that read the passage four times (SSSS) recalled most. After 1 week, the group that read it once and was tested three times (STTT) recalled most. The order of the three groups turns round.

ILLUSTRATION. You have just taught a batch of third-year MBBS students (the Bachelor of Medicine and Bachelor of Surgery course) how to classify body-mass index. At the end, you hand out a feedback form and a short quiz. The forms say the session was clear. Most students get the quiz right. You are about to write "objective achieved" in your teaching log.

Before you write it, open Roediger and Karpicke, "Test-enhanced learning", *Psychological Science* 2006, at https://doi.org/10.1111/j.1467-9280.2006.01693.x. Go to Experiment 2.

Students learned one short prose passage. Each student had four periods in a row, used in one of three ways. S stands for a study period: 5 minutes of reading. T stands for a test period: 10 minutes of writing down everything they could remember, with no feedback.

- SSSS: read it in all four periods, no test.
- SSST: read it three times, then one test.
- STTT: read it once, then three tests.

Then everyone took one final recall test. Half took it 5 minutes later. The other half took it 1 week later. Here are the percentages of the passage recalled on that final test.

group	after 5 minutes (%)	after 1 week (%)
SSSS	83	40
SSST	78	56
STTT	71	61

Read the first column. Reading four times wins.

Now read the second column. The order has turned round.

```
61 minus 40 = 21
```

The group that was ahead by 12 percentage points at 5 minutes is behind by 21 a week later.

Now find Table 2. At the end of the first session, before any final test, each student filled in a short rating form. One question asked how well they would remember the passage in 1 week, on a 7-point scale. The SSSS group gave 4.8. The STTT group gave 4.0. The group most sure it would remember was the group that remembered least a week later.

Your quiz is the first column: performance, minutes after the teaching. Your feedback form is the rating in Table 2: a feeling. Neither is the second column. To fill in the second column for your own batch, set a short test on the same objective a week or more later.

Figures quoted in this illustration are from [14].

WHERE THIS PICTURE BREAKS. The numbers come from undergraduates at one university. The numbers do not tell you what happens after a month. A test taken minutes after teaching, or a learner's own rating, can point the opposite way from a later test.

MUST KNOW POINTS FOR YOU.

- A high score on a quiz at the end of your session is not evidence that the batch learned. It measures performance minutes after the teaching.
- Do not read a learner's rating of how much they learned as a measure of learning. In Deslauriers's experiment, every student had one class taught actively and one taught by lecture, in an order decided by chance. In their active class they scored almost half a standard deviation (the spread of scores, Book 0, D5) higher on the end-of-class test. They rated their learning more than half a standard deviation lower.
- Before you call a session a success, book a second short test on the same objective, a week or more later.
- When you ask learners to work problems instead of watching you solve them, expect some to say they are learning less. Tell them at the start that effort is part of how it works. One course that did this saw most students say their view of the method improved, though nothing was compared.
- When a committee judges teachers by students' feedback ratings, say that the ratings measure feeling of learning.
- Test at the interval you actually care about, and do not treat a one-week result as proof of a six-month one.
- A learner's own rating is weakest where you most need it, in a beginner. Deslauriers found that ratings tracked test scores more closely in students who already knew more physics. With a new batch, give a test before you trust a show of hands.

EXERCISE 1 (interpretation). Your teaching log for one session reads: "End-of-session quiz, mean 8 out of 10. Feedback form, mean 4.5 out of 5 for 'I understood the topic'. Objective achieved." Say what each number measures. Say what you would add before writing "objective achieved".

2 · Testing a teaching method: the comparison, the delayed test and the size of the difference

DEFINITION. A fair test of a teaching method compares two groups of learners who differ in one thing only: the method. Both groups learn the same content, in the same time, and sit the same test after the same delay. That delay is the retention interval, as the last section defined it.

The learners are allocated to the two groups at random: a chance process decides who goes where, not the teacher or the learners. Anything else that affects the score is what the OpenStax statistics textbook calls a lurking variable. Random allocation spreads the lurking variables across the groups by chance.

The mean difference, MD, is the mean score of one group minus the mean score of the other. Its sign says which group scored higher; whether higher is better depends on the scale. So you always state which group was subtracted from which, and which way the score runs.

The standardised mean difference, SMD, is the MD expressed as a number of standard deviations (SD, Book 0, D5). In words, the SMD is the MD divided by the SD, written

```
SMD = MD ÷ SD
```

The choice of SD is part of the definition. The SMD the Cochrane Handbook uses, Hedges' adjusted g, divides by an SD pooled from both groups. Papers often report an SMD as an "effect size".

One group tested before and after its teaching, with no comparison group, gives a change, not a difference between methods. Divided by an SD it gives a number that looks like an SMD. But it measures one group's change over time, so it cannot be set beside an SMD between two groups. [15, 15, 16, 17, 18, 19, 20, 21, 22, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 44, 45, 46, 3, 47, 48, 49, 50]

IN PLAIN TERMS. Here is the trap with judging your own teaching. You try a new method with a batch. Their marks go up. It feels like proof.

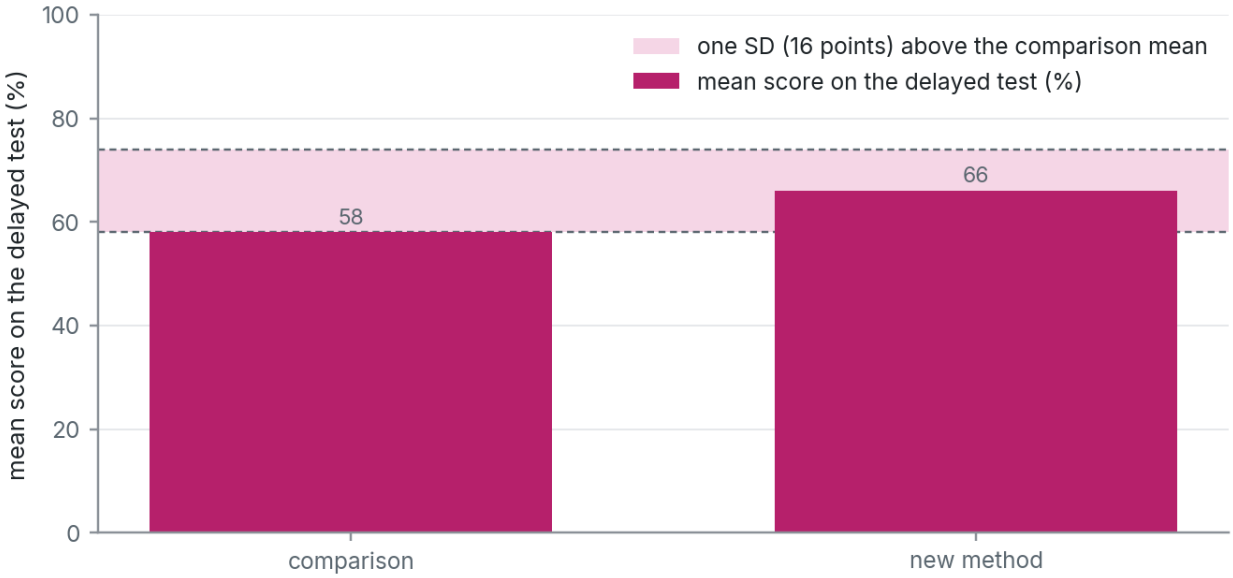
But marks go up after almost any teaching. A rise tells you something happened. It cannot tell you that your method did it.

So you need two groups, and they must differ in one way only: the method. Then you decide who goes in which group by chance, for example by drawing roll numbers from a box.

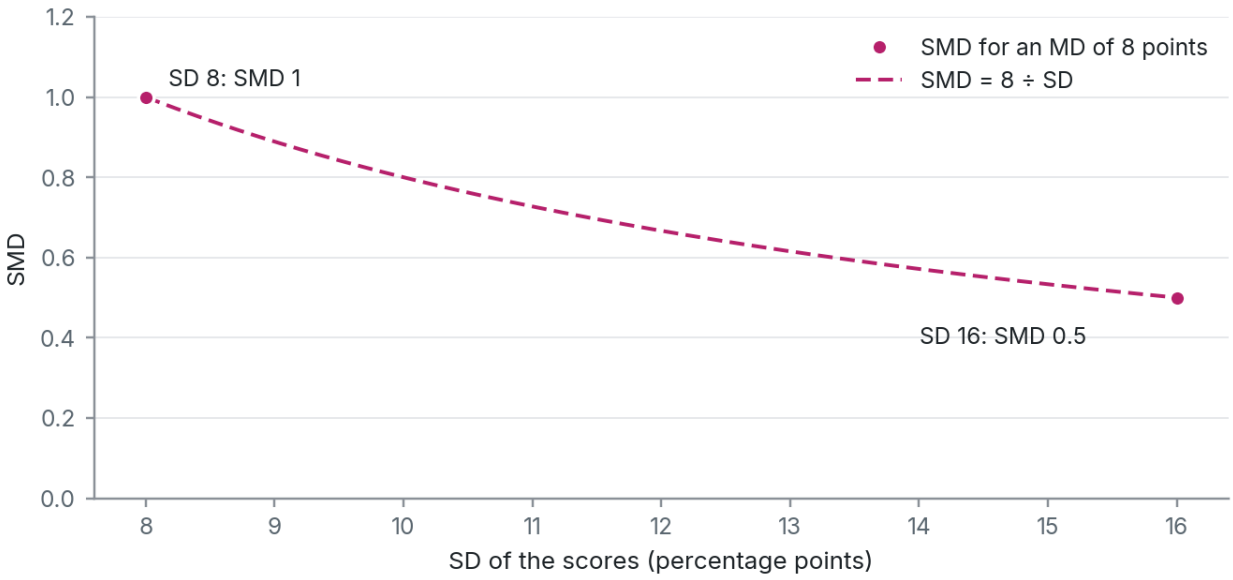
Now compare the two groups' average marks.

A gap of 8 marks can be large or small. It depends on how spread out the marks are.

So divide the gap by the standard deviation.



Made-up numbers. Mean scores on the delayed test: 58% for the comparison group and 66% for the new method's group, an MD of 8 percentage points. The shaded band runs from 58 to 74, one SD of 16 above the comparison mean. The new method's bar reaches halfway into it: an SMD of 0.5.



Made-up numbers. The same MD of 8 percentage points, divided by different SDs: SMD = 8 ÷ SD. With an SD of 16 the SMD is 0.5; with an SD of 8 it is 1. The gap in marks stays the same; the SMD grows as the scores bunch together.

Illustration 1. Start with the attempt that fails. The numbers in it are made up.

You teach a new way of running the nutrition sessions to a batch of 150 third-year MBBS students. Before the first session, they sit a 20-question test. At the end of the last session, they sit the same 20 questions again. The batch's mean score rises from 45% to 70%.

```
70 minus 45 = 25
```

A rise of 25 percentage points. A colleague asks whether the new way works. Before you answer, list what else could have raised that score.

- Any teaching at all. The old way would also have taught them something.
- The same 20 questions twice. The students had practised this exact test.
- The timing. The second test came at the end of a session, which measures performance, as the last section showed. It says nothing about next month.

Each of these would raise the score with no help from your method. With one group, you cannot tell how much of the 25 points belongs to which. The rise is real. It is not evidence about the method.

Now build the fair version. Put all 150 roll numbers in a box. Draw 75 of them. That group gets the new way. The other 75 get the usual way, the comparison group. Give both groups the same topic, the same number of minutes, and the same teacher if you can. Two weeks after the last session, give both groups one new test on the same objectives.

Does drawing from a box really give you two similar groups? Open Deslauriers and colleagues, "Measuring actual learning versus feeling of learning", *Proceedings of the National Academy of Sciences* (PNAS) 2019, at https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/. Go to Table 1. It describes two groups of physics students who were allocated at random. Later, each group had one topic as a lecture and one in which students solved problems in small groups. Here you look only at how alike chance made the groups before that. Find the row for each group's average score on two earlier exams.

semester	group A mean (%)	group A SD	group B mean (%)	group B SD
spring	77.9	13.4	78.1	14.4
fall	70.5	13.0	72.02	13.9

Work out each semester's gap, group B minus group A.

```
78.1 minus 77.9 = 0.2
72.02 minus 70.5 = 1.52
```

In spring the two groups sit 0.2 percentage points apart. In fall they sit 1.52 apart. Neither group was chosen for being stronger, and they came out nearly level. Nearly, not exactly: chance leaves small gaps, and a real study has to show they are small. That is what this table is for.

Figures quoted in this illustration are from [51].

WHERE THIS PICTURE BREAKS. A box of roll numbers balances a large group better than a small one. With 20 learners, chance alone can leave one group clearly stronger. Check the groups' earlier marks, as Table 1 does, before you trust the split.

The table shows balance on earlier exam marks only. It cannot show balance on things nobody measured.

Illustration 2. Now run the fair comparison and measure the difference. These numbers are made up too.

Two weeks after the sessions, both groups of 75 sit the same new test. Here are the results.

group	mean score on the delayed test (%)	SD (percentage points)
comparison	58	16
new method	66	16

First, the mean difference, the new method's group minus the comparison group.

```
66 minus 58 = 8
```

The MD is 8 percentage points in favour of the new method. Its sign is positive, so the new method's group did better.

Next, the size of that gap against the spread of the scores. Both groups have an SD of 16 percentage points, so use 16.

```
8 divided by 16 = 0.5
```

The SMD is 0.5. The new method's group scored half a standard deviation higher. To see what one SD is here, add it to the comparison group's mean.

```
58 plus 16 = 74
```

One SD above the comparison mean is 74%. The new method's mean, 66%, sits halfway there.

Keep the MD at 8 and change only the spread. Say both SDs had been 8 percentage points.

```
8 divided by 8 = 1
```

Now the same 8 points is a whole standard deviation. The gap in marks has not changed. What changed is how unusual that gap is, compared with how much learners differ from one another.

Now read a published SMD the same way. Open Freeman and colleagues, "Active learning increases student performance in science, engineering, and mathematics", *PNAS* 2014, at https://www.pnas.org/doi/10.1073/pnas.1319030111. The authors define active learning: it "engages students in the process of learning through activities and/or discussion in class, as opposed to passively listening to an expert."

They combined the results of 158 comparisons between active learning and lecturing into one average. The scores were on exams, concept inventories and similar tests. (A concept inventory is a test mostly designed to find known misconceptions.) The overall SMD is 0.47. It was larger on concept inventories than on course exams. The authors put it in words: performance rose "by just under half a SD". Turn it back into marks for a batch whose scores have an SD of 16 percentage points.

```
0.47 times 16 = 7.52
```

About 7.5 percentage points. That holds only for a test with that spread.

Last, go to Freeman's Table 1. It splits the studies by how the students were put into groups. In 16 studies, a chance process decided, or every student had both methods in turn. Their SMD was 0.514. In other studies, students chose their own class section, and the SMDs ranged from 0.362 to 0.534. The authors report that the randomised studies gave results no different from the others.

Figures quoted in this illustration are from [52].

WHERE THIS PICTURE BREAKS. The made-up batch gives both groups the same SD, so there is only one SD to divide by. Real groups rarely match exactly. A published SMD may divide by an SD pooled from both groups, or by the comparison group's SD alone (Glass's delta). Some papers name their version, such as Cohen's d. Each uses an exact formula this book does not work through. Your hand division gives the idea, not the published figure.

Freeman's studies were in university science courses in their own time and place. The timing of the tests varied from study to study. So 0.47 is not a promise about a delayed test in your college.

MUST KNOW POINTS FOR YOU.

- A batch's rise from a pre-test to a post-test is not evidence that your method works. Call it the change in this batch, and run a comparison before you claim more.
- To compare two ways of teaching, let chance decide who gets which. Then check the two groups' earlier marks, to see how evenly chance split them.
- Never quote an MD without saying which group was subtracted from which, and in what unit. A positive 6 and a negative 6 are opposite findings.
- An SMD of 0.5 does not mean "50% better". To turn it into marks, multiply it by the SD of your own test's scores. That assumes the effect would carry over to your test and learners, which the SMD cannot tell you. Report the MD in marks first, then the SMD with the SD you divided by.
- Freeman and colleagues combined 158 comparisons of active learning against lecturing into one SMD of 0.47, just under half an SD. It was larger on concept inventories than on course exams. Use that as your yardstick when someone offers you a new method's "effect size". It came from volunteer teachers in university science courses.
- The SMD shrinks when learners' scores are widely spread and grows when they are bunched, even if the gap in marks is the same.
- Random allocation makes the groups alike on average, not identical. With a small batch, chance can still leave one group stronger: the spread of a group's mean shrinks only with the square root of its size (Book 0, D6). How many learners a comparison needs is outside this book.
- When a committee is shown an "effect size" for a teaching method, ask three things. Compared with what? Who decided who got which? How long after the teaching was the test?

EXERCISE 1 (design). An accredited social health activist (ASHA) is a village health worker. Your department has a question about training them. Does a session built on short written ques-

tions teach them to read a growth chart better than the usual talk? A growth chart is the printed curve on which a child's weight is plotted against age. Sixty ASHA workers from one block will attend.

Plan the comparison. Say how you split them and what you keep the same. Say what you measure and when. Name the one number you will report first.

EXERCISE 2 (teaching). A journalist rings. A company says its training app has "an effect size of 0.8, the biggest in medical education". The company and its claim are made up for this exercise. You have three minutes and no jargon. Explain what that number means and what to ask the company.

PRACTICE. Twelve problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. Group P has a mean of 72 and group Q a mean of 63. The SD of the scores is 18 in both groups. Find the MD, P minus Q, and the SMD.

2. A new method's group has a mean of 47. The comparison group has a mean of 53. The SD is 12 in both groups. Find the MD, new minus comparison, and the SMD. Say what the sign tells you.

3. A study reports an SMD of 0.25. The SD of the scores is 20. What is the MD?

4. Two groups' means differ by 9, and the SMD is 0.6. What SD was used?

5. In Roediger and Karpicke's second experiment, the SSSS group read a passage in four study periods. The STTT group read it once and then took three recall tests. After 5 minutes, SSSS recalled 83% of the passage and STTT recalled 71%. After 1 week, STTT recalled 61% and SSSS recalled 40%. Find the MD, STTT minus SSSS, at each retention interval, with its unit. Say what the two signs tell you.

6. Deslauriers and colleagues allocated physics students to two groups at random. Before the experiment, the groups' averages on two earlier exams were as follows. Spring: group A 77.9% (SD 13.4 percentage points), group B 78.1% (SD 14.4). Fall: group A 70.5% (SD 13.0), group B 72.02% (SD 13.9). For each semester, find the MD, B minus A, and the SMD, dividing by group A's SD. Say what the two SMDs tell you about the random allocation.

7. Roediger and Karpicke report d, an SMD, for two of their comparisons. After 5 minutes, SSSS beat STTT by 12 percentage points, with d = 1.22. After 1 week, STTT beat SSSS by 21 percentage points, with d = 1.26. The paper does not give the SDs. Taking d as the MD divided by the SD, find the SD each d implies. Why is the second MD nearly twice the first, while d hardly moves?

8. Here is a worked answer. The numbers are made up. Find the step that broke.

```
comparison group mean 58%, new method group mean 66%, SD 16 percentage points in
each group
MD = 66 minus 58 = 8 percentage points
SMD = 8 divided by 58 = 0.14
so the new method's group scored 0.14 SD higher
```

9. Here is a worked answer. The numbers are made up. Find the step that broke.

```
usual method group mean 64%, new method group mean 58%, SD 12 percentage points in
each group
MD = 64 minus 58 = 6 percentage points
SMD = 6 divided by 12 = 0.5
so the new method is half an SD better
```

10. Here is a worked answer. The numbers are made up. Find the step that broke.

```
ASHA workers each plot ten children's weights on a growth chart; the score is the
number of mistakes
question-based group: mean 2.1 mistakes, SD 1.4; usual talk group: mean 3.5 mis-
takes, SD 1.4
MD = 2.1 minus 3.5 = -1.4 mistakes
SMD = -1.4 divided by 1.4 = -1
the SMD is negative, so the question-based session did worse, by one SD
```

11. A speaker at a medical-education meeting says: "Active learning raises exam scores by 6%. Freeman proved it." Freeman's own words are that a change of 0.47 SDs "would produce an increase of about 6% in average examination scores".

Suppose your course's exam marks have a mean of 77.9% and an SD of 13.4 percentage points. Those are the figures for one group in Deslauriers's Table 1. Decide what to compute, compute it, and say what your answer does not establish.

12. A colleague writes this in the department's annual report. The words are made up for this problem.

"My case-based sessions raised the batch's mean score from 52% on the pre-test to 68% on the post-test. The pre-test SD was 16 percentage points. So that is an effect size of 1.0, more than double Freeman's 0.47."

Decide what to compute, compute it, and say what the claim does not establish.

3 · Retrieval practice: pulling it out beats putting it in again

Checked against the sources as they stood on 2026-09-25.

DEFINITION. Retrieval practice is an attempt by a learner, during learning, to bring material they have already met back to mind from memory: to recall it or to recognise it, without looking at it. Restudy is the alternative it is compared with: reading or hearing the same material again, for the same time.

The testing effect is the finding that a retrieval attempt improves memory on a later test more than restudy does. It holds on a delayed test. On a test a few minutes later, restudy can come out ahead.

Roediger and Karpicke's (2006) first experiment shows both halves. Each student reread one passage and tried to recall the other, with no feedback.

In classrooms, Agarwal, Nunes and Blunt (2021) screened nearly 2,000 abstracts and kept 50 experiments in which each student retrieved alone, closed-book, during class time. From those 50 they computed 49 effect sizes.

Feedback is showing the learner the correct answer after the attempt. Dunlosky and colleagues (2013) call retrieval practice "practice testing". In their review, practice testing with feedback beats restudy more reliably than practice testing without it. A low-stakes test is one whose score counts for little or nothing.

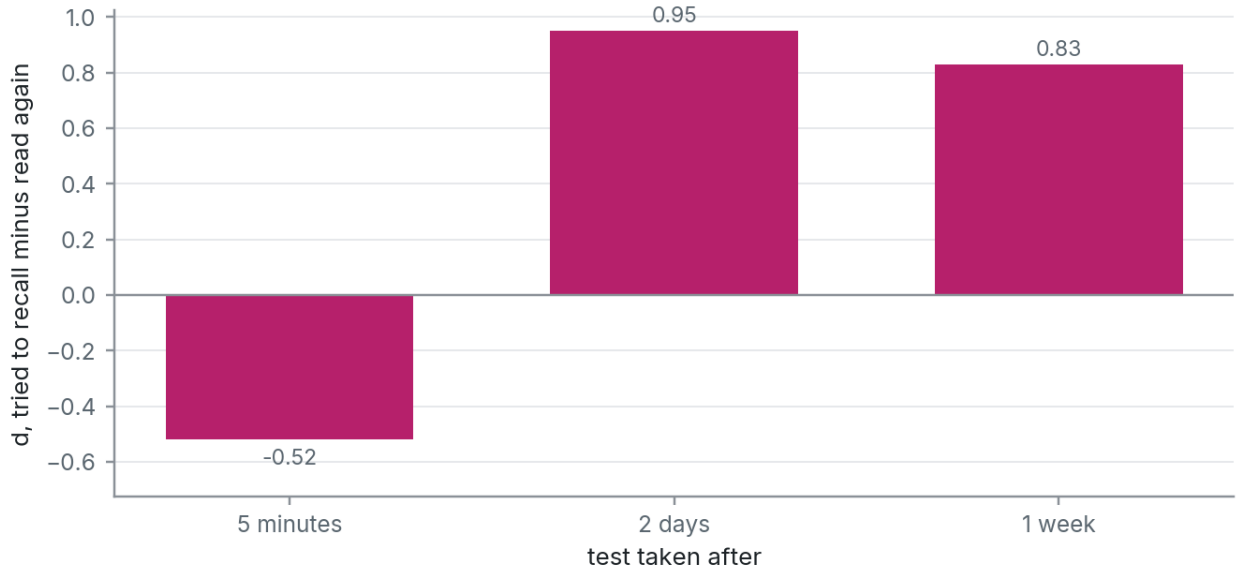
What would overturn this for your setting: well-run comparisons in Indian classrooms, with a delayed test on new questions, that find no gain. [53, 54, 55, 56, 57, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 72, 73, 74, 75, 76, 77, 78]

IN PLAIN TERMS. Most of us were taught that a test measures learning. It also causes it. Pulling something out of memory makes it easier to pull out again later. Reading it again does much less, even though it feels like more.

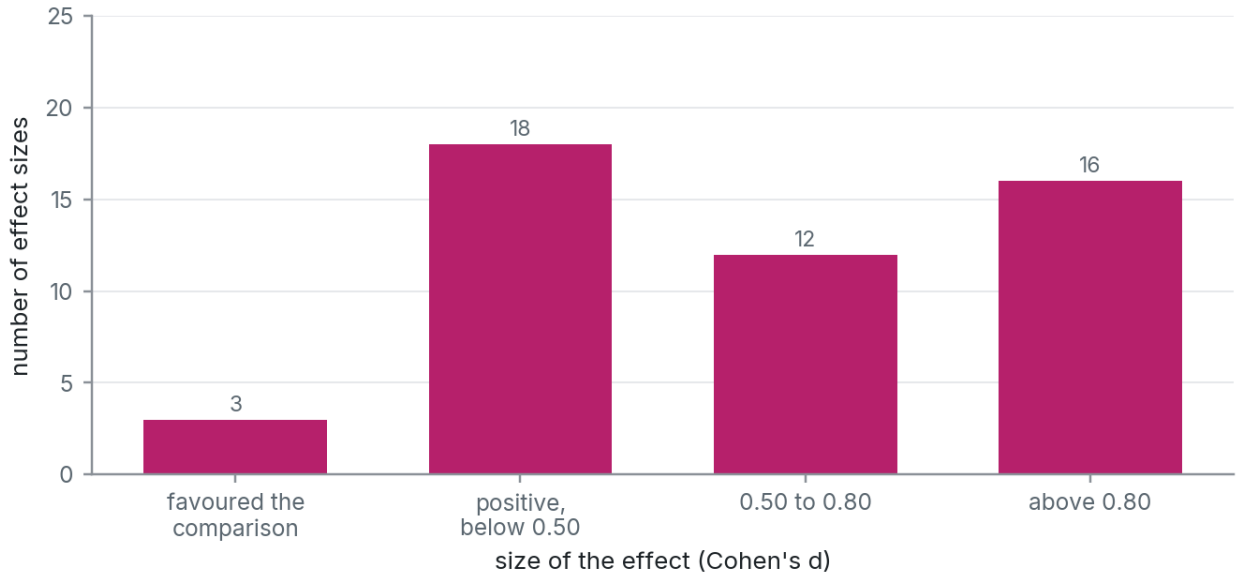
The catch is timing. Straight after a session, what was reread looks better.

In class, use three conditions. Every learner answers alone, without notes; that is how the classroom studies below were run. The score counts for little or nothing, so the test is practice and not judgement. After they have tried, they see the right answer. Of the three, feedback has the clearest evidence behind it.

The classroom evidence is large and mostly from the United States and Europe. Ten of the experiments were in medical schools. None was in India.



Roediger and Karpicke (2006), first experiment: the difference between the passage students tried to recall and the one they read again, in standard deviations (d), with the sign of the direction. At 5 minutes it is -0.52, rereading ahead. At 2 days it is 0.95 and at 1 week 0.83, recall ahead.



Agarwal, Nunes and Blunt (2021): the 49 effect sizes from classroom experiments on retrieval practice, counted by size. 3 favoured the comparison, 18 were small benefits, below 0.50, 12 were between 0.50 and 0.80, and 16 were above 0.80. The comparison was re-reading in 19 experiments, lessons without quizzes in 14, questions that were never quizzed in 12, and fewer quizzes in 5. The bars are counts, not one pooled result.

Illustration 1. In section 1, Roediger and Karpicke's second experiment showed that a check in the last five minutes can pick the wrong method. Their first experiment answers a different question: how big is the benefit of trying to recall, in the unit section 2 gave you for comparing results?

Open Roediger and Karpicke, "Test-enhanced learning: taking memory tests improves long-term retention", *Psychological Science* 2006, at https://doi.org/10.1111/j.1467-9280.2006.01693.x. Each stu-

dent read two short passages. One they read again. For the other, they wrote down all they could remember, and nobody told them what they got wrong. Search the text for 81%.

> *restudying produced better recall than testing (81% vs. 75%)*

Beside it the paper prints d = 0.52. Read on to the 2-day and 1-week results.

> *the initially tested group recalled more than the additional-study group (68% vs. 54%)*

> *The tested group recalled 56% of the material, whereas the restudy group recalled only 42%*

Beside these it prints d = 0.95 and d = 0.83. The d is a standardised mean difference: the gap between the two means, in standard deviations. The paper gives each d as a size, with no sign. Give it the sign of the direction, tried to recall minus read again, as section 2 taught you to do for a mean difference. The gap in percentage points (Book 0, A4) goes beside it.

```
75 minus 81 = -6
68 minus 54 = 14
56 minus 42 = 14
```

test taken after	read it again (%)	tried to recall it (%)	gap (points)	d, recall minus reread
5 minutes	81	75	-6	-0.52
2 days	54	68	14	0.95
1 week	42	56	14	0.83

Two things to see. First, the benefit of trying to recall is not small once there is a delay: close to one standard deviation at 2 days. Second, the same 14-point gap gives two different d values. A d divides the gap by a spread. The paper does not report that spread, but it must have differed between the two delays. That is why you quote the gap in points beside the d.

Figures quoted in this illustration are from [14].

WHERE THIS PICTURE BREAKS. These were undergraduates at one university reading two short passages, one on the Sun and one on sea otters, tested by writing out what they recalled. Each student had one passage each way. The numbers give the direction and a rough size in that setting, not what your batch will score.

The d values come from one small laboratory experiment. The classroom evidence, next, is the better guide to size.

Illustration 2. A laboratory is not a lecture hall. So look at the classroom evidence, and at where it was gathered.

Open Agarwal, Nunes and Blunt, "Retrieval practice consistently benefits student learning", *Educational Psychology Review* 2021, at https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf.

The authors kept only experiments where every student retrieved alone, without notes, in class time. Search for 5374.

> *Our review yielded 49 effect sizes and a total n = 5374, the majority of which (57%) revealed medium or large benefits from retrieval practice.*

Each effect size is a Cohen's d, the same kind of number as the d values above. Search for 0.80 and read the next two sentences.

> *Overall, 16 effect sizes indicated large benefits from retrieval practice (d > 0.80), 12 indicated medium benefits (0.50 < d < 0.80), and 18 were small or very small (d < 0.50). Only three out of 49 effect sizes revealed a negative effect*

Check that the four groups account for every effect size, then turn the paper's 57% back into a fraction.

```
16 plus 12 plus 18 plus 3 = 49
16 plus 12 plus 18 = 46
16 plus 12 = 28
28 divided by 49 = 0.571428571
```

So 46 of the 49 favour retrieval practice, as the paper says elsewhere. The 18 small ones are small benefits, not losses.

Now search for Pakistan. It is in Table 2, under Location.

where the experiments were run	experiments
United States, Denmark, Netherlands, Sweden	47
Pakistan, Taiwan, Turkey	3

The paper calls the first four countries western, educated, industrialized, rich and democratic (WEIRD). Those seven countries are every country in the review. India is not one of them. Search once more, for medical schools.

> *and 10 experiments were conducted in medical schools*

Ten of 50 experiments were run on students like yours, and none of the 50 in your country. That is where your own measurement comes in (section 2).

Last, search for without feedback.

> *Studies without feedback resulted in mostly small or very small effects on learning.*

Figures quoted in this illustration are from [79].

WHERE THIS PICTURE BREAKS. The 49 effect sizes are not one experiment repeated 49 times. The designs, subjects and delays differ, and the authors chose not to average them. Read "28 of 49 above 0.50" as a count, not as the size of the effect you should expect.

The classroom effects were smaller when the final test came at the end of a semester than when it came one to three days after the last quiz. A quiz this week does not by itself carry to an examination months away. The next section is about that.

Illustration 3. Now build the last ten minutes of your own session with the three conditions in it. Say the session taught a batch how to read the nutrition panel on a packet of biscuits.

Ask them to put their notes and phones away. Each student, alone, writes down every step of reading the panel they can remember. No talking to a neighbour. Weinstein and colleagues describe exactly this move: "asking students to put their class materials away and try to write out everything they know about a topic".

Tell them nobody will collect the sheets. Then show the steps on the board and let each student mark their own sheet.

Each choice has a reason. Alone: this is what the classroom evidence covers. Agarwal's review left out pair work, because in a pair one student may do the remembering while the other watches. The one pair study it describes did better than students working alone. So "alone" is the safe choice, not a proven one. Low-stakes: Weinstein reports a study where students quizzed under high pressure did worse on a later test than students quizzed under low pressure. The answer at the end: practice testing with feedback beats restudy more reliably than testing alone.

Next week, open the class with three short questions on this week's panel before you teach anything new. Then you have a second retrieval, on a delay.

The same design works outside a lecture hall. At a monthly meeting of accredited social health activist (ASHA) workers, each person answers aloud in turn, before the answer is given.

WHERE THIS PICTURE BREAKS. If almost nobody can answer, the attempt teaches little on its own. Weinstein's review says retrieval that is "extremely low" in success is unlikely to help, "particularly in the absence of feedback". Give the answer every time, and teach again what most of the room missed.

Answering aloud in turn is not the same as each person writing alone. Later speakers hear the earlier answers. The classroom evidence above is for individual retrieval.

MUST KNOW POINTS FOR YOU.

- "A test is for finding out what they know; the teaching happens before it" is the error. The attempt to recall is itself a learning event.
- Run retrieval so that every learner answers alone and without notes, the score counts for little or nothing, and the correct answer follows the attempt.
- Keep quizzes that are for learning low-stakes.
- In classrooms, 28 of 49 effect sizes were above 0.50 and 46 of 49 favoured retrieval practice (Agarwal and colleagues, 2021). The comparison was re-reading in 19 experiments,

lessons without quizzes in 14, questions that were never quizzed in 12, and fewer quizzes in 5.

- None of the 50 classroom experiments in that review (which searched up to January 2018) was run in India. Measure it in your own batch, with a delayed test.
- Retrieval with no feedback and little success is where the method stops paying. Classroom studies without feedback showed mostly small or very small effects. When almost nobody can answer, give the answer and reteach; do not count the silence as practice.
- If your final test repeats the quiz questions word for word, a gain may be memory of those questions. To find out whether they learned the idea, test it with new questions.

EXERCISE 1 (design). Take one hour you will teach in the next month: a batch of MBBS students, a group of residents, or a meeting of ASHA workers. Plan its last ten minutes.

Write down four things. What will each learner try to recall? How will you make sure each one answers alone? How will you keep the stakes low? When will they see the right answer? Then write down when they will recall it a second time, and on what day you will test whether it lasted.

EXERCISE 2 (teaching). A journalist rings you for three minutes. She has read that "tests help students learn" and wants to know whether that is true, and whether it has been shown in India. Say what you would tell her, including one quotable sentence.

Record your confidence as a percentage before turning to the answer.

4 · Spacing: the same practice, spread out, and how far apart

Checked against the sources as they stood on 2026-09-25.

DEFINITION. Spacing, which Dunlosky and colleagues (2013) call distributed practice, is a schedule: the same material is studied or retrieved on two or more occasions spread out in time. Massed practice, often called cramming, puts the same occasions back to back.

The gap is the time between two study sessions of the same material. The retention interval, from section 1, is the time from the last study session to the test.

The spacing effect is the finding that, for the same total study time, spaced practice gives better long-term retention than massed practice.

Cepeda and colleagues (2008) taught 1,354 people 32 trivia facts. Their numbers here are from the authors' manuscript. Final recall first rose as the gap grew, then fell. Of the gaps tested, the best were 1, 11, 21 and 21 days for retention intervals of 7, 35, 70 and 350 days.

As shares of the retention interval, those best gaps were 14%, 31%, 30% and 6%. They are rough, because only six or seven gaps were tried. The abstract rounds them to about 20% for a test a few weeks away and about 5% for a year. The authors' fitted curve smooths them into a best gap that grows while its share falls. This book uses 20% to 30% as a planning window for tests a few weeks away.

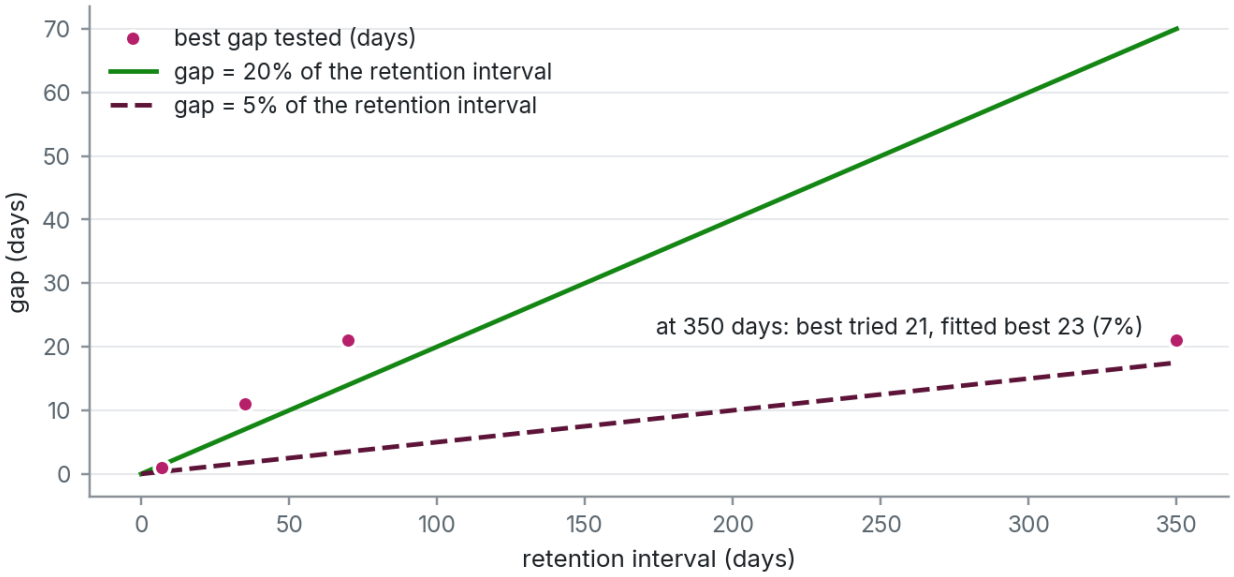
Overall, recall at the best gap was 64% higher than with no gap: a relative increase, not percentage points. By retention interval it ran from 10% for a test a week away to 111% for 70 days. A gap longer than the best cost less than a gap shorter than the best. [80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115]

IN PLAIN TERMS. Take the same three hours of study. Spend them in one evening, or spread them over three weeks. The spread-out hours leave more behind a month later.

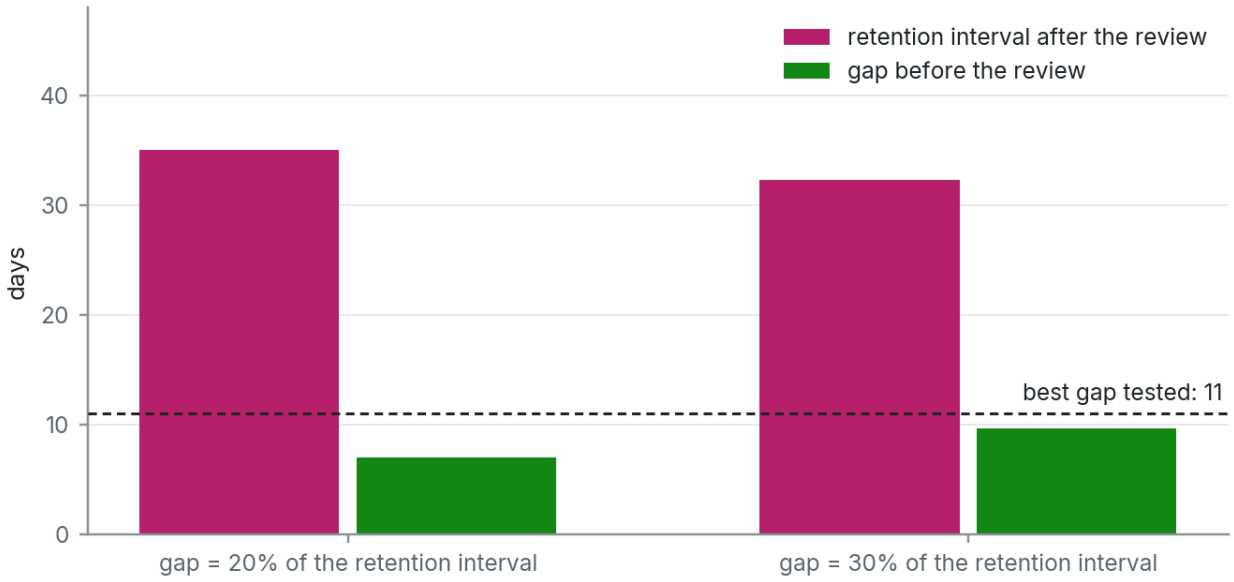
It does not feel that way. Straight after a cramming session, you remember a lot, and on a test that day cramming can win.

Say the test is a few weeks away. This book suggests a gap before the review about 20% to 30% as long as the wait from the review to the test. That window comes from one large study of facts. Treat it as a starting point, and if you are unsure, err on the long side.

Make the review a retrieval, not a reread (section 3): close the book, write what you remember, then check.



Cepeda and colleagues (2008), recall: the best of the gaps tested was 1, 11, 21 and 21 days for retention intervals of 7, 35, 70 and 350 days. The two straight lines are a gap of 20% and of 5% of the retention interval. The points for 35 and 70 days sit above the 20% line, the 7-day point just below it, and the 350-day point near the 5% line. The points are the best of the few gaps tried, so nothing is drawn between them.



The worked calendar example: 42 days from teaching to test, split into the gap before the review and the retention interval after it. A gap of 20% of the retention interval puts the review on day 7, leaving 35 days; a gap of 30% puts it on day 9.69, leaving 32.31. Each pair adds up to the whole 42. The dashed line is the best gap Cepeda tested for a 35-day retention interval, 11 days.

Illustration 1. Here is the plan most of us were taught by. Cover a topic in one week of the course. Revise it all in the week before the examination. Every revision session goes well, because you have only just seen the material. That good feeling is the trap.

Open Dunlosky and colleagues, "Improving students' learning with effective learning techniques", *Psychological Science in the Public Interest* 2013, at https://doi.org/10.1177/1529100612453266, and go to section 9, Distributed practice. It describes an old experi-

ment. Students learned Spanish words, then relearned them in six more sessions. The sessions were back to back, 1 day apart, or 30 days apart. With 30 days between sessions, students forgot the most in between, and they never caught up with the others during practice. Search for 30 days later.

> *the pattern reversed on the final test 30 days later, such that the best retention of the translations was observed in the condition in which relearning sessions had been separated by 30 days.*

The schedule that looked worst during practice was best a month later. Read on.

> *the condition with the most intersession forgetting yielded the greatest long-term retention.*

Why spacing works is debated. One account is that a review soon after learning is too easy to do much work. Cepeda and colleagues doubt it fits gaps of days and weeks.

So how far apart should the sessions be? Open Cepeda and colleagues, "Spacing effects in learning: a temporal ridgeline of optimal retention", *Psychological Science* 2008. The copy free to read is the authors' own manuscript, at https://files.eric.ed.gov/fulltext/ED505660.pdf. Its first line warns that it "may differ from the final published version".

Each learner studied the facts, waited a gap, went through them once more by answering each question and then seeing the answer, waited a retention interval, and took a final test. RI is the paper's short form for retention interval. Search for respectively.

> *For the RIs of 7, 35, 70, and 350 days, the optimal gaps (of those included in the study) were 1, 11, 21, and 21 days, respectively, for recall data*

Divide each best gap by its retention interval. That is a ratio of two times in the same unit (Book 0, A5), and 100 times it is a percentage (Book 0, A4).

```
1 divided by 7 = 0.1428571429
11 divided by 35 = 0.3142857143
21 divided by 70 = 0.3
21 divided by 350 = 0.06
```

Multiplied by 100 and rounded: 14%, 31%, 30% and 6%.

retention interval (days)	best gap tested (days)	gap as % of retention interval
7	1	14
35	11	31
70	21	30
350	21	6

For a test a year away, the best gap tested was 6% of the retention interval, not 30%.

The authors also fitted one smooth surface to all 26 conditions. Search for 23 days.

> *the optimal gap for RI = 350 equals 23 days, which is just 7% of the RI.*

On that surface the best gap grows as the test moves away, while its share falls. That shape was one of the constraints the authors built the surface to meet.

Their abstract rounds the shares off: about 20% for a test a few weeks away, falling to about 5%, or 0.05, for a test a year away. Last, search for too small a gap.

> *while there are costs to using a gap that is longer than the optimal value, these costs are much smaller than the costs of using too small a gap value*

Too long is the cheaper mistake.

Figures quoted in this illustration are from [116, 117].

WHERE THIS PICTURE BREAKS. The best gaps in the table are the best of a few tried: six or seven gaps at each retention interval. The real best can sit anywhere between two tried gaps. That is why the share jumps from 14% to 31% and back down to 6%. Read the shares as rough, not exact.

These were adults of many ages and countries, learning trivia facts online, with one review, tested up to 350 days later. For a test years away, for a skill, or for a large body of connected material, the numbers are a guess carried past the data.

Illustration 2. Now put it on a calendar. On Monday 5 October 2026 you teach a batch of MBBS (Bachelor of Medicine and Bachelor of Surgery) students how to plot a child's weight on a growth chart. Their test is on Monday 16 November, 42 days later. There is room for one review in class. When should it go?

The 42 days are the gap plus the retention interval. For a test a few weeks away, use this book's planning window: a gap of about 20% to 30% of the retention interval. Work out both ends.

At 20%, the gap is 0.2 times the retention interval. So the whole 42 days is 1 times the retention interval plus 0.2 times it, which is 1.2 times it. Divide to undo the multiplication (Book 0, C2).

```
42 divided by 1.2 = 35
0.2 times 35 = 7
7 plus 35 = 42
```

At 30%, the whole is 1.3 times the retention interval.

```
42 divided by 1.3 = 32.31
0.3 times 32.31 = 9.69
```

The review goes on day 7 to day 10 after teaching. Count forward from 5 October: day 7 is Monday 12 October, and day 10 is Thursday 15 October. If the timetable gives you a choice, take

the later day, because too long is the cheaper mistake. Check it against the data: for a 35-day retention interval, the best gap tested was 11 days. Day 10 sits close to it.

Make the review a retrieval (section 3). Each student, alone, writes the steps of plotting the chart from memory, then you show the answer. Weinstein and colleagues note that spacing costs little if you "set aside a few minutes per class to review information from previous lessons".

Do the same for yourself. You will teach the topic to residents without notes on Monday 30 November. You study it on Monday 2 November: 28 days before.

```
28 divided by 1.2 = 23.3
0.2 times 23.3 = 4.66
28 divided by 1.3 = 21.5
0.3 times 21.5 = 6.45
```

Round 4.66 up to 5, since too long is the cheaper mistake. Review on day 5 or day 6: Saturday 7 or Sunday 8 November. Close the book, write out what you would teach, and only then open it to check.

Figures quoted in this illustration are from [117].

WHERE THIS PICTURE BREAKS. The 20% to 30% is a rough figure for tests a few weeks away. It comes from one study of trivia facts with a single review. It gives you a window for a date, not a measured best day for your topic.

For an examination months or years away, the share is smaller, and the data stop at 350 days. That study did not compare one review with several. Add later reviews if you can, but no figure here says how far apart they should be.

MUST KNOW POINTS FOR YOU.

- "Three hours is three hours, however you spread them" is the error. With study time held equal, spaced practice leaves more behind on a delayed test.
- Never judge a schedule by how the practice sessions go.
- Before you pick a review date, write down when the learning must be used: the test date, or the day you will teach it. Then set the gap from that.
- Do not plan a long interval with a fixed share. Dunlosky and colleagues summarise Cepeda and colleagues' study as a fixed "10 to 20%" of the retention interval. On the authors' fitted curve the share falls: 23 days, 7%, for a test 350 days away (21 days was the best they tried).
- When you are unsure of the best gap, choose the longer one.
- Test often and lightly, so that students space their own study. In one study, when tests came only every 3 weeks, students left their study until just before each test. With daily tests, their study was spread out.
- These gaps were measured for trivia facts, one review and tests up to 350 days away. Past a year, treat any gap figure as a starting guess. Then check it with a delayed test of your own (section 2).

EXERCISE 1 (design). Pick a topic you will teach in the next month, to MBBS students, residents or accredited social health activists. Write down the date you teach it and the date it must be used: a test, a posting or a field visit. Then set the date of one in-class review, show the arithmetic, and say what the learners will do at the review. Finally, set your own review date if you will teach the topic from memory.

EXERCISE 2 (teaching). You have ten minutes and a whiteboard with a first-year resident. She asks: "Why not teach each topic once and do a full revision week before the exam? It works for me." Explain the mechanism and the evidence, and what you would change in her plan.

PRACTICE. Ten problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. Write each ratio as a percentage, to the nearest whole number: 11 divided by 35, then 21 divided by 350, then 23 divided by 350.

2. Find 20% of 45. Then find 5% of 360.

3. Split 36 into two parts. The smaller part must be 0.2 times the larger part. Find both parts.

4. Cepeda and colleagues (2008) also gave a multiple-choice recognition test. For retention intervals of 7, 35, 70 and 350 days, the best gaps tested for recognition were 1, 7, 7 and 21 days. Work out each best gap as a percentage of its retention interval, to the nearest whole number. Say which way the percentage moves as the retention interval grows. Then compare the gaps with the recall gaps of 1, 11, 21 and 21 days.

5. You teach a batch of MBBS students a topic on 1 December 2026. Their test is 60 days later. Take the gap as 20% to 30% of the retention interval, this book's planning window for a test some weeks away. Work out the test date, and the range of days after teaching on which the one review should fall, with its dates.

6. The timetable fixes your one review 9 days after you first teach a topic. You want the gap to be 20% to 30% of the retention interval. How many days after the review should the test fall, and how many days after the first teaching is that?

7. Here is a worked answer. Find the step that broke.

```
Cepeda and colleagues: the best gap gave a 64% increase in final recall over no gap
suppose recall with no gap was 40%
40 plus 64 = 104
104% recall is impossible, so the 64% figure must be wrong
```

The 40% is made up for the sake of the sum. The paper's recall scores for each condition are in a figure this book does not hold.

8. Here is a student's worked plan. Find the step that broke.

```
I start the topic on 20 October; the exam is 30 days later, on 19 November
the gap should be 20% of the time
20% of 30 days = 6 days
so I revise 6 days before the exam, on 13 November
```

9. A colleague plans reviews for a batch of MBBS students. “Dunlosky and colleagues say: space reviews at 10 to 20% of how long you need to remember. The final examination is 3 years away. So each topic gets its review 3.5 to 7 months after it is taught.” Decide what to compute, compute it, and say what your answer does not establish.

10. A coaching advertisement says: “Research shows spacing your study raises what you remember by 64%. Our spaced timetable will raise your university examination marks by 64%.” Decide what to compute, compute it, and say what your answer does not establish.

5 · What does not work: rereading and highlighting

Checked against the sources as they stood on 2026-09-25.

DEFINITION. Rereading is studying a text again after a first reading: the restudy of section 3, done with a text. Highlighting is marking the parts of a text that seem important while reading it.

Dunlosky and colleagues reviewed ten study techniques and gave each an overall utility rating: low, moderate or high. The rating asks two things. How widely does the benefit hold, across learners, kinds of material, kinds of final test and real classrooms? And how does the technique compare with the others? They rated both rereading and highlighting low utility. Practice testing, the retrieval practice of section 3, and distributed practice, the spacing of section 4, were rated high.

Highlighting, in most situations studied and with most learners, did little to boost performance. It may help when learners know enough to pick out what matters, or when the text is hard.

Rereading does improve recall over a single reading. But in direct comparisons, rereading has consistently done worse than practice testing and than two techniques of explaining: explaining why a stated fact is true, and explaining how new material connects to what you already know.

So low utility is not zero effect. It means the same time spent another way does more.

Fluency is the ease with which a text reads or a fact comes to mind. Pashler and colleagues (2008) warn that this ease and familiarity can come from things unrelated to understanding. [118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157]

IN PLAIN TERMS. Ask a batch how they revise, and you will probably hear two things. They read their notes again. They mark the important lines with a highlighter.

Read a page a second time and it goes down easily. That ease is fluency, and it can feel like knowing. But it only tells you the page is familiar.

Read once. Then close the book, write down what you remember, and check it against the page.



Roediger and Karpicke (2006), second experiment, group means. The SSSS group, which only reread, expected to remember best (4.8 out of 7) and recalled least after 1 week (40%). The STTT group, which studied in one period and was tested in three, expected least (4.0) and recalled most (61%). The ratings differ by under 1 point of the 1-to-7 scale; the bottom axis shows only that part of it.

Illustration 1. A student in your batch shows you her textbook. The chapter on nutritional anaemia is yellow from top to bottom. She has been through it twice. What does the yellow tell you?

Open Dunlosky, Rawson, Marsh, Nathan and Willingham, "Improving students' learning with effective learning techniques", *Psychological Science in the Public Interest* 2013, at https://doi.org/10.1177/1529100612453266. Go to section 4, Highlighting and underlining, and search for Fowler and Barker.

Undergraduates read articles of about 8,000 words. One group only read them. A second group highlighted as much as they liked. A third read copies the second group had marked. Every group had 1 hour. A week later, each student had 10 minutes to look over their copy, then took a 54-question multiple-choice test. Search for did not outperform.

> *Overall, the highlighting groups did not outperform the control group on the final test, a result that has unfortunately been echoed in much of the literature*

The control group is the group that only read. Read on. The students who highlighted did better on questions about lines they had marked. They paid for it with a small loss on questions about lines they had not marked. Highlighting moved what they remembered around the text. It did not add to the total.

Now search for marking almost everything.

> *with some students marking almost nothing and others marking almost everything*

A marked line helps because it stands out from the rest of the page. When most of the page is marked, nothing stands out. So her yellow chapter tells you she spent time on it. It does not tell you what she can recall with the book shut.

Figures quoted in this illustration are from [116].

WHERE THIS PICTURE BREAKS. Fowler and Barker's students were undergraduates reading magazine articles, tested with multiple-choice questions a week later. That is one experiment. The low rating rests on many.

The review also reports exceptions. In one classroom study, statements the instructors had underlined in red were answered better on the exam. Highlighting may also help a reader who knows enough to pick out what matters. So "highlighting never helps" goes further than the evidence. "Highlighting as students usually do it adds little" does not.

Illustration 2. Now see how the review rates the techniques side by side. In the same article, go to the Closing Remarks and find Table 4. Before you read a single letter, settle what a row is and what a column is (Book 0, F1).

A row is one technique. The Utility column is the overall rating. The next three columns ask whether the benefit holds for different learners, for different materials, and for different kinds of final test. The review calls the final test the criterion task. The last two ask whether the technique is easy to put into use, and whether it has been shown in real classrooms. Each cell holds a letter. Search for negative (N) rating to read the key.

> *A positive (P) rating indicates that available evidence demonstrates efficacy of a learning technique with respect to a given variable or issue. A negative (N) rating indicates that a technique is largely ineffective for a given variable.*

Q means it worked in some conditions or groups and not in others. I means there is not enough evidence to say. A cell with two letters, such as P-I, is positive on one part of that column and short of evidence on another. Here are three of the ten rows.

technique	utility	learners	materials	criterion tasks	issues for implementation	educational contexts
Highlighting	Low	Q	Q	N	P	N
Rereading	Low	I	P	Q-I	P	I
Practice testing	High	P-I	P	P	P	P

Read the Rereading row first. Materials has a P: a second reading helped across many kinds of text. Issues for implementation has a P: it is easy to do. Yet the rating is Low. Search the rereading section for largest strike.

> *The relative disadvantage of rereading to other techniques is the largest strike against rereading*

So Low is a comparison. Rereading does something. The same minutes spent on practice testing do more. Now read the Highlighting row. It has an N under criterion tasks: across the kinds of final test studied, highlighting did not raise the score.

Last, put both rows beside the Practice testing row, which holds no Q and no N. That is the swap. Take the minutes a learner would spend on a second reading or a highlighter. Spend them on a closed-book attempt at recall, then a check (section 3).

WHERE THIS PICTURE BREAKS. A letter in this table is the reviewers' judgement of a body of studies, not a measurement. Two techniques with the same letters need not have effects of the same size. The table gives you no effect sizes at all.

The ratings are from 2013. For rereading, all but two of the studies the review knew of were of undergraduates. Your own batch is not in any of them.

Illustration 3. section 1 walked you through Roediger and Karpicke's second experiment. Look at it again with one group in mind: SSSS, the students who spent all four periods reading the passage. It is the purest case of rereading in this book.

Open "Test-enhanced learning", *Psychological Science* 2006, at https://doi.org/10.1111/j.1467-9280.2006.01693.x. Search for 14.2. On average, each SSSS student read the passage through 14.2 times across the four periods. Now find Table 2. Before any final test, each student rated how well they would remember the passage in 1 week, from 1 (not very well) to 7 (very well).

group	read it through (mean times)	expected to remember (mean, 1 to 7)	recalled after 1 week (%)
SSSS	14.2	4.8	40
SSST	10.3	4.2	56
STTT	3.4	4.0	61

Read down the last two columns together. The group that reread most was the surest it would remember. It remembered least. The other two groups' ratings did not differ reliably.

Why would reading more make a group surer? Pashler and colleagues give one reason.

Open Pashler, McDaniel, Rohrer and Bjork, "Learning styles: concepts and evidence", *Psychological Science in the Public Interest* 2008, at https://doi.org/10.1111/j.1539-6053.2009.01038.x. Near the end, search for sense of familiarity.

> *As learners, we can also be fooled by subjective impressions, such as the ease or sense of familiarity we gain on reading expository text*

Expository text is text that explains, like a textbook. That ease is what this section calls fluency. The ease is real. Pashler's point is that it can come from things that have nothing to do with understanding.

Figures quoted in this illustration are from [14].

WHERE THIS PICTURE BREAKS. The SSSS group did not only reread more. It also took no test before the final one. So the comparison is rereading instead of recall practice, not rereading against doing nothing.

The ratings are group means on a 7-point scale. They show which group felt surest, not that any one student's feeling was wrong. And on a test 5 minutes after the last period, the SSSS group recalled the most (section 1).

MUST KNOW POINTS FOR YOU.

- "I read it three times and it all looks familiar now" describes fluency, not learning. When a learner says this, ask them to shut the book and say it back.
- Replace the second reading with a test. Teach your batch the order: read once, close the notes, write what you remember, then check it against the notes.
- Do not tell learners that rereading does not work. It is rated low because practice testing does more with the same time.
- A heavily highlighted book is a sign of time spent, not of learning. Check what a learner can recall with the book shut.
- Expect rereading to be your learners' default. At one highly selective university, 84% of students listed it among their study methods. Ask a batch how they revise before you teach them how.
- These ratings come mostly from laboratory studies of undergraduates. By 2013 no experiment had tested rereading's effect on learning in a real course. Later classroom experiments often set rereading against retrieval practice (section 3). Your own batch was in none of them.

EXERCISE 1 (teaching). A journalist writing for a newspaper's education page gives you three minutes. Her question: "Is highlighting a waste of students' time?" Say what you would tell her, including one quotable sentence, and say how sure the evidence lets you be.

Record your confidence as a percentage before turning to the answer.

6 · Learning styles: a popular claim that fails its own test

Checked against the sources as they stood on 2026-09-25.

DEFINITION. "Learning styles" is the idea that people differ in which mode of teaching or study works best for them.

Pashler, McDaniel, Rohrer and Bjork (2008) keep two claims apart. The first is that people will state preferences about how they like to take in new material. The second, the learning-styles hypothesis, is that teaching fitted to a learner's style gives a better learning outcome. Its commonest version is the meshing hypothesis: teaching works best in the format that matches the learner's preference, pictures for a "visual learner", say.

They set four conditions an experiment must meet before it supports the hypothesis. Learners are divided into two or more groups by a measure of style. Within each group, they are allocated at random to one of two or more teaching methods. Everyone takes the same test; if the tests differ, no support is possible. And the method that gives one group its best test score is not the method that gives another group its best. They call that last pattern a crossover interaction, drawn with the style groups along the bottom axis of a graph.

With two styles and two methods, the mean scores make a 2 × 2 table, read "two by two": two rows and two columns. Each row is a style group and each column a method. Within one row, the two cells are two groups allocated at random, so the difference between their means is the mean difference (MD), as in section 2. Take the MD the same way round in both rows, say the first method minus the second. A crossover is present when one row's MD is positive and the other's is negative. The best column then differs from row to row.

A crossover can be present even when one style group scores higher than the other on both methods. Pashler's criterion does not ask that each group's best method be its matched one. Meshing asks that as well. Pashler calls the version without that demand the broad version of the hypothesis: the best method differs between style groups, whichever way round. So a crossover the wrong way round, each group best with the method that does not match its label, supports the broad version and contradicts meshing.

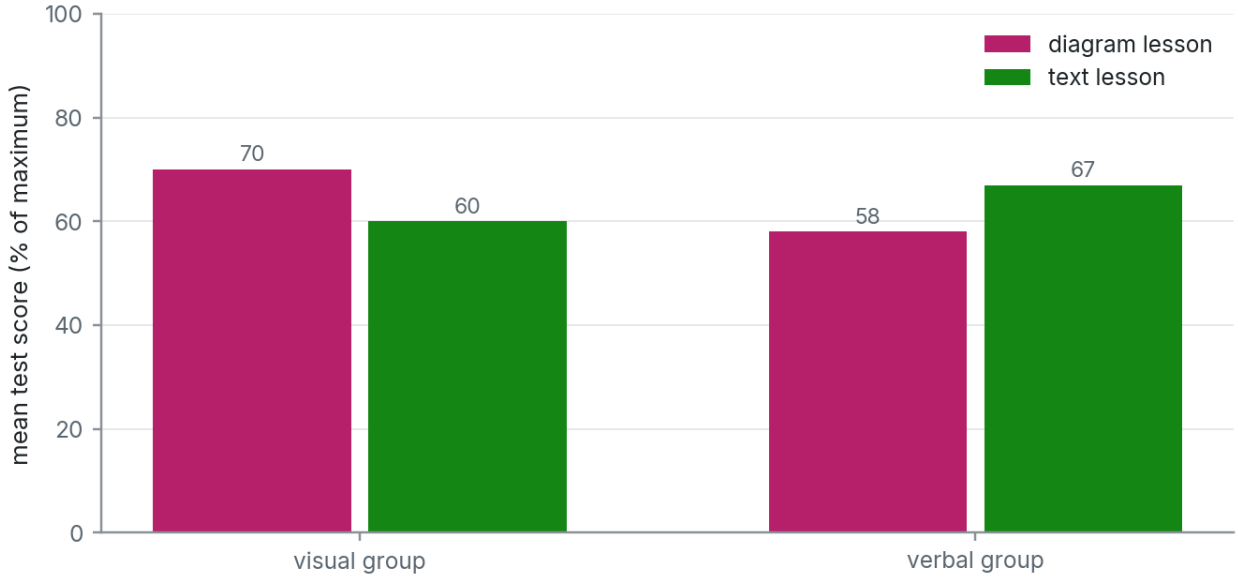
When both MDs have the same sign, the same method is best for every learner, and the result is no evidence for learning styles, however different the two MDs are in size. Read down the columns instead of across the rows, and such a result can look like a crossover.

Studies with the right design found no support, one of them with 123 internal medicine residents. Pashler and colleagues conclude that applying learning styles in classrooms is unwarranted. [158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194]

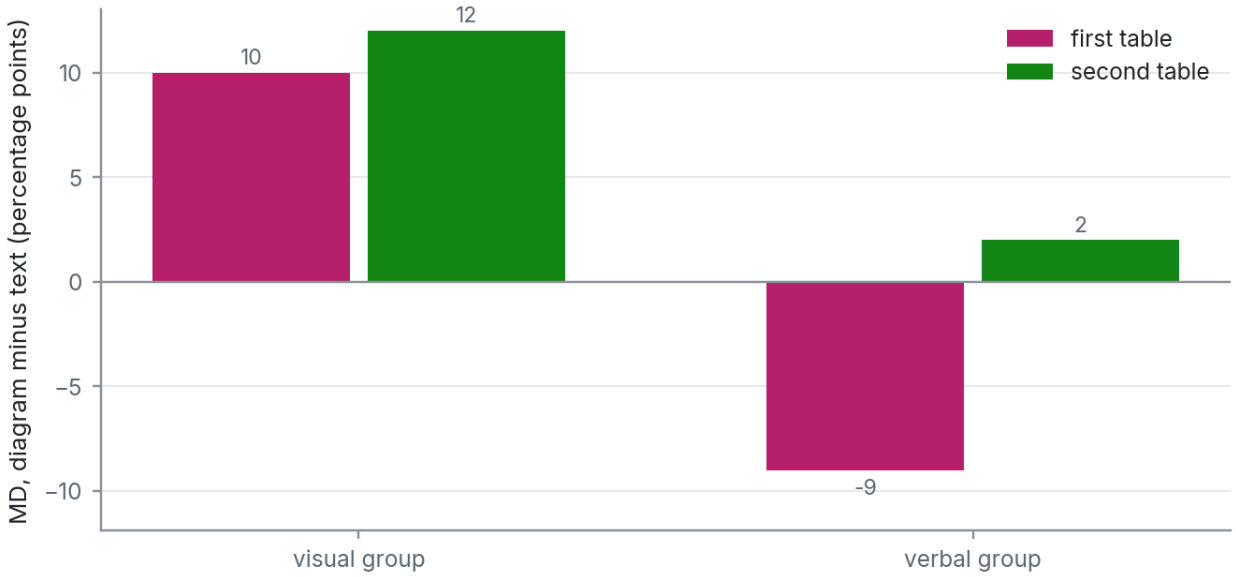
IN PLAIN TERMS. The claim goes like this: some people learn best from pictures, some from words, some by doing, so find out which kind each learner is and teach each one their way. Here is the test it has to pass.

Take learners of two styles. Split each group in two by lottery, and teach one half one way and the other half the other way. Test everyone the same way. The best method has to change from one group to the other. That change is the crossover.

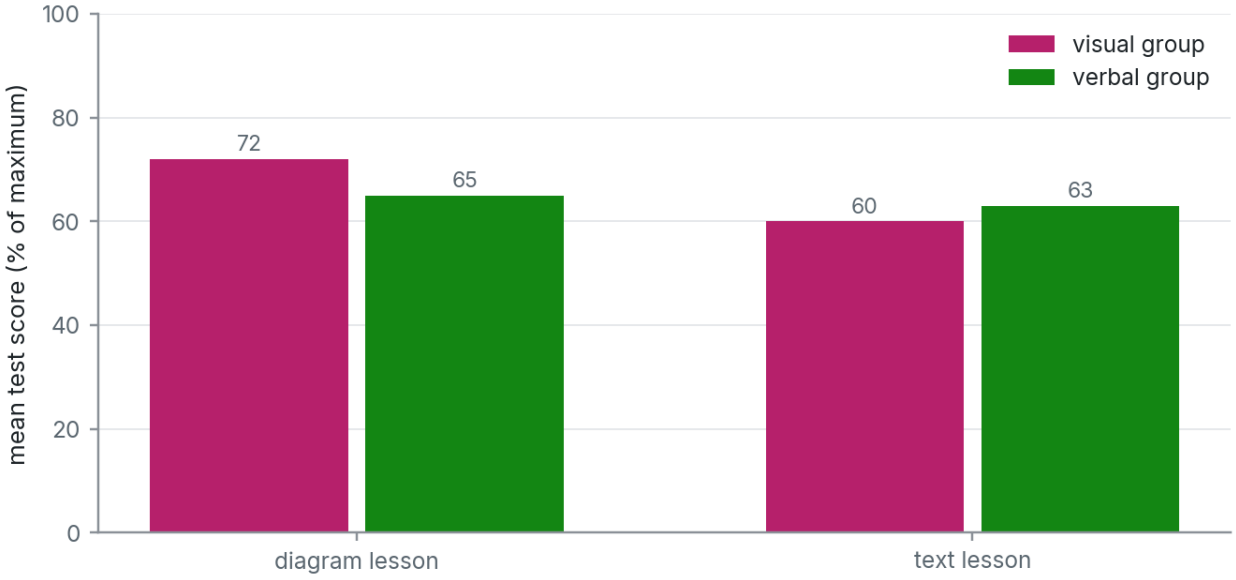
Notice what does not count. Pictures working better for everyone does not count, even if they help visual learners more. Visual learners scoring higher than verbal learners does not count either.



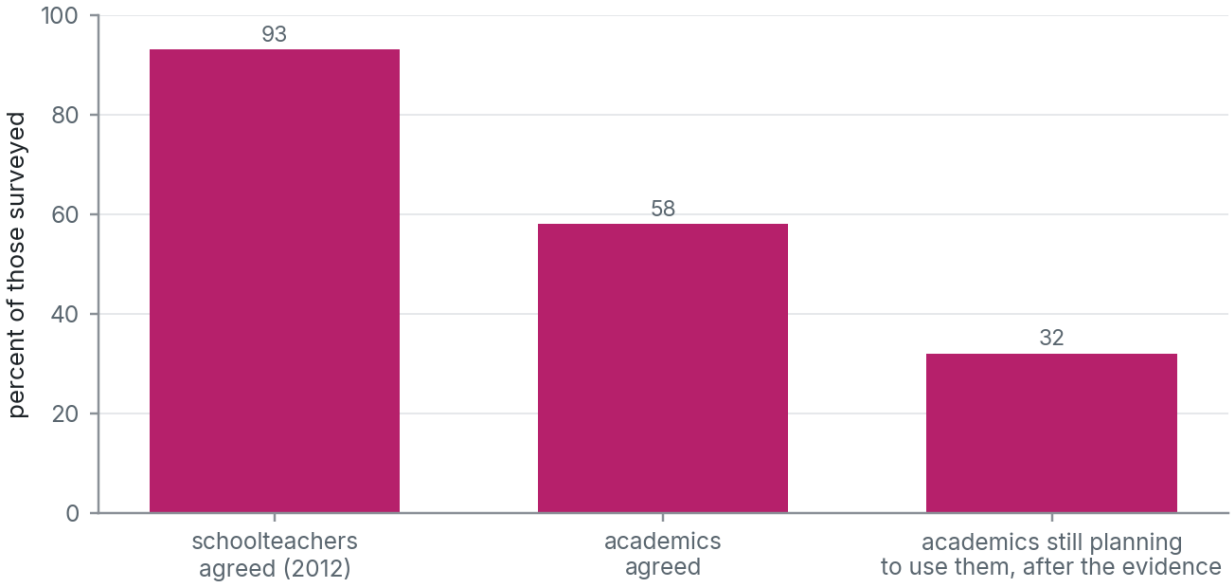
Made-up numbers, not data, drawn with the style groups along the bottom as Pashler's test asks. Diagrams give the visual group its best score, 70% against 60%, and text gives the verbal group its best, 67% against 58%. The best lesson changes with the group: a crossover.



Made-up numbers, not data. The MD, diagram minus text, within each group. In the first table the bars point opposite ways, 10 and -9: a crossover. In the second both point up, 12 and 2: diagrams are best for both groups, and there is no crossover.



The second made-up table, drawn with the lessons along the bottom instead. Visual beats verbal on diagrams, 72 against 65, and verbal beats visual on text, 63 against 60, so the pattern seems to cross. Read within each group and diagrams are best for both: 72 against 60, and 65 against 63.



93% of 137 UK schoolteachers (2012) and 58% of 114 UK academics agreed that people learn better in their preferred style. After being shown that the evidence is missing, 32% said they still planned to account for learning styles in their teaching. The first bar is a different sample from the other two.

Illustration 1. Open Pashler, McDaniel, Rohrer and Bjork, "Learning styles: concepts and evidence", *Psychological Science in the Public Interest* 2008, at https://doi.org/10.1111/j.1539-6053.2009.01038.x. Search for Fourth, the results.

> *Fourth, the results need to show that the learning method that optimizes test performance of one learning-style group is different than the learning method that optimizes the test performance of a second learning-style group.*

That is the whole test. Now put it in a table. The numbers below are made up to show the logic. They come from no study.

Eighty MBBS students fill in a learning-style questionnaire. Forty come out visual and forty verbal. Within each group, a lottery sends half to a lesson on reading a food label taught with diagrams. The other half get the same lesson taught as text. A week later all eighty sit the same test. Here are the mean scores, as a percentage of the maximum.

style group	diagram lesson	text lesson
visual	70%	60%
verbal	58%	67%

Work across each row, not down each column. The question a teacher faces is which lesson to give this learner, so compare the two lessons within one group. Take the MD, diagram minus text, in percentage points.

```
visual: 70 minus 60 = 10
verbal: 58 minus 67 = -9
```

One MD is positive and the other negative. Diagrams are best for the visual group and text is best for the verbal group. The best lesson changes with the group. That is a crossover, and in this made-up table it runs the way meshing predicts.

Now change the bottom row. Same design, different made-up numbers:

style group	diagram lesson	text lesson
visual	72%	60%
verbal	65%	63%

```
visual: 72 minus 60 = 12
verbal: 65 minus 63 = 2
```

Both MDs are positive. Diagrams are best for both groups, so there is no crossover. The difference is bigger for the visual group, but a bigger difference is not a different best lesson. Teach everyone with diagrams.

Now read the second table down its columns, and watch it argue the opposite. On diagrams, visual beats verbal, 72 against 65. On text, verbal beats visual, 63 against 60. “Each group wins on its own method.” Draw it with the lessons along the bottom and the bars seem to cross. Search the review for horizontal axis.

> *the choice of variable for the horizontal axis can affect whether an interaction appears to “cross over.”*

Reading down a column compares two different sets of people, who differ in more than their style. It does not tell you which lesson to give the verbal student in front of you.

WHERE THIS PICTURE BREAKS. Both tables are invented. They show what the test reads, not what any study found. A real table from one study can show a crossover by chance, when both MDs are small beside the spread of the scores. Pashler and colleagues assume every pattern in their figure is statistically significant, which roughly means too large to put down to chance. Telling a real crossover from a chance one needs a statistical test this book does not teach, and a repeat of the study.

Illustration 2. Now see what happened when researchers looked for the crossover. In the same review, search for only one study.

> *we found only one study that could be described as even potentially meeting the criteria described earlier*

That study, by Sternberg and colleagues (1999), gave an ability test to 324 high school students. Search for 112.

> *the authors selected a subset of 112 subjects (35%) for whom one of these three abilities was much higher than the other two*

Check the share yourself.

```
112 divided by 324 = 0.345679012
```

About 35%, as printed. So the reported crossover rests on about a third of the students tested. The reviewers add that it appeared only in heavily processed scores, and only after some students' scores were dropped for reasons not given. They call it tenuous.

Now search for internal medicine.

> *Studying a sample of 123 internal medicine residents and presenting modules on four ambulatory medicine topics, they found no support for this prediction.*

The prediction was that residents of one style would learn better with the problem shown before the content, and residents of another style with the reverse. This study met the design, on doctors in training, and found no support for the predicted crossover.

One more, whose pattern you can now read. Search for Visual presentations produced.

> *Visual presentations produced better free recall than did purely verbal presentations*

In that study, words shown as line drawings were recalled better than words only heard. A person's questionnaire score was not related in any clear way to which way suited them. That is close to the same-sign pattern of the second made-up table: one method best for all. Now read the review's conclusion. Search for unwarranted.

> *lead us to conclude that any application of learning styles in classrooms is unwarranted*

Figures quoted in this illustration are from [195].

WHERE THIS PICTURE BREAKS. The review is from 2008 and covers what had been published by then. The authors say plainly that the negative results are not a conclusive refutation. Many versions of learning styles have never been tested at all. "No adequate evidence" is the claim the review supports. "Proven false for every version" is not.

Illustration 3. Open Newton and Miah, "Evidence-based higher education – is the learning styles 'myth' important?", *Frontiers in Psychology* 2017, at https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.00444/full. They surveyed 114 academics at UK universities. Search for 2012 study.

> *In a 2012 study, 93% of a sample of 137 UK school teachers agreed with the statement "Individuals learn better when they receive information in their preferred learning style (e.g., auditory, visual, kinesthetic)." In our sample of academics in UK Higher Education, 58% agreed with that same statement*

Now search for Thirty-three.

> *Thirty-three percent of participants reported having used Learning Styles in the last 12 months*

Belief ran well ahead of use. The survey then explained the missing evidence, and asked again. Search for 32%.

> *a substantial number of participants (32%) stated that they would continue to use Learning Styles despite being presented with the lack of an evidence base to support them*

So expect the idea in your own staff room. Expect, too, that the evidence alone will not move everyone. The weakness these academics agreed with most was that the theory is conceptually flawed. The authors give an example. Search for heart sounds.

> *teaching medical students to recognize heart sounds would be impossible using visual methods, whereas teaching them to recognize different skin rashes would be impossible using sounds*

The content decides how it must be taught, not a label on the learner. Every student listens to a murmur. Every student looks at a rash. When you teach accredited social health activist (ASHA) workers to measure a waist with a tape, every one of them holds the tape, whatever each says her style is.

Figures quoted in this illustration are from [196].

WHERE THIS PICTURE BREAKS. The survey is of UK university teachers who chose to answer it; the authors call it a convenience sample. It measures what they believe, not whether learning styles work. None of its figures describe Indian teachers. And "would continue" is a stated plan, not something the survey watched them do.

MUST KNOW POINTS FOR YOU.

- "Learners have learning styles, so teaching to them helps" joins two claims. Accept the first; refuse the second.
- To test any claim that matching helps, ask for all four cells. Take the MD within each group, the same way round. Only opposite signs make a crossover.
- Comparing down a column, visual learners against verbal learners on the same lesson, is not the test. Compare the two lessons within one group.

- A lesson that is better for one group by a lot and for the other by a little is still better for both. Teach everyone with it.
- Rejecting learning styles is not a vote for teaching everyone one way. Choose methods for the material and the level.
- Cook and colleagues, reported by Pashler, split 123 internal medicine residents by style and by method. They found no support for matching.
- Expect the idea among colleagues. Lead with the content argument: a murmur has to be heard and a rash has to be seen.
- Say "there is no adequate evidence for it", not "it has been proven false". A crossover in one small table can also arise by chance: the MD check reads the pattern, not whether it is real.

EXERCISE 1 (design). A colleague wants a learning-style questionnaire to decide how your department teaches community medicine to MBBS students. Specify the study that could settle it for two styles, visual and verbal, and two ways of teaching one topic. Say what result would support matching, and what result would not.

PRACTICE. Ten problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. A 2 × 2 table has rows P and Q and columns X and Y. Row P: X = 60, Y = 70. Row Q: X = 50, Y = 65. In each row, which column is higher? Is there a crossover?

2. Row P: X = 72, Y = 64. Row Q: X = 58, Y = 66. Work out X minus Y in each row, and say whether there is a crossover.

3. Row P: X = 55, and X minus Y = -6. Row Q: Y = 66, and X minus Y = 4. Find the missing Y in row P and the missing X in row Q. Is there a crossover?

4. Made-up numbers, from no study. Sixty MBBS students take a learning-style questionnaire. Within each style group, a lottery sends half to a lesson on reading a food label taught with diagrams. The other half get the same lesson as text. A week later all of them take the same test.

Mean scores: visual group, diagram 80%, text 72%. Verbal group, diagram 55%, text 63%. A colleague says: "The verbal group did worse on both lessons, so this tells us nothing about matching." Is there a crossover, and is the colleague right?

5. Made-up numbers, from no study. The same design as before. Visual group: diagram 75%, text 60%. Verbal group: diagram 64%, text 62%. Is there a crossover, and what should the teacher do?

6. Made-up numbers, from no study. Forty ASHA workers answer a questionnaire that labels them visual or auditory. Within each group, a lottery sends half to a session on the child growth chart taught with picture cards. The other half get the same session as a spoken explanation. Everyone takes the same test a week later.

Visual group: picture cards 58%, spoken 66%. Auditory group: picture cards 71%, spoken 63%. Is there a crossover? Does it support learning styles, and does it support meshing?

7. Here is a worked answer. Find the step that broke. Made-up numbers: visual group, diagram 68%, text 57%; verbal group, diagram 62%, text 59%.

```
on diagrams, visual 68 beats verbal 62
on text, verbal 59 beats visual 57
each group wins on its own method
so there is a crossover, and matching works
```

8. Here is a worked answer. Find the step that broke. Made-up numbers.

```
visual learners taught with diagrams: mean 71%
verbal learners taught with text: mean 69%
both matched groups scored well above the 50% pass mark
so matching teaching to style works
```

9. A colleague says in a department meeting: “We tried it. The visual learners in our batch scored 74% after the diagram session, and the whole batch averaged 66% after the text session. So visual learners do better when we teach visually.” Decide what to compute, compute it, and say what your answer does not establish.

10. A senior teacher says in a curriculum committee: “Most teachers believe in learning styles. Newton and Miah found 58% of 114 UK academics agreed, and an earlier study found 93% of 137 schoolteachers agreed. That many people cannot all be wrong.” Decide what to compute, compute it, and say what your answer does not establish.

7 · Explaining is how you find out whether you understand

Checked against the sources as they stood on 2026-09-25.

DEFINITION. People feel they understand how things work in far more detail than they do. Rozenblit and Keil named this the illusion of explanatory depth in 2002.

It is measured in three steps. A person rates how well they understand something. They write out a detailed, step-by-step explanation of how it works. Then they rate their understanding again.

In Rozenblit and Keil's first study, 16 graduate students rated 48 things taken from books on how things work. Each then explained four devices, such as a zipper or a flush toilet, and rated again. The ratings fell after explaining.

The drop has held up. In a second study, 33 undergraduates showed the same pattern. When students were warned at the start that they would have to write explanations, the drop was smaller, but it did not go away. The drop was large for devices and natural processes, small for facts such as capital cities, and absent for procedures and for the plots of films.

Expertise does not reliably remove the illusion either. A later study from the same laboratory found that people with more education show it on topics from their own subject.

Two study techniques put explaining to work. In self-explanation, a learner explains how new information relates to what they already know, or explains the steps they take while solving a problem. In elaborative interrogation, a learner works out why a stated fact is true. A 2013 review by Dunlosky and colleagues rated both of moderate utility, below the high utility of practice testing.

Self-explanation has helped across different materials, a wide range of ages, and tests of memory, understanding and transfer to new problems. Almost all the studies of both tested learners within minutes, and few were run outside the laboratory. [197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 216, 217, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 228, 228, 229, 229, 218]

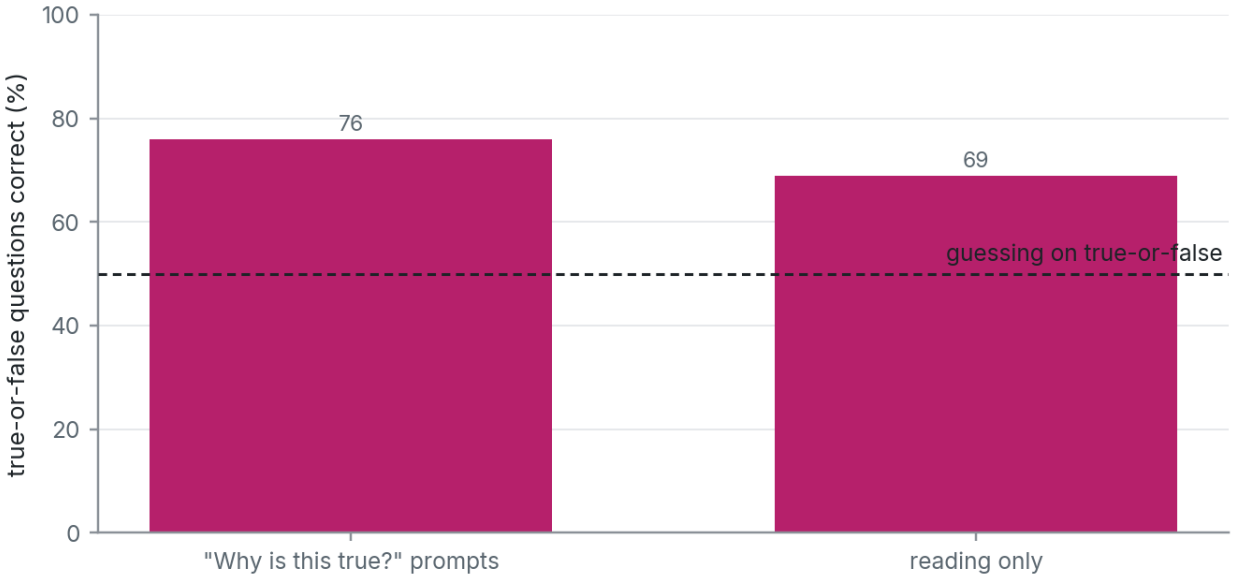
IN PLAIN TERMS. Ask yourself whether you understand how a ceiling fan works, and you will probably say yes. Now write down, step by step, what happens between pressing the switch and the blades turning. Then ask yourself the first question again. Your answer is likely to come down.

That drop is the illusion of explanatory depth. You felt you understood. Trying to explain showed you where you do not.

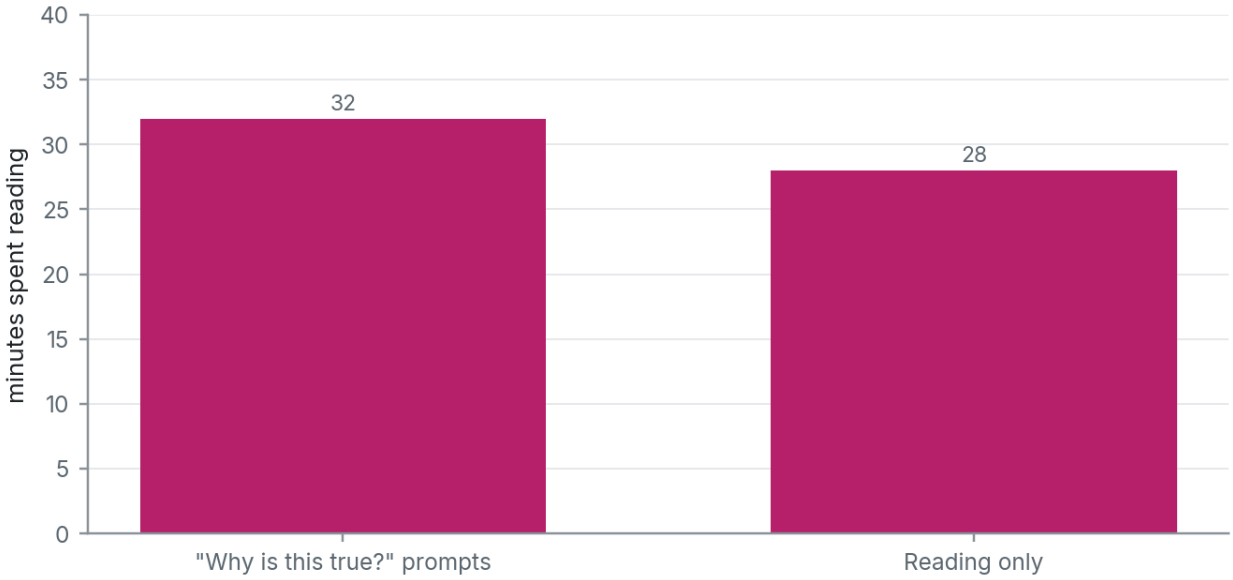
The illusion lives in "how does it work" and "why". In one study, people judged their knowledge of everyday procedures, such as how to bake chocolate chip cookies from scratch or tie a bow

tie, about right. The steps for weighing a child may behave the same way; nobody has tested it. Ask why each step is there, and you are asking how something works.

So use explaining as a check. When a student nods, ask them to explain the step.



One introductory biology class reading a textbook passage on human digestion. Students who met a “Why is this true?” prompt about every 150 words got 76% of 105 true-or-false questions right. Students who read without prompts got 69%. Guessing alone would score about 50%. The test was not reported as delayed.



What the prompts cost in the same class: 32 minutes of reading with “Why is this true?” prompts against 28 without, 4 minutes more for 7 more percentage points correct.

Illustration 1. Start with yourself. Pick something you use every day and believe you understand: how a pressure cooker cooks dal faster than an open pan. Give yourself a mark from 1 to 7 for how well you understand it. 1 means you have almost no idea and 7 means you could explain every part. That is the scale Rozenblit and Keil gave their students. Write the mark down.

Now write out the explanation as a chain of steps, each one following from the one before. Book 0's section on premise and inference (Book 0, F4) gave you the test: each step has to follow from the last. Wherever you find yourself writing "and then somehow", a step is missing.

Mark yourself again, from 1 to 7. If the second mark is lower, you have just met the illusion.

Open Rozenblit and Keil, "The misunderstood limits of folk science", *Cognitive Science* 2002, at https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/. Search the page for genuine surprise.

> *many participants reported genuine surprise and new humility at how much less they knew than they originally thought*

These were graduate students who had just tried to explain a zipper or a flush toilet. Now search for extreme warning.

> *Importantly, the drop was still significant—the illusion held even with the extreme warning.*

Here the students had been told at the start that they would have to write explanations. The drop was smaller than without the warning, and it was still there. Now search for cookbook.

> *People appear to be quite accurate in assessing their own knowledge of "cookbook" procedures in contrast to their knowledge of how things work.*

This tells you where to aim the check. Say you will teach a batch of MBBS students to measure a child's height next week. The steps in the manual are a procedure, and you probably know them about as well as you think. Now write, from memory, why each step is there, and what goes wrong in the reading if it is skipped. That is a question about how something works. Mark every place you stall. That list is your reading for the evening.

Figures quoted in this illustration are from [230].

WHERE THIS PICTURE BREAKS. The paper plots its students' mean ratings and does not print them, so you cannot compare the size of your own drop with theirs. A lower second mark tells you that you overrated yourself. It does not tell you what the right explanation is. A smooth explanation can still be wrong, and only checking it against a good source shows that. The students explained devices. Nobody in these studies tested the reasons behind the steps of a clinical method, so treat the height example as a place to look, not a measured case.

Illustration 2. Now turn explaining into something your learners do. You are taking a batch of MBBS students through a textbook passage on human digestion.

Dunlosky and colleagues describe a class study of this kind, by Smith and colleagues in 2010. Open their review at https://doi.org/10.1177/1529100612453266 and search for Why is this true. Undergraduates in an introductory biology course read a long passage on human digestion from their course textbook.

> *For half of the students, 21 elaborative interrogation prompts were interspersed throughout the text (roughly one prompt per 150 words)*

Each prompt was a statement from the text followed by "Why is this true?" The other half read at their own pace without prompts. Everyone then answered 105 true-or-false questions, none of them the same as the prompts.

Group	Minutes spent reading	Questions correct
"Why is this true?" prompts	32	76%
Reading only	28	69%

Guessing alone would score about 50% on a true-or-false test.

Work out what the prompts bought, and what they cost.

```
76 minus 69 = 7 percentage points more correct
32 minus 28 = 4 minutes more reading
```

To use it, pick the statements in your passage that carry the mechanism. After each one, write "Why is this true?" Your students answer in writing. Then they check the answer against the text, or with you, because an explanation that is wrong and left unchecked can hurt learning.

Figures quoted in this illustration are from [116].

WHERE THIS PICTURE BREAKS. This is one class, one passage and a true-or-false test. The account does not say the test was taken days later, and almost all studies of this technique tested within minutes, so the 7 points say nothing about what lasted. Prompts spread thinly, one every page or two, have shrunk the benefit or reversed it. Learners who know little about a subject may gain less from "why" questions.

MUST KNOW POINTS FOR YOU.

- A learner who nods, or says "yes, understood", has told you how they feel, not what they can do. Ask them to explain the step in their own words, aloud or on paper.
- Before you teach a topic, write your explanation out from memory, and read up on the places where you stall.
- When you ask learners to explain, have them check the explanation against the text or against you. A poor explanation, left unchecked, can hurt learning (Weinstein and colleagues).
- Asking someone to list the steps of a procedure may not show you the illusion. In one study, people judged their knowledge of everyday procedures about right.
- Do not claim that "why" questions or self-explanation produce lasting learning. Only a few studies tested after a delay, and their results were promising. Test your own learners days later before you say they worked.

- Self-explanation took 30% to 100% more time in the few self-paced studies that reported it. The "why" prompts in the biology class took 32 minutes against 28 for reading alone. Weigh fewer problems explained against more problems solved.

EXERCISE 1 (teaching). A first-year resident tells you: "I read the chapter on anthropometry, body measurement, twice last night. I understood all of it, so I can take the MBBS batch on Monday." You have ten minutes and a whiteboard. Explain why "I understood it" is not yet evidence. Then give them one thing to do tonight instead of a third reading.

8 · Working memory is small: cognitive load, and effort that helps

Checked against the sources as they stood on 2026-09-25.

DEFINITION. New information is handled first in working memory, which can hold and work on only a limited number of new elements (pieces of information held as one) at a time, and only for a short time. What has been learned is kept in long-term memory, which has no known limit.

Cognitive load is the demand a learning task makes on working memory. Cognitive load theory, set out by Sweller and colleagues, separates two sources of it. Intrinsic load comes from the material itself and from how much the learner already knows; it changes only if what is taught or the learner's knowledge changes. Extraneous load comes from how the material is presented and what the learner is made to do; the teacher can change it. If the load becomes too high, learning and transfer to new problems suffer.

For novices, studying a worked example, a problem shown with its full solution, can teach more than solving the same problem. That held for problems whose steps depend on each other. For mathematical definitions it reversed: students made to produce the answer did better than students shown it. So, this book suggests, retrieval practice (section 3) still has a place with beginners.

A diagram alone has taught more than the same diagram with text that repeats it. Spoken information is gone in seconds, so the learner may have to hold it in working memory, which adds load and can reduce learning.

These effects shrink, then vanish, and can reverse as learners gain expertise. Interleaving, mixing different problem types in one session instead of doing one type in a block, gave college students better test results a week later. [231, 232, 233, 234, 235, 236, 237, 238, 239, 239, 240, 241, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 216, 255, 256, 218]

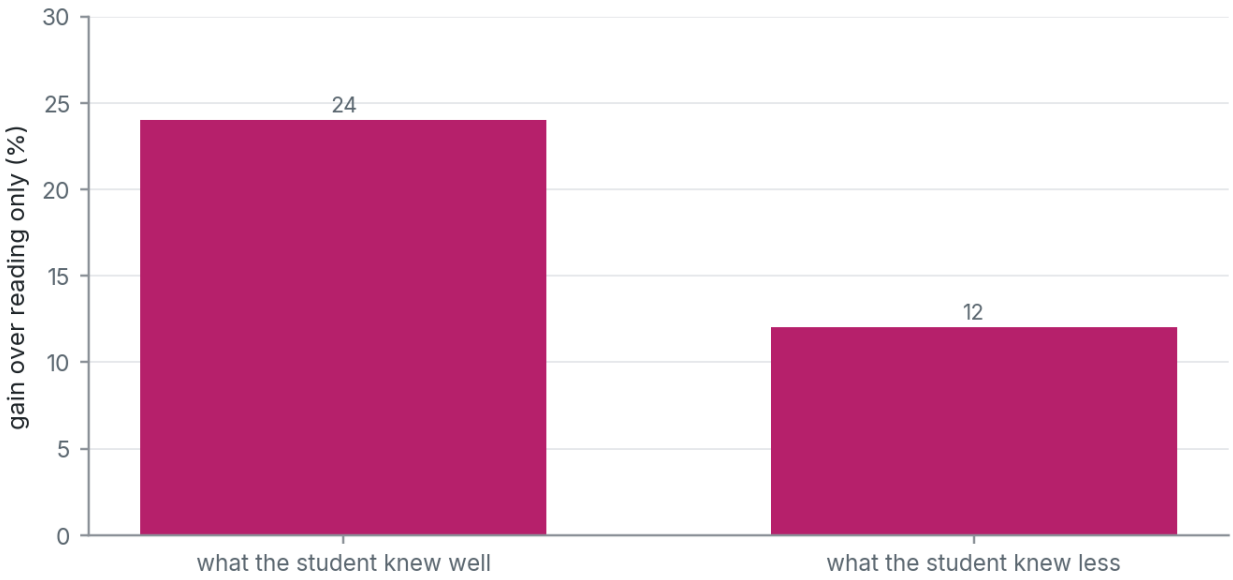
IN PLAIN TERMS. Your working memory is the part of your mind that holds what you are thinking about right now. It is small. It can juggle only a limited number of new pieces at once, and only for a short time. The review behind this section gives no number for either.

Your long-term memory is where learned things are kept, and it has no known limit. Anything already learned comes back from it as one piece. For you, "percentage" is one piece. A learner meeting it for the first time has several pieces to hold.

Overload working memory and less is learned. Some of the load comes from the material itself, and you cannot remove it without changing what is taught. Some comes from how you teach it, and that part is yours to cut.

For a beginner, a problem shown worked out, step by step, often teaches more than being made to solve it cold. As they learn, take the help away a step at a time.

Once the basics are held, effort helps. The same effort that swamps a beginner is what moves a more advanced learner on.



In one study, students answered "why" questions about facts on Canadian provinces and German states. Over reading alone, the review reports a 24% increase on the facts they knew well and a 12% increase on those they knew less (whether points or a share is not stated): twice as much where knowledge was greater.

Illustration 1. Start with the usual way it goes wrong. You explain aloud, once, how to turn a score into a percentage, then set a batch of first-year medical students three problems. Your words were gone within seconds. Each student now has to hold what you said in working memory while also holding the numbers and trying the steps. That is load you added, and it is load you can take away.

Instead, write a worked example on the board and leave it there. The numbers here are made up. 26 of 40 students answered question 3 correctly. What percentage is that?

```
26 divided by 40 = 0.65
0.65 times 100 = 65%
```

Then give a partly worked problem, with the last step left for them. 18 of 24 students answered question 4 correctly.

```
18 divided by 24 = 0.75
0.75 times 100 = ?
```

Then give a whole problem with nothing worked: 31 of 50 students answered question 5 correctly. What percentage is that? They should reach 62%.

You have faded the help: a full example, then one with a step left out, then a bare problem. A percentage is the skill of Book 0, A4. For you it is a single step. For these students it may not be yet.

Now take the same three problems to a batch of residents who have done this a hundred times. For them the worked example is text to read and ignore, and reading it costs time. Give them the bare problem from the start.

WHERE THIS PICTURE BREAKS. The numbers are made up, to show the order of the steps; they report no class. The review behind this section says that worked examples teach novices more than problems do, and that the benefit fades and can reverse with expertise. It does not say at what point a particular learner has crossed over. You find that by watching whether they can do the bare problem.

Illustration 2. Now the other side. Once learners can do each kind of problem on its own, make them work.

Mix the problem types. Suppose your batch can now find a percentage, a ratio and a mean when each is asked in a block of its own. Next session, shuffle them: a percentage, then a mean, then a ratio, then a percentage. This is interleaving. Open Weinstein and colleagues, "Teaching the science of learning", 2018, at https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ and search for shuffling.

> *Rohrer and Taylor (2007) found that shuffling math problems that involved calculating the volume of different shapes resulted in better test performance 1 week later*

The reason the authors give is that mixing makes the student choose the method each time, which a block of one type never asks them to do.

"Why" questions also pay more to learners who already know more. Take the "why is this true?" questions of section 7. Open Dunlosky and colleagues' review at https://doi.org/10.1177/1529100612453266 and search for Canadian provinces. Woloshyn and colleagues gave Canadian and German students facts about Canadian provinces and German states, so each student knew one set well and the other less. Students who answered "why" questions about the facts gained more over reading alone where they knew more.

> *in their high-knowledge domain (a 24% increase) than in their low-knowledge domain (a 12% increase)*

Facts about	Gain over reading only (%)
what the student knew well	24
what the student knew less	12

The same "why" questions paid twice as much where the learner already knew more.

```
24 divided by 12 = 2
```

Figures quoted in this illustration are from [257, 116].

WHERE THIS PICTURE BREAKS. Interleave problems that are related, like three calculations from the same unit. Mixing different subjects in one sitting has not been shown to help. The 24% and 12% come from one study of facts about places, and the review does not say whether they are percentage points or a share of the reading-only score; read them as "about twice as much", not as a size to expect. They show that "why" questions help less where knowledge is thin, not that they never help a beginner.

MUST KNOW POINTS FOR YOU.

- Saying more is not teaching more. Put the steps where they stay: on the board, a chart on the wall, a handout. Written steps that add what the picture cannot show (the order, what to check) are not redundant; a paragraph that repeats the labels is.
- For beginners, show a worked example before you set a problem. Then give a problem with part of the solution done, then a whole problem. Take the help away as they gain skill.
- What helps a learner new to a topic can hold back one who already knows it. For them, set problems.
- Before you put the same content on a slide twice, as a diagram and as text that repeats it, ask what the second copy adds.
- Once the basics are held, make learners work: mix related problem types and have them produce answers rather than read them.
- Do not tell a colleague a session "overloaded" learners as though you measured it. Say what the learners could and could not do afterwards.

EXERCISE 1 (critique). A colleague plans a one-hour session for 30 accredited social health activists. The topic is reading a weighing scale and recording a child's weight. For the first 40 minutes the colleague talks through every step. Each slide shows a labelled picture of the scale. Beside it, a paragraph says the same thing in words.

In the last 20 minutes each worker weighs a child alone and records the weight. Nobody has seen a weighing worked through first. Name what in the plan adds load that could be cut, and say what you would change.

9 · Writing a learning objective that can be assessed

DEFINITION. A learning objective is one sentence that states what a named learner will be able to do at the end of a piece of teaching, in a form that someone else could watch or mark. It has six parts.

1. Who: the learners.
2. The action: one verb whose result can be seen or marked.
3. The content the action is done to.
4. The condition: what the learner is given, or is not allowed, while doing it.
5. The standard: how much, or how well, counts as done.
6. By when: the point at which it will be checked. In this book that is the end of the session and again on a delayed test.

Chatterjee and Corral name five of these parts: who, will do, how much or how well, of what, and by when. The condition is this book's addition.

A verb that names a state of mind cannot be watched or marked. Understand, know, learn, appreciate, believe, be familiar with and comprehend are such verbs, and they do not belong in an objective. Each objective carries one action. Two actions make two objectives.

Bloom's taxonomy is a sorted list of kinds of thinking skill. It was revised in 2001. The revision sorts the skills into six levels, from lowest to highest: remember, understand, apply, analyse, evaluate and create. Chatterjee and Corral's table says what the learner does at each level.

level	what the learner does
remember	Retrieve relevant knowledge from long term memory
understand	Construct meaning from instructional messages, including oral, written and graphic communication
apply	Carry out or use a procedure in a given situation
analyse	Break material into its constituent parts and determine how the parts relate to one another and to an overall structure or purpose
evaluate	Make judgments based on criteria and standards
create	Put elements together to form a coherent or functional whole

IN PLAIN TERMS. Many session plans open with a line like "Students will understand obesity." Nobody can check that. You cannot see understanding. You can see what a person does with it.

Here is the test of a finished objective. Hand it to a colleague and ask them to write the question that checks it. If they cannot, the sentence is not finished.

Illustration 1. Start with a list that fails. Chatterjee and Corral open their article with it. An anaesthesiologist is giving grand rounds, a teaching talk for a whole department. The topic is ma-

lignant hyperthermia (MH), an emergency that can arise during an operation under anaesthesia. You need to know nothing more about MH to follow this. He opens with these four objectives:

1. Understand the pathophysiology of MH.
2. Review the clinical presentation of MH.
3. Discuss the treatment of MH.
4. Become familiar with caffeine-halothane contracture testing for MH.

Now run the first line through the six parts. Who? Not said. The action? "Understand". You cannot watch it, so you cannot write a question that checks it. The content is there: the pathophysiology, the way the disease works. The condition, the standard and the by-when are missing.

Run the other three lines through. "Review" and "discuss" could name what the learners do, but here they read as the speaker's plan for the hour. "Become familiar with" names a state of mind. None of the four lines says who, how well or by when. The slide tells the audience what the talk will cover. It does not tell them what they will be able to do afterwards.

Now read the authors' rewrite that covers the treatment. The stem, "Upon completion of the grand rounds presentation, participants will be able to", supplies who and by when.

> *Summarize, in order, the critical steps in the intraoperative management of an MH crisis according to recommendations from the Malignant Hyperthermia Association of the United States.*

Find the parts. The action is "summarize". The content is the steps for managing a crisis during an operation. "According to recommendations from" the association names the list the answer is marked against. "In order" is part of the standard.

Now write the check. Give each participant a blank sheet: "Write the steps for managing an MH crisis during an operation, in order." You can mark that. Notice one thing the rewrite still leaves open: how many steps must be right. Add it and the objective is finished.

WHERE THIS PICTURE BREAKS. The rewrite is the authors' own, for one talk to trained anaesthesiologists. It shows the form of an objective. It says nothing about what an Indian MBBS batch needs to be able to do; the level and the standard come from your learners' job, not from this example.

Illustration 2. Now write one for a session of your own. You are teaching a batch of final-year MBBS (Bachelor of Medicine and Bachelor of Surgery) students on body mass index, BMI, which Book 0 built (Book 0, B6). Your first draft is the one most people write.

> *Students will understand BMI.*

Take the parts in turn.

Who: “each student in this final-year batch”. Say each, so the objective is about every learner, not the batch on average.

The action: ask what the students will have to do with BMI in the job. In a clinic or a survey they will be handed a weight and a height and will have to say which category the person falls in. That is two actions, working out the number and placing it in a category. Pick the one the job ends on: classify. Working out the number becomes a step inside it.

The content: an adult’s BMI, from a measured weight in kilograms and a height in centimetres.

The condition: given a printed table of BMI categories, without notes. Handing over the table matters. Book 0 showed that a cut-off is a decision some body made, not part of the arithmetic. The objective tests the classifying, not the memory of one body’s cut-offs.

The standard: correct for at least eight of ten adults.

By when: at the end of the session, and again on a test one week later. As section 1 showed, the second check is the one that counts: learning is what lasts.

Put it together.

> *At the end of this session, each student in this final-year batch, given the weight in kilograms and the height in centimetres of ten adults and a printed table of BMI categories, will classify each adult’s BMI into the table’s category, without notes, correct for at least eight of the ten, on the day and again on a test one week later.*

It is long. Chatterjee and Corral hear the same complaint from teachers. The length is where the check lives: a colleague can now write your test from this sentence alone.

Last, find its level. Here the two sources disagree, and the disagreement teaches something. Chatterjee and Corral’s verb table lists “classify” under analyse. Adams lists “classifying items in groups” as a sign of comprehension, the old name for understand. Neither verb list settles it. Look at the task instead. Reading a number against a printed table is carrying out a procedure in a given situation, which is the table’s own description of apply. So the level is apply, whatever the verb.

Figures quoted in this illustration are from [277].

WHERE THIS PICTURE BREAKS. Eight of ten is a standard chosen for this session, not a figure from any study. Set yours from what a wrong answer would cost. For a classification a student will use on patients next month, you might demand ten of ten.

Placing the objective at apply works because the student classifies adults they have not seen before. If the ten adults on the test are the ten from the session, the same sentence is tested at remember.

MUST KNOW POINTS FOR YOU.

> - “Understand” is the name of a Bloom level, and it is also a verb Chatterjee and Corral tell you never to use in an objective. The level is shown by actions you can see: explain,

summarise, estimate. When you want understanding, write the action that would show it.

- A verb does not fix the level. "Describe" sits under both remember and understand in Chatterjee and Corral's table, and the two sources in this section put "classify" at different levels. Decide the level from the task. Ask whether the learner meets a case they have not seen before, or repeats something they were shown.
- One action per objective. When a draft says "diagnose and manage", write two objectives.
- Before you settle on remember, ask what the learner will have to do with the material in the job. Write the objective at that level.
- Bloom's levels sort thinking skills. So the list cannot grade weighing a child. For a skill like that, still write an action you can watch, such as "weigh" or "demonstrate". Then judge it by watching, not by the verb list.
- An objective states what you intend. Never report a well-written objective as an outcome.
- For your build session, write one objective, not a list. The one-objective rule is this book's choice, not a research finding.

EXERCISE 1 (critique). Here are three objectives of the kind you will find in teaching plans. For each one, name every part that is missing or broken. Then rewrite it as a finished objective.

1. For a batch of MBBS students: "Students will know the causes of childhood obesity."
2. For a session for accredited social health activists (ASHAs), village health workers: "ASHAs will appreciate the importance of weighing children every month."
3. For a journal club of residents: "Residents will understand and critically appraise the trial."

EXERCISE 2 (design). Choose the session you will teach for this book's build: a real group of learners and one topic. Write its single objective with all six parts. Then hand it to a colleague and ask them, without talking to you, to write one test question from it. Compare their question with what you meant. Revise the objective until the question they write is the one you would have written.

EXERCISE 3 (teaching). A first-year resident is preparing her first teaching session. She asks why "understand" is forbidden in an objective, when it is the name of one of Bloom's levels. You have ten minutes and a whiteboard.

10 · A test that matches the objective

DEFINITION. An item is one question or task on a test, marked on its own. A test matches a learning objective when every item asks the learner to carry out the objective's action, on its content, under its condition. Each item is marked right or wrong against a key written in advance, and the objective's standard is applied to the learner's total.

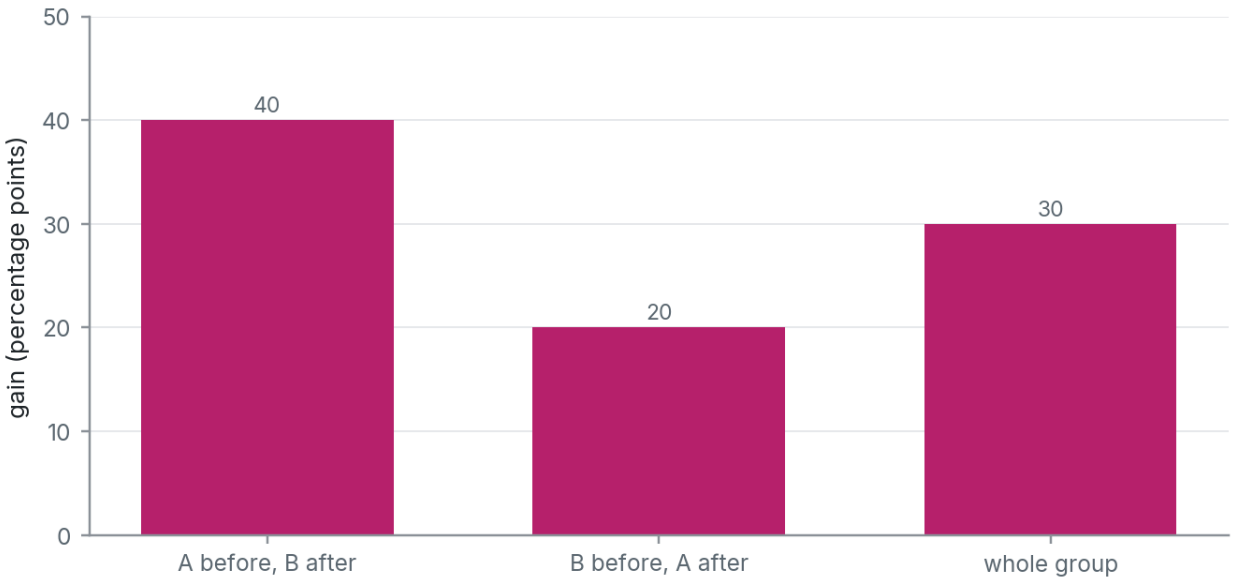
A test that counts towards a mark or a decision about the learner is summative. In this book, a test used only to show the learner and you where they stand is formative.

An item is at the objective's level when it cannot be answered by repeating something said or shown in the session. For an objective above remember, that means the item hands the learner a case they have not met before.

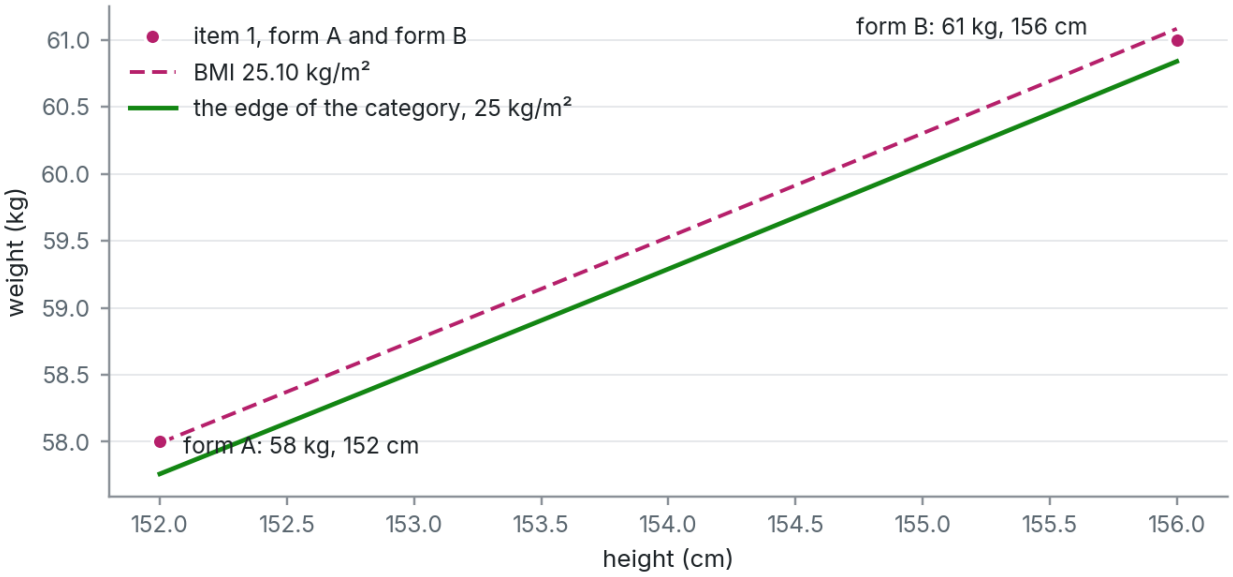
Two forms of a test are parallel when each item on one form has a partner on the other. Only the case or the numbers differ.

The pre-test is taken before the teaching and the post-test at its end. [278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289]

IN PLAIN TERMS. A test can be well made and still test the wrong thing. Suppose your objective says students will classify a new adult's BMI. A test that asks them to write the BMI formula checks something else: whether they remember a line from your session.



Made-up numbers. A session raises every score by 30 points, and form B is 10 points easier than form A. Half the group taking A then B gains 40; half taking B then A gains 20. The whole group's mean gain is 30, the session's own effect.



Item 1 on each form: 58 kg at 152 cm on form A, 61 kg at 156 cm on form B. Both sit on the curve of weights that give a BMI of 25.10 kg/m^2, to the nearest kilogram, just above the edge of a category at 25 kg/m^2. The numbers differ; the item is as hard to place.

Illustration 1. Start with a test that fails. You wrote this objective for a final-year batch in the last section.

> *At the end of this session, each student in this final-year batch, given the weight in kilograms and the height in centimetres of ten adults and a printed table of BMI categories, will classify each adult's BMI into the table's category, without notes, correct for at least eight of the ten, on the day and again on a test one week later.*

A colleague offers you his end-of-session test.

1. Write the formula for BMI.
2. What is the unit of BMI?
3. Name the categories in the table we used today.
4. True or false: BMI uses height in metres.
5. List two limits of BMI.

Hold each item against the objective. Item 1 asks for a line from the session. So do items 2, 3 and 4. A student can score full marks and still be unable to classify one new adult. Item 5 asks for something the objective never promised. None of the five asks for the action, classify. None gives a case the student has not seen. None hands over the table, which the condition promised.

Now turn the objective into items. Its action and condition give you the item's shape: one adult, weight and height given, table given, notes put away.

> *Item 1. An adult weighs 58 kg and is 152 cm tall. Using the table, write this adult's BMI category.*

Its standard, eight of ten, tells you there are ten such items, one adult each. It also tells you how to mark: each item right or wrong, and eight right meets the objective.

Last, check the level. Would a student who had only memorised your slides get item 1 right? Only by working out the BMI and reading the table. That is carrying out a procedure on a new case, which is apply. The test now asks for what the objective promised, at the level it promised.

WHERE THIS PICTURE BREAKS. This objective has one action and a clear right answer, so its items are easy to write. An objective at evaluate, such as judging whether a trial's comparison was fair, needs a new case and a marking key written in advance. Otherwise two markers will disagree about the same answer.

The five weak items are not useless. Items 1 to 4 test remember, and a quick check of them during the session can tell you who is lost. They are simply not a test of this objective.

Illustration 2. Now build the second form. The post-test cannot repeat the pre-test. A student who met "58 kg, 152 cm" at the start may remember working it out, and a rise in score could then be memory of the item. section 3 found the same danger with repeated quiz questions.

So give each item a partner. The partner changes the numbers and keeps what makes the item hard. Here that is where the BMI falls: an adult close to the edge of a category is harder to place than one far from it. So choose a partner whose weight and height differ but whose BMI is almost the same. BMI is the weight in kilograms divided by the height in metres times itself, as Book 0 built it (Book 0, B6).

```
1.52 times 1.52 = 2.3104
58 divided by 2.3104 = 25.10
1.56 times 1.56 = 2.4336
61 divided by 2.4336 = 25.07
```

Form A, item 1: 58 kg, 152 cm. Form B, item 1: 61 kg, 156 cm. The two BMIs differ by less than 0.1 kg/m^2. Suppose the table's category changes at 25 kg/m^2. Both adults sit just above that edge, so the two are equally hard to place. Unless the table puts an edge between 25.07 and 25.10, the two adults land in the same category. Do the same for all ten pairs.

However carefully you pair them, one form may still turn out easier. That is a systematic error, the kind Book 0 described (Book 0, D7): it pushes every score on that form the same way. Here is how it spoils a gain, and how to cancel it. The numbers are made up to show the arithmetic.

Suppose your session raises every student's score by 30 percentage points. Suppose form B is 10 points easier than form A. You have 30 students. Fifteen take form A before and form B after. The other fifteen take form B before and form A after.

```
A before, B after: 30 plus 10 = 40
B before, A after: 30 minus 10 = 20
mean gain of the whole group: (40 plus 20) divided by 2 = 30
```

order of the forms	students	gain (percentage points)
A before, B after	15	40

order of the forms	students	gain (percentage points)
B before, A after	15	20
whole group	30	30

Had all 30 taken A before and B after, every gain would read 40. You would report 40 points for a session that gave 30. Swapping the order for half the group makes the easier form help one half and hurt the other by the same 10 points. In the mean of the whole group, the two cancel.

WHERE THIS PICTURE BREAKS. The cancelling works for the mean of the whole group, and exactly only when the two halves are the same size. It does nothing for any one student: in this example, every student's own gain is still 10 points wrong. Do not judge one student's progress from one pair of forms.

Matching partners on BMI makes them equally hard for a table of BMI categories. For another objective, find what makes its items hard, and match on that.

MUST KNOW POINTS FOR YOU.

- A test of recall cannot show that learners reached an objective above remember. Before you use a test, ask of each item whether a learner could answer it by repeating something from your session. If yes, it is testing memory of the session.
- Write the test straight after the objective, before you plan the teaching. That order is this book's advice. Biggs's point is that the objective, the activities and the test are designed as one. Then choose activities that make learners do what the test will ask.
- "Use the same test before and after, so the comparison is fair" is the error. A repeated item can be answered from memory of the first attempt. Use two parallel forms: partner items that ask for the same action on different cases.
- Give half the group each order of the two forms, with the halves the same size. A difference in difficulty then cancels in the group's mean gain. With an odd number of learners, one half has one more. The cancellation is then close, not exact. Say so when you report the mean gain. It does not cancel for one learner, so never judge one learner from one pair of forms.
- A short test moves in big steps. On five items, each item is 20 percentage points. One lucky guess or one slip moves a learner a whole step.
- India's undergraduate medical course follows the Competency Based Medical Education (CBME) Curriculum 2024, from the National Medical Commission. Its tables give each competency a level. At knows (K), the learner enumerates or describes; at knows how (KH), discusses or analyses; at shows how (SH), demonstrates. Performs (P) is certified after a set number of supervised performances. Each row also names a suggested assessment method. Read both before you write the test. A competency at SH or P has to be watched.
- The CBME Curriculum 2024 builds internal assessment from day-to-day tests, and quizzes can count. A student needs at least 50% of those marks, and 40% in theory and

in practical separately, to sit the university exam in that subject. So a quiz whose marks count there is not low-stakes. Keep your practice questions out of it.

- A matched test shows what learners can do when they take it. It does not show that your teaching caused it. For that you need the comparison of section 2. Do not report a well-matched post-test as proof that the method works.

EXERCISE 1 (critique). A trainer writes this objective for a half-day session with accredited social health activists (ASHAs), village health workers:

> *At the end of this session, each ASHA, given a child and the scale used at her centre, will weigh the child, completing every step on the programme's weighing checklist. Writing the weight in the register is its last step. The trainer observes, at the end of the session and again at a supervision visit a month later.*

Her test is a written quiz of ten multiple-choice questions on the steps of weighing a child. Say whether the test matches the objective. If not, describe a test that does, and say how you would make a second form of it for the end of the session.

EXERCISE 2 (design). Take the objective you wrote for your build session. Write a pre-test of at least five items and a parallel post-test, each item with its partner. For each pair, say what makes the item hard and how the partner keeps it. Then write down how you will split your learners between the two orders of the forms.

EXERCISE 3 (teaching). Most ASHAs scored well on the written quiz at the end of last year's training. A state health secretary asks why that is not evidence that they now weigh children correctly. One page, answer first, no methods section.

11 · Running one session around retrieval rather than transmission

DEFINITION. A session built around retrieval is one in which the learners spend much of the time bringing material back from memory, each alone and in writing, and seeing the correct answer afterwards.

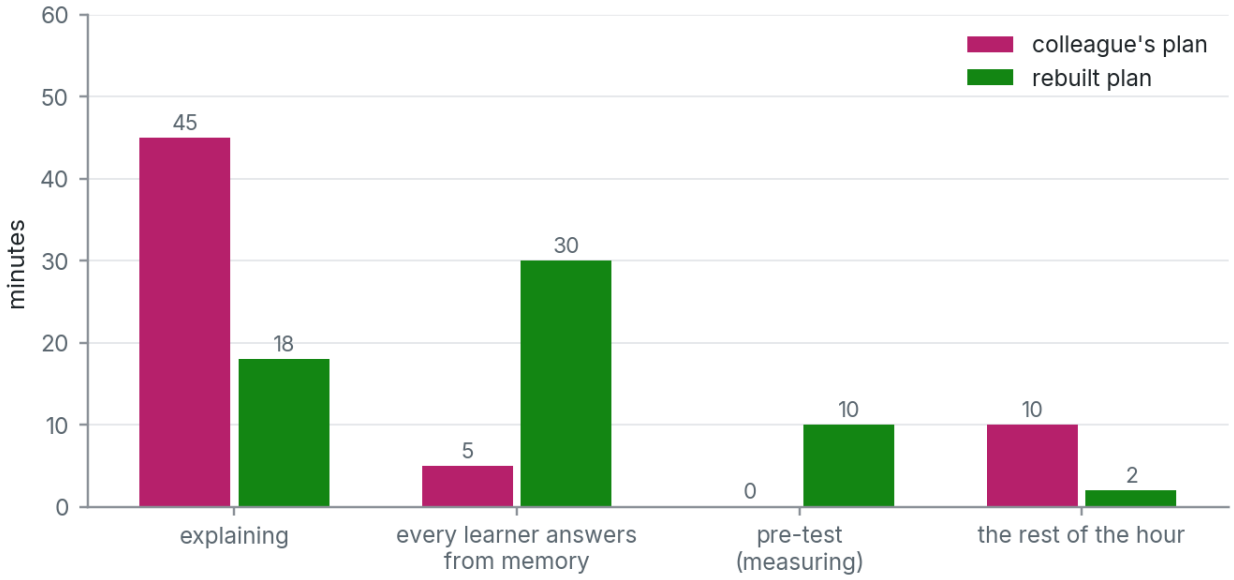
The session plan is a written table of the session's parts, in order. A plan for one session built around retrieval has these parts.

1. The objective, stated to the learners at the start.
2. A pre-test on one of two parallel forms.
3. Short teaching parts. Each is followed by one question that every learner answers alone, in writing, without notes, and then by the correct answer.
4. One question on material from an earlier session, answered from memory.
5. A closing free recall: notes away, each learner writes down all they can about the topic, then checks it against the key points shown on the board.
6. A post-test on the other form.
7. A delayed test on a third form, or the first form again, booked before the session for a date written into the plan. [279, 290, 291, 292, 293, 294, 69, 295, 296, 297, 298, 299, 300]

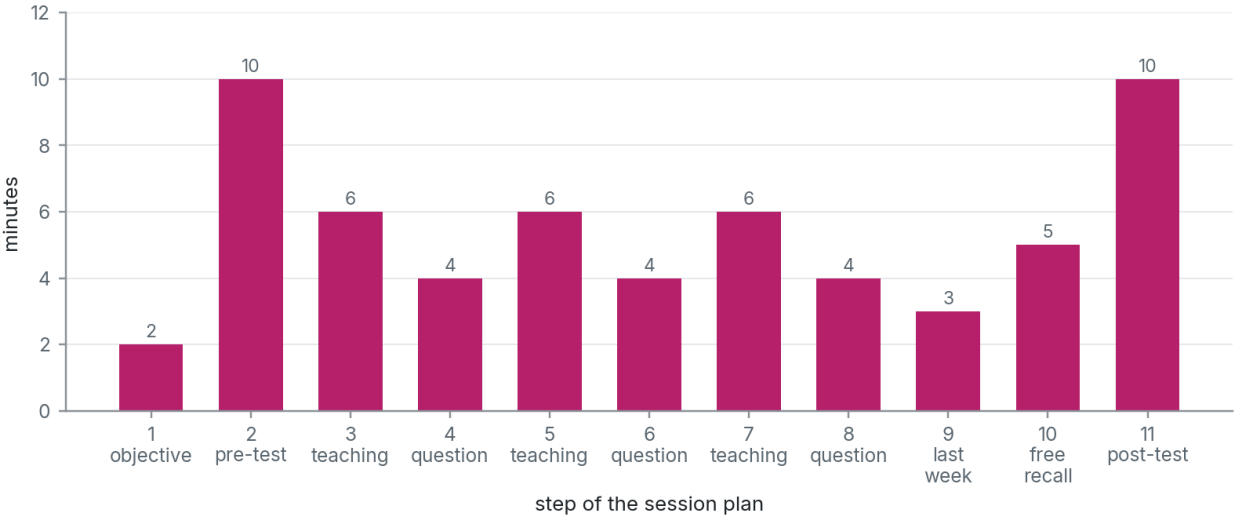
IN PLAIN TERMS. Most sessions are built around the teacher talking. The learners listen, and at the end someone asks "any questions?". This section builds a session the other way round.

The earlier sections gave you the pieces. This section puts them in order on one page, the session plan, with the minutes for each part.

Then you count. How many of the 60 minutes did every learner spend answering from memory? In a lecture the answer is close to none. In this plan it is 30.



The same 60 minutes in two plans. The colleague's: 45 minutes of slides, 5 of a quiz and 10 of "Any questions?". The rebuilt plan: 18 minutes of explaining, 30 with every learner answering from memory, 10 of pre-test and 2 on the objective.



The rebuilt plan step by step. A 10-minute pre-test (step 2) and post-test (step 11) bracket three teaching parts of 6 minutes, each followed by a 4-minute question and its answer.

Illustration 1. Start with a plan that fails. A colleague shares his plan for a one-hour session on body mass index (BMI) for a final-year MBBS batch of 30.

- 45 minutes: slides on BMI, with examples.
- 10 minutes: "Any questions?"
- 5 minutes: a quiz. "Write the formula for BMI."

Count the minutes in which every learner pulls something from memory. The slides: none. The questions: a few students ask, the teacher answers, and the rest listen. Nobody is made to recall anything. The quiz: 5 minutes, and it asks for a line from the slides. The last section showed that such an item tests memory of the session, not the objective.

Now look for what is missing. There is no objective, so nothing to test against. The pre-test gets 0 minutes, so there is no starting point. There is no feedback: the quiz is collected and nobody sees the right answer. There is no delayed test, so the plan cannot show learning. section 1 said learning is what lasts, and it is what this plan never measures.

This is a transmission session. The slides may be excellent. The design gives the learners almost nothing to do.

WHERE THIS PICTURE BREAKS. The colleague's plan is made up, though its shape is common. A lecture is not worthless: it can be the fastest way to lay out new material. What fails here is the hour's balance, 45 minutes of telling and 5 of recall, and the missing checks.

Illustration 2. Now rebuild it around the objective you wrote two sections back: each student classifies ten new adults' BMI, given a table, eight of ten right. The plan runs for 60 minutes. The batch is split in half for the two forms, as the last section showed.

step	min utes	the teacher	every learner
1. objective	2	writes it on the board, reads it out	reads it
1. pre-test, form A or B	10	hands out the forms, half the batch each	classifies ten adults, alone, notes away
1. teaching part 1: working out BMI from kg and cm	6	explains, works one example	watches
1. question 1, then its answer	4	gives one new adult, waits, then shows the worked answer	writes the BMI, then checks it
1. teaching part 2: reading the table	6	explains, works one example	watches
1. question 2, then its answer	4	gives one new adult, then the answer	writes the category, then checks it
1. teaching part 3: adults near the edge of a category	6	explains, works one example	watches
1. question 3, then its answer	4	gives one new adult near an edge, then the answer	writes the category, then checks it
1. question from last week	3	asks one question on last week's topic, then the answer	writes an answer, then checks it

step	min utes	the teacher	every learner
1. free recall	5	“Notes away. Write all you can about working out and reading BMI.” Then shows the key points	writes, then checks against the key points on the board
1. post-test, the other form	10	hands each learner the form they did not take first	classifies ten adults, alone, notes away

The delayed test is booked now, for the first 10 minutes of next week’s session. Write its date into the plan. Use a third parallel form if you can build one. If not, give each learner the form they took first. A week later they may still remember some of its items, so the delayed score may run slightly high. Say so when you report it.

Walk through why each part is there.

The objective goes first so every learner knows what they will be asked to do.

The pre-test gives each learner a starting score. For students who met BMI in an earlier year, it is also a first attempt at recall. For anyone who never met it, it is only a measurement. Trying to recall something you never learned is unlikely to strengthen memory, especially when no answer follows.

The teaching parts are short, 6 minutes each. That is this plan’s choice, not a figure from a study. section 8 showed that working memory holds only a few new elements at once. A short part followed by a question stops the new elements piling up, and it gives every learner a turn. Weinstein and colleagues also report that frequent testing can cut mind-wandering.

Each question gives a new adult, not one from the slides, and asks for what the tests ask for. Every learner writes an answer, alone, before anyone speaks. Then you show the worked answer and each learner checks their own. section 3 showed why all three conditions matter.

The question from last week puts spacing, from section 4, into the hour. Last week’s material comes back after a gap of a week.

The free recall is retrieval of the whole topic in the learners’ own order. It is the task Roediger and Karpicke’s students did in section 3: write down as much as you can. Writing it out also shows each learner what they cannot yet explain. section 7 showed that this is where understanding gets tested.

Now count where the hour goes. The pre-test is counted apart: it measures. The post-test counts as retrieval, because it asks for what was just taught.

what fills the time	minutes
explaining	18
every learner answers from memory	30
measuring (the pre-test)	10
the objective	2

```
6 plus 6 plus 6 = 18
4 plus 4 plus 4 plus 3 plus 5 plus 10 = 30
18 plus 30 plus 10 plus 2 = 60
```

Against the colleague's plan, explaining falls from 45 minutes to 18. So you cover less. That is the trade: in the classroom studies Agarwal and colleagues reviewed, which you met in section 3, retrieval practice beat the same time spent on other activities, lectures without quizzes among them. One session is a fair place to start. In 10 of their 50 classroom experiments, all the retrieval practice fell within a single session.

WHERE THIS PICTURE BREAKS. The minutes are this plan's own choices, sized for one objective and a batch of 30 who have calculators. A topic with more new elements needs more or longer teaching parts, and fewer items per test.

Ten minutes for a ten-item test assumes the learners work fast. Time a colleague on one form before the session. If they need 15 minutes, cut the teaching, not the questions.

MUST KNOW POINTS FOR YOU.

- "Any questions?" and a question to one student are not retrieval practice for the room. Ask a question every learner answers alone, in writing, so that you know every learner did the recalling.
- Write the plan as a table before the session: minutes, what you do, what every learner does. Then add up the minutes in which every learner answers from memory.
- A pre-test on material the learners never met measures where they start. It is not practice. Do not count it as part of the retrieval in your plan.
- For a hands-on objective, the retrieval is doing it. Each learner performs the step from memory while someone ticks a checklist of its steps. With one trainer, watch a few learners each session and say how many you watched.
- Questions spread through a session have not been shown to beat the same questions at its end. Spread them for other reasons: short teaching parts, attention, a turn for every learner. Do not tell colleagues the spreading itself is proven.
- Give the in-session questions the same format as the tests. Change the case, not the kind of question.
- Book the delayed test before the session and write its date in the plan.
- A plan that meets every part here says what the learners did. Present your plan as a design, not as a result.

EXERCISE 1 (critique). A senior resident runs a journal club for eight residents. Her plan: she presents the trial for 30 minutes. Then she picks residents one at a time and asks each a question about it. Their answers are marked, and the marks count in their formal assessment. The session ends there.

Name each place where this plan departs from a session built around retrieval. For each, say what you would change.

EXERCISE 2 (build). This book's build: deliver one teaching session built around retrieval, with a written objective. Choose a real group: a batch of MBBS students, residents, or a session for accredited social health activist (ASHA) workers. Before the session, prepare and date these.

1. The objective, with all six parts.
2. Two parallel forms of the test, and how the group splits between them.
3. The session plan as a table: minutes, what you do, what every learner does.
4. The questions for each teaching part, each with its answer.
5. The date of the delayed test.

Deliver the session. Afterwards, keep every learner's pre-test and post-test. Give the delayed test on its date. Write three lines on what departed from the plan.

EXERCISE 3 (teaching). A district health officer asks whether the district's one-day refresher trainings for ASHA workers should switch from lectures to sessions built around retrieval. One page, answer first, no methods section.

12 · Measuring what changed: pre-test, post-test, gain

DEFINITION. A learner's score on a test is written here as a percentage of the test's maximum, from 0 to 100%. The pre-test score, pre, is taken before the teaching. The post-test score, post, is taken at the end of it.

The raw gain is post minus pre, written post − pre, in percentage points.

The mean gain of a group is the mean of its learners' raw gains. It equals the mean post-test score minus the mean pre-test score. Its spread is the standard deviation (SD) of the raw gains.

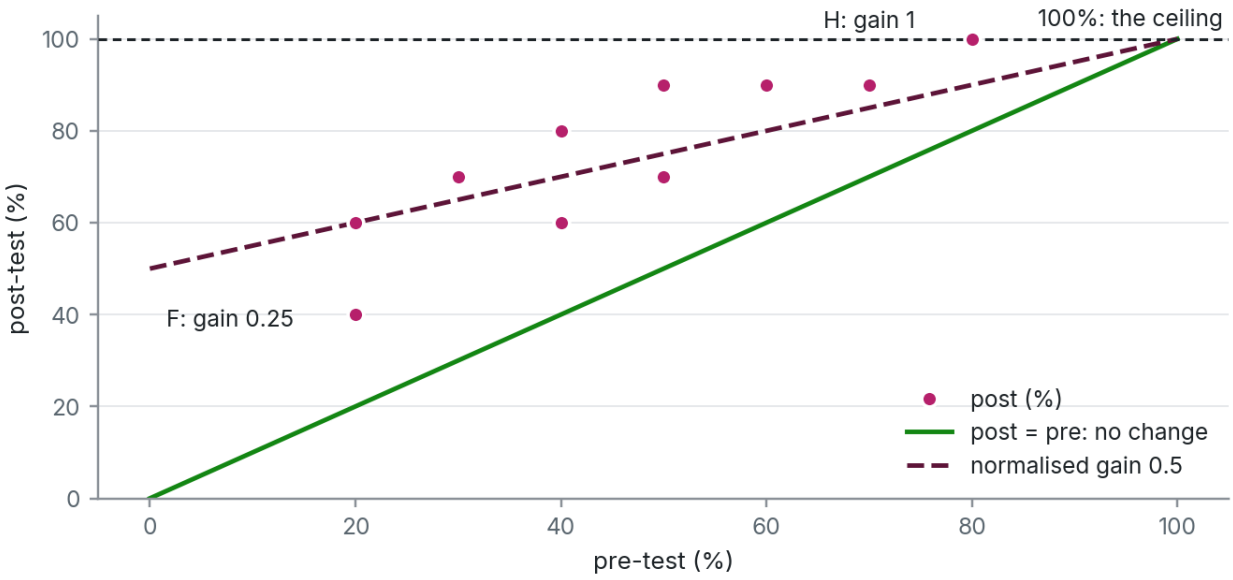
The room to gain is 100 − pre: the most a learner, or a class, could have gained. The normalised gain, written g, is the raw gain divided by the room to gain.

```
g = (post - pre) ÷ (100 - pre)
```

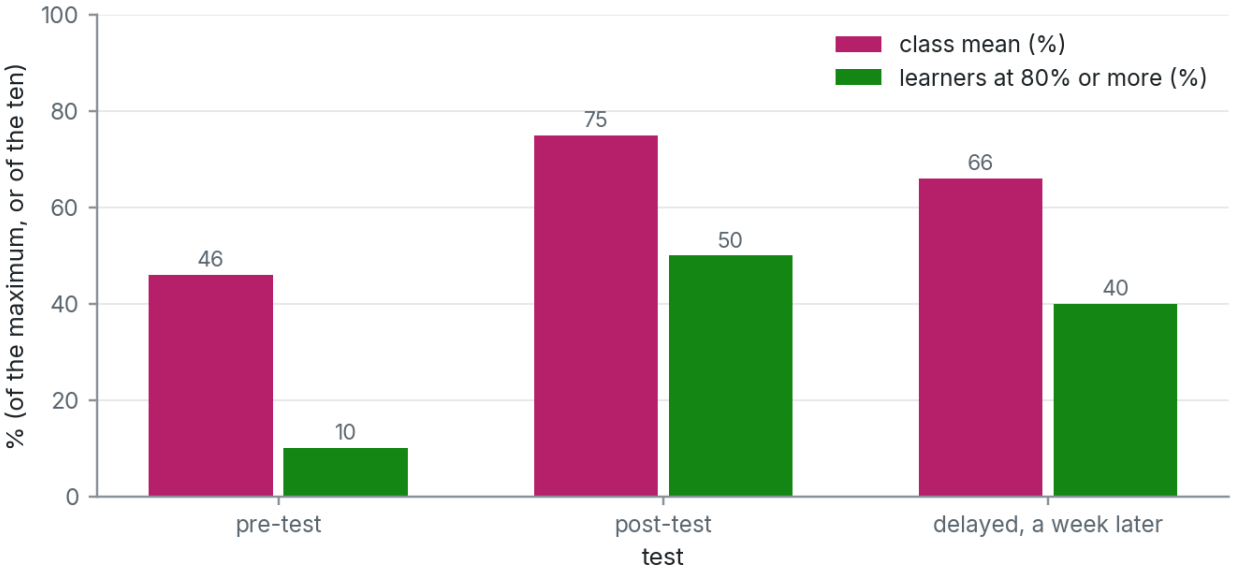
It is a fraction with no unit. It does not exist for a pre-test of 100%. Hake defines the class form from class-average scores. He writes it ⟨g⟩, with angle brackets round g, which here mean "the class average".

The proportion reaching the standard is the number of learners whose score meets the standard the objective set, divided by the number tested. [301, 302, 303, 304, 305]

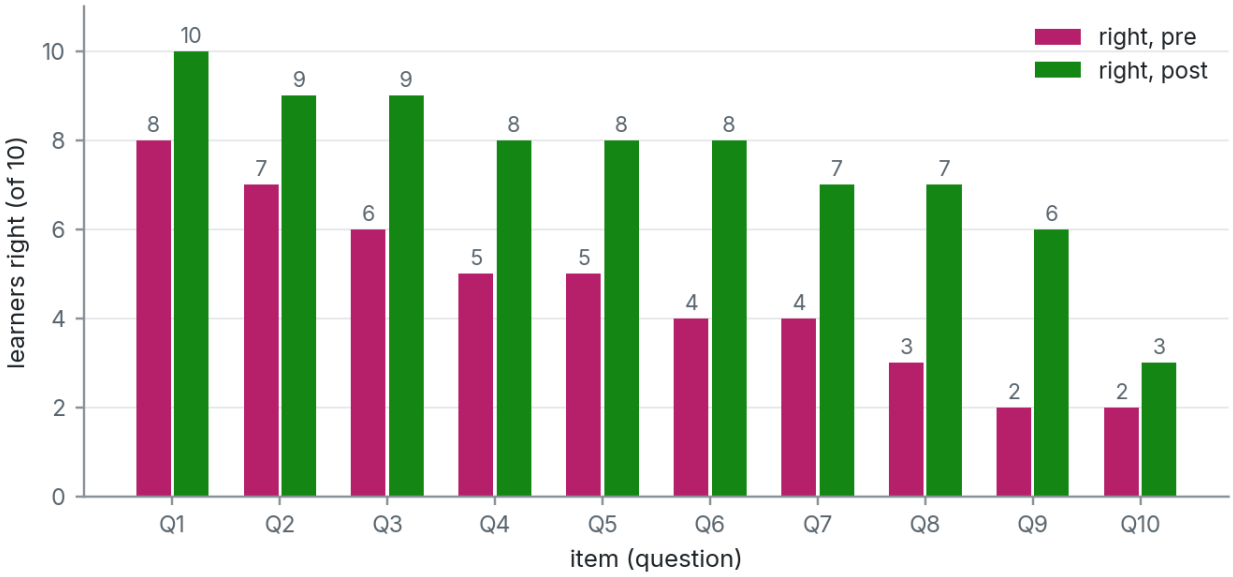
IN PLAIN TERMS. The raw gain has a catch. A learner who starts at 80% can gain at most 20 points, because no score passes 100%. A learner who starts at 20% has 80 points of room. So a gain of 20 points means everything to the first learner and a quarter to the second.



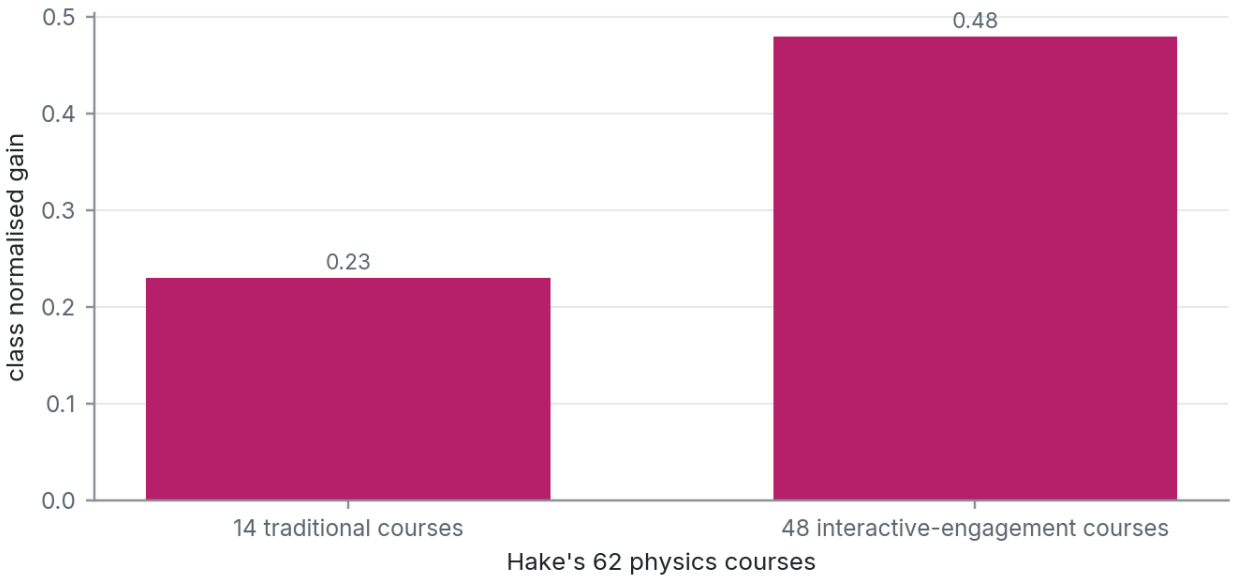
Ten students' pre-test and post-test scores, in %. The scores are made up for the example. Points on the lower line, post = pre, would mean no change. The upper line is every pair with a normalised gain of 0.5: post = 50 + 0.5 × pre. F, at 20 and 40, sits below it with a gain of 0.25. H, at 80 and 100, reaches the top of the scale with a gain of 1, though both gained 20 points. The dashed line at 100 is the ceiling: no score passes it.



The same ten made-up learners at three tests. The class mean rose from 46% to 75%, a gain of 29 points, and was 66% a week later: 20 of the 29 points, about 69%, were kept. At the end of the session 5 of 10 met the standard of 80%; a week later, 4 did.



How many of the ten learners got each question right, before and after the session. The counts are made up for the example. Item 1, Q1, was right for 8 of 10 before any teaching. Item 10, Q10, moved only from 2 to 3: read that question again before you teach the topic again.



Hake's survey of 62 physics courses, class normalised gain. The 14 courses that made little or no use of interactive engagement averaged 0.23 (SD 0.04); the 48 that made substantial use of it averaged 0.48 (SD 0.14). These are physics courses on physics tests, not a target for one session in another subject.

ILLUSTRATION. You teach a batch of ten third-year medical students one session. The standard the objective sets is 8 of the test's 10 questions right, which is 80%.

The scores below are made up for this example. The delayed test is one week later, on a parallel form: different questions asking for the same action.

learner	pre (%)	post (%)	delayed (%)
A	20	60	50

learner	pre (%)	post (%)	delayed (%)
B	30	70	60
C	40	80	70
D	50	70	60
E	60	90	80
F	20	40	30
G	70	90	90
H	80	100	90
I	40	60	50
J	50	90	80

Add each column and divide by 10.

```
460 divided by 10 = 46
750 divided by 10 = 75
660 divided by 10 = 66
```

The class went from a mean of 46% to 75%.

```
75 minus 46 = 29
```

A mean gain of 29 percentage points. Now look at two learners inside it.

```
F: 40 minus 20 = 20
H: 100 minus 80 = 20
```

F and H each gained 20 points. Divide each by the room to gain.

```
F: 100 minus 20 = 80
F: 20 divided by 80 = 0.25
H: 100 minus 80 = 20
H: 20 divided by 20 = 1
```

F's normalised gain is 0.25. H's is 1.

Turn the rule round to set a target. What post-test would F have needed for a gain of 0.5? Half of F's room is 0.5 times 80.

```
0.5 times 80 = 40
20 plus 40 = 60
```

A post-test of 60%.

Now the whole class, the way Hake defines it: from the class means.

```
100 minus 46 = 54
29 divided by 54 is about 0.537
```

The class normalised gain is about 0.54. The mean of the ten learners' own normalised gains is a different number. Work out each, add them, and divide by 10: it comes to about 0.59. Say which one you computed.

Next, the spread. The ten raw gains are 40, 40, 40, 20, 30, 20, 20, 20, 20 and 40. Their mean is 29. Four sit 11 above it, five sit 9 below it, and one sits 1 above it. Square and add. Divide by 9, one less than the ten, as Book 0 did for a sample.

```
4 times 121 = 484
5 times 81 = 405
484 plus 405 plus 1 = 890
890 divided by 9 is about 98.9
```

The square root of 98.9 is about 9.9. The SD of the gains is about 9.9 percentage points.

Now the standard. Count the learners at 80% or more.

```
pre: 1 of 10 = 10%
post: 5 of 10 = 50%
delayed: 4 of 10 = 40%
```

Half the batch met the standard at the end of the session. A week later, four did.

The delayed test also shows how much of the gain lasted.

```
66 minus 46 = 20
20 divided by 29 is about 0.69
```

A week on, the class kept 20 of its 29 points, about 69%.

```
100 minus 46 = 54
20 divided by 54 is about 0.37
```

The delayed normalised gain is about 0.37: the room to gain is still counted from the pre-test.

Last, the items. Count, for each question, how many of the ten got it right.

item	1	2	3	4	5	6	7	8	9	10
right, pre	8	7	6	5	5	4	4	3	2	2
right, post	10	9	9	8	8	8	7	7	6	3

Item 1 was right for 8 of 10 before you taught it: that time could go elsewhere. Item 10 moved from 2 to 3. Either the teaching missed it or the question is broken. Read the question again before you teach the topic again.

Hake surveyed 62 physics courses. The 14 courses that "made little or no use of interactive-engagement (IE) methods" reached a class normalised gain of 0.23, with an SD of 0.04. The 48 courses that made "substantial use" of them reached 0.48, with an SD of 0.14.

Figures quoted in this illustration are from [306].

WHERE THIS PICTURE BREAKS. Hake's two averages are for physics courses measured with two physics tests. They are not a target for one session in another subject.

With a short test, one question moves a learner's g a long way. Read a single learner's g with that in mind.

MUST KNOW POINTS FOR YOU.

- Twenty points from 80% is all the room there was. Twenty points from 20% is a quarter of it. Report the normalised gain beside the raw gain whenever you compare learners, batches or sessions.
- The class normalised gain worked out from the class means and the mean of each learner's own normalised gain are different numbers. Say which one you computed, and compare with Hake's figures only the first kind.
- The normalised gain stops working at the top of the scale. If several learners start that high, the test is too easy for them. Report them apart, and next time write harder items that match the objective.
- A gain in one group, tested before and after, is a change, not an effect of your teaching. Call it "the change in this batch", and use a comparison group allocated at random, as in section 2, before you claim more.
- The delayed test measures what lasted. Book it before the session, on a parallel form, and report both.
- Report four things for any session you evaluate. First, the mean pre-test and post-test, with the SD of the gains. Second, the class normalised gain. Third, the share who reached the objective's standard. Fourth, the same at the delayed test. Add the item table.
- Read the item table before you plan the next session. An item most learners got right before you taught it is time you can spend elsewhere.
- Hake's averages were 0.23 for 14 traditional physics courses and 0.48 for 48 courses using interactive methods. Quote them as a rough scale for what a course can do, never as a pass mark for one session.

EXERCISE 1 (design). Plan how you will measure the session you are building for this book.

Write down four things. The objective and its standard. How many items the test has, and what each asks for. When you will give the pre-test, the post-test and the delayed test. The numbers you will report. Then say what you will do if three or more learners score 100% on the pre-test.

EXERCISE 2 (interpretation). Read the abstract of Hake's 1998 survey, cited below. Say what the "±0.14 (std dev)" beside the interactive courses' 0.48 describes, and what it does not tell you about any one student.

PRACTICE. Fourteen problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. pre = 35 and post = 62. Work out the raw gain and the normalised gain.

2. Learner P scores 70% before and 85% after. Learner Q scores 20% before and 60% after. Work out each learner's raw gain and normalised gain.

3. Five learners' raw gains are 10, 20, 15, 25 and 30 percentage points. Work out the mean gain and the SD of the gains.

4. pre = 40 and g = 0.25. What is post?

5. post = 85 and g = 0.75. What was pre?

6. Hake's 14 traditional physics courses averaged a class normalised gain of 0.23. His 48 courses using interactive methods averaged 0.48. Take a class whose mean pre-test is 45%. What mean post-test does each gain give it, and how many percentage points apart are the two?

7. In Deslauriers and colleagues' study, one group of students (group A in the spring semester) averaged 23.3 on a physics pre-test, the Force Concept Inventory, scored out of 30. Write that as a percentage. How many percentage points of room did that group have? What raw gain, in percentage points, would equal Hake's interactive-course average of 0.48, and what post-test would that be?

8. Twelve accredited social health activist (ASHA) workers take a 10-item test before and after a session. The objective's standard is 7 items right, or 70%. Their scores, made up for this problem, in %, are:

worker	1	2	3	4	5	6	7	8	9	10	11	12
pre (%)	40	50	30	60	70	20	50	40	60	30	80	50
post (%)	70	80	60	90	90	50	70	60	80	70	90	60

What share reached the standard before, and after? By how many percentage points did the share change?

9. Twenty residents take a 5-item test before and after a journal club on reading a trial's results table. The number who got each item right, made up for this problem:

item	1	2	3	4	5
right, pre	17	6	4	5	9
right, post	19	16	15	6	18

Work out the proportion correct on each item before and after, and the change in percentage points. Say which two items you would act on, and what you would do.

10. A batch's mean pre-test is 38%. Its mean post-test at the end of the session is 74%. Its mean delayed test, a week later on a parallel form, is 61%. The three scores are made up. Work out the raw gain and the class normalised gain at the post-test. Do the same at the delayed test. Then work out the share of the post-test gain still there a week later.

11. Here is a worked answer. Find the step that broke.

```
pre = 50, post = 80
raw gain = 80 minus 50 = 30
room to gain = 100 minus 80 = 20
g = 30 divided by 20 = 1.5
the learner did better than gaining everything there was to gain
```

12. Here is a worked answer. Find the step that broke.

```
four learners, pre and post (%): 90 to 100, 20 to 30, 30 to 40, 40 to 50
their normalised gains: 1.0, 0.125, 0.143, 0.167
mean normalised gain = 1.435 divided by 4 = 0.36
Hake's traditional courses averaged 0.23
so this session did better than the average traditional course
```

13. A colleague reads Hake's abstract and says: "Interactive courses were almost two standard deviations better. So any student taught interactively gains about twice as much as one taught traditionally."

Decide what to compute, compute it, and say what the abstract does not establish.

14. A made-up report on a training day for ASHA workers says: "Before the session, 60% of workers passed the test. After it, 90% did. The session improved performance by 50%."

Decide what to compute, compute it, and say what the claim does not establish.

13 · Liking a session is not learning from it

DEFINITION. To evaluate teaching is to measure some outcome of it. Kirkpatrick's four-level model sorts those outcomes by what is measured. So each level names a different response variable: the thing the teaching is meant to change. As Frich and colleagues (2015) set it out, the four levels are reaction (Level 1), knowledge (Level 2), behavioural change (Level 3) and system results (Level 4).

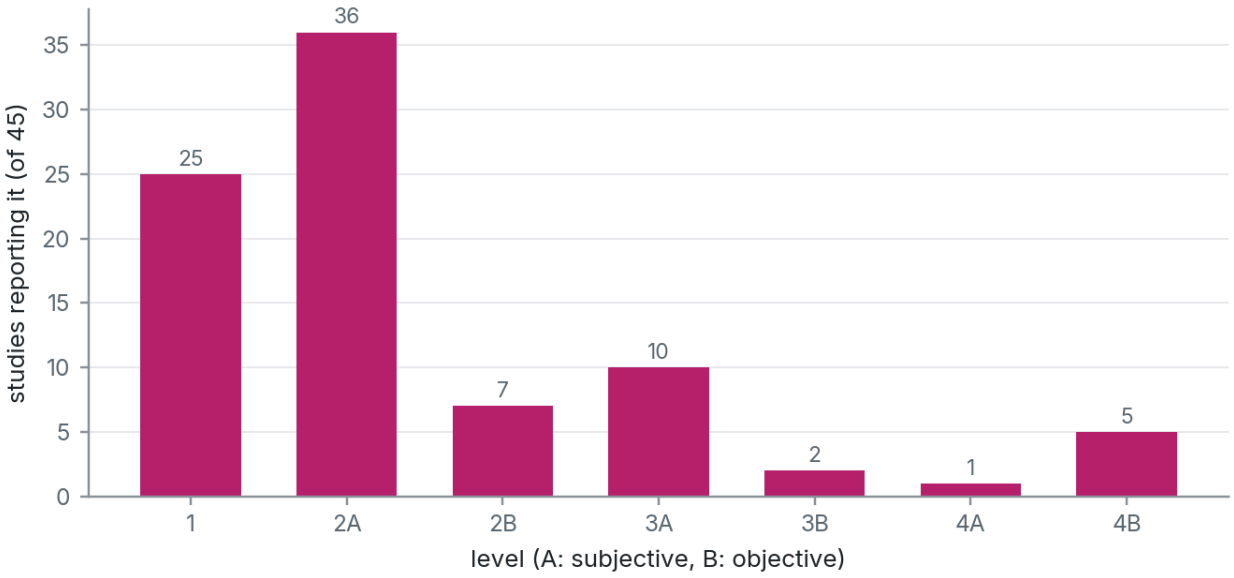
Frich and colleagues split Levels 2, 3 and 4 by how the outcome is measured. Subjective means stated by the learner or someone near them, as an opinion or a rating (2A, 3A, 4A). Objective means measured by a test, an observation or a record (2B, 3B, 4B).

A feedback form given at the end of a session measures Level 1. An item on it such as "I learned a great deal" measures Level 2A. A test that asks for the objective's action, scored against a key, measures Level 2B. [307, 308, 309, 310, 311, 312, 313, 311, 313, 314, 315, 316, 2]

IN PLAIN TERMS. At the end of a session you hand out a feedback form.

Ask what the form measured. It measured whether the learners liked the session. There are three more. Did they learn what you set out to teach? Do they now do their work differently? Did anything change for the people they serve?

These four questions are called Kirkpatrick's levels. Each needs its own measure. A form answers the first. A test answers the second. Watching people at work answers the third. Records of the service answer the fourth.



Frich and colleagues' review of 45 studies of leadership programmes for physicians: how many studies reported an outcome at each level. One study can report several levels. 36 studies reported self-assessed knowledge; 7 tested it. 2 measured behaviour at work objectively. 6 reported system results: 1 as perceived, 5 measured objectively.

Illustration 1. Here is the trap. You teach a two-hour session to a batch of accredited social health activist (ASHA) workers. The objective: each worker weighs a child and plots the weight on the child's growth chart correctly. At the end you hand out the usual feedback form. Nearly every form ticks "agree" or "strongly agree". Your report says the session was a success.

Now sort the form. Take each item and ask what it measures.

item on the form	what it measures	level
The session was useful.	liking the session	1
The trainer was clear and well prepared.	liking the trainer	1
I learned a great deal from this session.	own view of learning	2A
I feel confident plotting a child's weight.	own view of learning	2A
I will use this in my work.	an intention	1

Nothing on the form reaches Level 2B. The last item is not Level 3 either. Frich's levels do not name intentions. This book files it with reaction: it is a promise made in the room, before the worker is back at work.

The third item is worth a second look. It is close to the question Deslauriers and colleagues used to measure students' feeling of learning: "I feel like I learned a great deal from this class." The first section showed that feeling moving the wrong way.

Now add what each higher level would need.

- Level 2B: at the end of the session, and again a week later, give each worker three children's weights to plot. Mark each plot right or wrong against a key. That is the kind of test, given more than once, that the section on measuring change scored.
- Level 3B: a month later, a supervisor watches each worker weigh and plot one child at work, with a short checklist.
- Level 4B: over months, the service's own records show whether children whose weight stops rising are found sooner. A single session is unlikely to move this alone, and you would need a comparison area to tell.

The form was not wrong. It answered the question it asked. Your report asked a different one.

WHERE THIS PICTURE BREAKS. The form and the batch are made up for this example. The form's items are typical of end-of- session forms, not copied from any one.

Sorting an item by level says what it measures, not whether it measures it well. A Level 2B test with items that do not match the objective is still a poor test.

Illustration 2. Now read a real literature the same way. Frich and colleagues (2015) reviewed 45 published studies of leadership programmes for physicians. Their abstract says positive outcomes "were reported in all studies". Before you repeat that, ask what the positive outcomes were measured at.

Their Results count how many studies reported an outcome at each level. One study can report several levels, so the counts add to more than 45.

level	what was measured	studies reporting it (of 45)
1	reaction	25
2A	knowledge, self-assessed	36
2B	knowledge, tested	7
3A	behaviour at work, self-assessed	10
3B	behaviour at work, measured objectively	2
4A	system results, as perceived	1
4B	system results, measured objectively	5

Turn two of those counts into proportions of the 45 studies.

```
36 divided by 45 = 0.8
7 divided by 45 is about 0.156
```

Eight in ten studies reported what physicians said they had learned. About one in six tested it. Two measured behaviour at work objectively, for example with a checklist scoring a team's performance.

Now read the abstract's sentence to the end: "although the majority of studies relied on learner satisfaction scores and self-assessed knowledge or behavioral change." The positive outcomes rest mostly on Levels 1 and 2A. The authors say so themselves. Only six studies reached system results.

One more thing to check before you quote it. The same review says most assessments "occurred immediately after the program". A test given the moment a course ends measures performance, not what lasted.

Figures quoted in this illustration are from [317].

WHERE THIS PICTURE BREAKS. Frich's counts describe published studies of leadership programmes for physicians, found in one database up to November 2013. They say what that literature measured. They are not a count of how teaching in India, or any teaching session, is evaluated.

The counts also say nothing about how good each study was at its level. A Level 4 result from one programme with no comparison group is still only a change, not an effect.

MUST KNOW POINTS FOR YOU.

- A warm set of feedback forms tells you the learners liked the session. It does not tell you they can do anything new.
- An item that asks "Did you learn a lot?" or "Do you feel confident?" is still the learner's opinion, Level 2A. When you read "participants' knowledge improved", find out which of the two it was.
- This book asks you to show that your learners can do something they could not do before, and that you measured it. Any test you score against a key is Level 2B, even one at

the end of the session. This book asks for more: give it a week or more after the session, because section 1 counts only what lasts. Plan it before you teach.

- Ratings and a test score at the end of class can point opposite ways. So never switch methods, or judge a colleague's teaching, on ratings alone.
- "Every study reported positive outcomes" means little until you know the level. Ask any programme report which level its good news came from.
- The levels say what an evaluation measured, not what caused the result. Before you credit the course, ask for the design as well as the level.
- Level 1 is not worthless. It tells you whether learners could hear, kept up, and would come back, and a session nobody attends teaches nothing. Use the form for those questions and a test for learning. Do not rewrite a session on the form alone.

EXERCISE 1 (critique). A department's report on its teaching sessions for second-year MBBS students reads: "Feedback was collected from all students. The mean rating for 'I have learned a great deal' was high, showing the sessions achieved their learning objectives."

Sort what was measured by level. Say what the conclusion needs that the report does not have, and what you would add next year.

EXERCISE 2 (interpretation). Frich and colleagues' abstract says "four studies used a comparison group". Their Results section says "Only five studies" did, and their Table 2 gives "Comparison group" as 4, or 9%. You want to quote the review in your own write-up. What do you write, and why?

EXERCISE 3 (design). Take the session you are planning for this book's build. Write one line for each of the four levels: what you would measure, how, and when. Mark which lines you will actually carry out and which you cannot, and say why.

EXERCISE 4 (teaching). For a health secretary. A state programme reports that its refresher training for ASHA workers on child nutrition "was rated excellent by participants". The secretary asks whether the training worked and whether to fund it again. One page, answer first.

14 · Where teaching sits in Indian medical education

Checked against the sources as they stood on 2026-09-25.

DEFINITION. Undergraduate medical education in India is governed by the Graduate Medical Education Regulations, 2023 (GMER 2023). The National Medical Commission (NMC) made them under the National Medical Commission Act, 2019, sections 24(1) and 57(2). NMC notified them in the Gazette on 2 June 2023. They direct NMC's undergraduate board to publish the model curriculum on the NMC website from time to time.

The curriculum in force is NMC's Competency Based Medical Education (CBME) Curriculum 2024, enclosed with a letter of 12 September 2024. It calls itself guidelines, and it was issued by letter. GMER 2023 requires every student to be trained as the board's published guidelines say. So the guideline binds through the regulation.

The CBME Curriculum 2024 names seven roles for the Indian Medical Graduate (IMG): clinician; leader and member of the health care team and system; communicator; lifelong learner; professional; critical thinker; and researcher. The Regulations on Graduate Medical Education (Amendment), 2019 named the first five.

The CBME Curriculum 2024 limits large group teaching to one third of a subject's allotted hours. It counts any large group method, an interactive lecture included. Its definition sets no group size. Phase I opens with a Foundation Course of two weeks and 80 hours.

The Medical Institutions (Qualifications of Faculty) Regulations, 2025 were notified on 30 June 2025. They superseded the Teachers Eligibility Qualifications in Medical Institutions Regulations, 2022. Their Table E covers faculty in the broad specialties: the subjects of the Doctor of Medicine and Master of Surgery degrees. There a Professor or an Associate Professor "shall be required to undergo" the Basic Course in Medical Education. The proviso is that the specialty is covered under undergraduate training. [318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345]

IN PLAIN TERMS. At the top sits a regulation, the Graduate Medical Education Regulations, 2023 (GMER 2023). It was notified in the Gazette, so it binds.

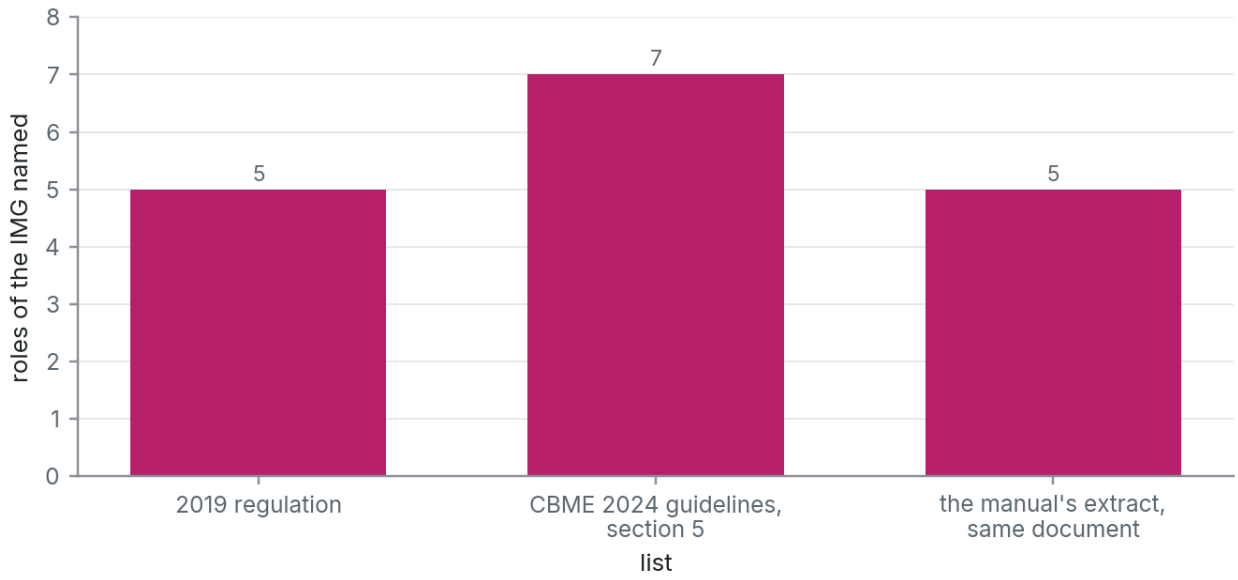
Under it sits the detail, in a guideline — a document NMC's undergraduate board sent out by letter and put on its website. The regulation says every student is trained as that guideline says, so the guideline binds too. "Competency based" means the course is built from what a graduate must be able to do, not from a list of topics to cover.

That guideline names the doctor the course is meant to produce: the Indian Medical Graduate, or IMG. It gives the IMG seven roles.

Teaching to a large group, by any method, may take at most a third of a subject's hours.

A third paper decides what you need to hold as a teacher. It is the faculty regulations of 2025. They tie a course, the Basic Course in Medical Education, to two posts: Associate Professor and

Professor. This holds in subjects of the Doctor of Medicine and Master of Surgery degrees that are taught to undergraduates. Community Medicine is one of them.



How many roles of the Indian Medical Graduate each list names. The 2019 regulation named five. The CBME Curriculum 2024's guidelines, section 5, name seven, including critical thinker and researcher, which were added after 2019. The manual inside the same document says it copies the guidelines, yet still lists five. Teach seven.

Illustration 1. You are asked to take the Foundation Course session on the roles of the Indian Medical Graduate for a new batch of first-year MBBS students. A senior hands you last year's slides. They show five roles.

Open NMC's CBME Curriculum 2024 at https://nmc.org.in/storage/cms/rules-regulation-nmc/YT-W423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf. Find Goals for the Learner in the guidelines at the front, section 5. Read the list. It runs from a to g: clinician; leader and member of the health care team and system; communicator; lifelong learner; professional; critical thinker; researcher. That is seven.

Now search for "Roles (global competencies)". It takes you further down the same document, to the curriculum volume's "How to use the Manual". Its Section 1 says it gives the roles "extracted from" the CBME Guidelines, 2024. Read its list, under 2.3. It stops at e, professional. That is five.

So the slides were not made up. The five are exactly the roles the 2019 regulation named. Critical thinker and researcher came later, and one copy inside NMC's own document was never brought up to date.

Now decide which to teach. The manual says it is copying the guidelines, so the guidelines are the thing to quote. Teach seven. Then show the batch the second list, and tell them why a copy can disagree with what it copies.

Figures quoted in this illustration are from [346, 347].

WHERE THIS PICTURE BREAKS. This settles what NMC's document of 12 September 2024 says. It does not tell you whether a later letter has changed the list. Before the session, open NMC's

Rules and Regulations page, https://nmc.org.in/page/rules-regulations-rules-regulations-nmc, and look for anything under the undergraduate board dated after 12 September 2024.

Illustration 2. You will hear it said that every medical teacher must do the Basic Course in Medical Education. Check it against the instrument.

Open the Medical Institutions (Qualifications of Faculty) Regulations, 2025 at https://nmc.org.in/storage/cms/rules-regulation-nmc/QZSDeGu7WXS9MMI0dL5ZR1EbS48kLaFbvMTpww2B.pdf. The first paragraph says they are made "in supersession of" the teachers' eligibility regulations of 2022. Supersession means the new instrument takes the old one's place. So anything you learnt from the 2022 regulations needs checking again.

Go to the Schedule and find Table E, for faculty in the broad specialties. Read it one row at a time and write down what each row says about the course.

The Professor row says the Professor "shall be required to undergo Basic Course in Medical Education provided their broad specialty subject is covered under undergraduate training". The Associate Professor row says the same. The Assistant Professor row does not name the course at all.

Now apply the proviso to yourself. Is Community Medicine covered under undergraduate training? Search the CBME Curriculum 2024 for Community Medicine. It is one of the subjects of the Phase III Part 1 university examination. So yes. Under this table, the course reaches you at Associate Professor.

Go back to the Professor row and read it again, slowly. The research course is worded "have completed". The teaching course is worded "shall be required to undergo". Those are different words. May you take the teaching course after the promotion? For one route the table says so. Its Note 2 lets an experienced non-teaching consultant, specialist or medical officer become an Associate Professor. That person must take the teaching course "within two years of the appointment". For the ordinary route the table does not say. Do not settle that from the wording. Read NMC's published FAQs on these regulations. If they do not answer it, ask NMC in writing.

WHERE THIS PICTURE BREAKS. This reads one table of one regulation, as notified on 30 June 2025. Dental faculty have their own table, which words the course differently. The table does not tell you what your own college or state requires on top of it. And an amendment or a published answer from NMC can change the reading.

MUST KNOW POINTS FOR YOU.

- "The Indian Medical Graduate has five roles" is out of date. The CBME Curriculum 2024 lists seven: critical thinker and researcher were added to the 2019 five.
- When a department asks you to lecture, count its large-group hours first. Over one third of a subject's hours in large groups puts a timetable out of line with the curriculum. Good lectures do not change that.
- Mentoring is in the curriculum, not an optional extra. Each mentor gets three students during the Foundation Course and keeps them until they finish internship. Mentors are drawn from Assistant Professor upwards, so expect to be one from your first post as Assistant Professor.

- The Basic Course in Medical Education is tied to posts, not to teaching as such. Its Assistant Professor row does not name it. Say that, not "every teacher must do it".
- Know which kind of paper you are quoting. Mahajan and Gupta (2024) report that a circular of 12 June 2023, issuing guidelines under GMER 2023, was withdrawn eleven days later. The regulations themselves were not. A CBME guideline of 1 August 2023 was later revised by the 2024 curriculum. So "the 2023 guidelines" can mean either paper. Ask for the date, and check which one it is.
- Two papers in force can disagree. GMER 2023 bars a student from continuing after nine years from admission. The CBME Curriculum 2024 says a learner may not graduate later than ten years from joining, internship included. When a student's case turns on the limit, name both papers and their words.
- These instruments tell you what a college must provide and what a teacher must hold. They say nothing about whether any session works.

EXERCISE 1 (retrieval). Name the seven roles of the Indian Medical Graduate in the CBME Curriculum 2024. Say which two were not among the five named in 2019.

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (interpretation). A new Assistant Professor of Community Medicine is told by a colleague: "You cannot teach until you have done the Basic Course in Medical Education. It is in the 2022 regulations." Using only what this section gives, say what is wrong with each half of the sentence, and what you would check before saying anything more.

EXERCISE 3 (teaching). For a first-year resident, ten minutes with a whiteboard. The CBME Curriculum 2024 caps large group teaching at one third of a subject's hours. What does the cap force a department to do, what argument from this book supports it, and what does the cap not guarantee?

15 · Leading people is a craft of its own

DEFINITION. To lead people is to get a group to move towards a change and to keep them at the work it takes.

Stoller, reviewing physician leadership, argues that skill at the work and skill at leading others in it are different competencies. Kotter (1990) argued that leadership has nothing to do with charisma or other exotic personality traits, and is not the province of a chosen few.

Leadership is taught to physicians, and the teaching has been evaluated. Frich and colleagues' systematic review found 45 published studies of leadership development programmes for physicians, searched in one database up to November 2013. All 45 reported positive outcomes. Four had a comparison group (the Results text says five; see section 13).

That such training changes what physicians do at work, or the results of the service they work in, has been tested in very few studies.

In this book, a teaching lineage means the people you teach, and the people they go on to teach. A team you run means a set of people whose work on one job you direct, for as long as the job lasts. An institution that outlives you means a standing group, with written terms, roles and a named successor, that keeps working after you leave it.

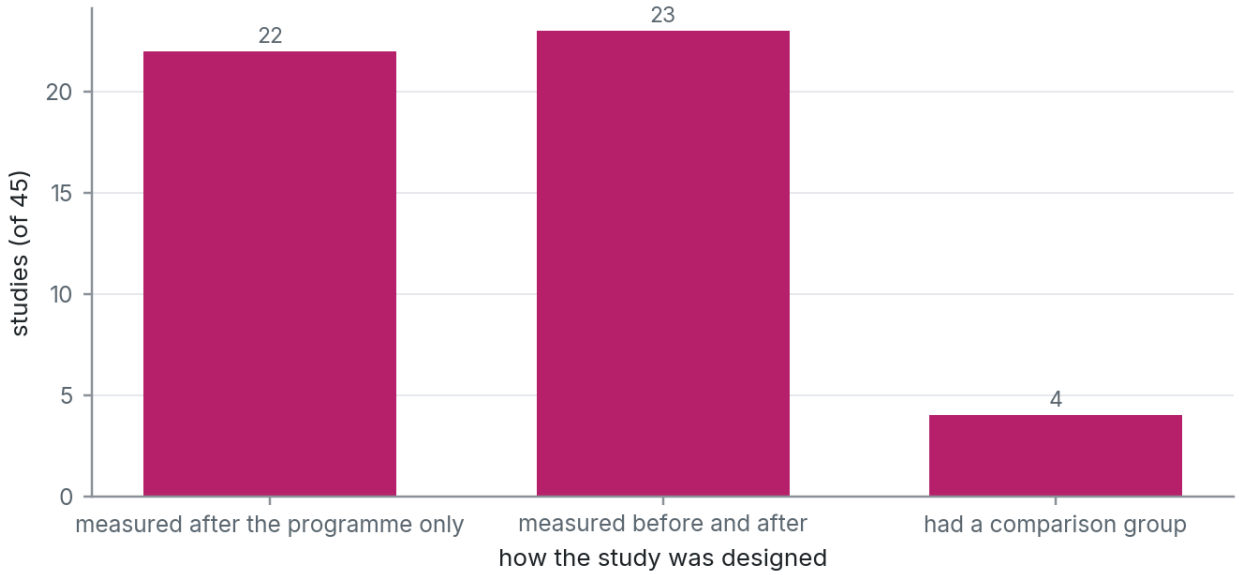
In this book, a professional body means an association whose members share one profession or discipline. A civil society organisation means a group formed by citizens, outside government and outside business, to work for a public cause. [348, 349, 350, 351, 352, 353, 312, 311, 354, 355, 356, 357, 358, 313, 359, 360, 361, 362, 363]

IN PLAIN TERMS. Here is the usual path. The resident who knows the most is made chief resident. Everyone assumes the skill carries over. It need not. Stoller argues that knowing the work and getting others to do it well are different skills.

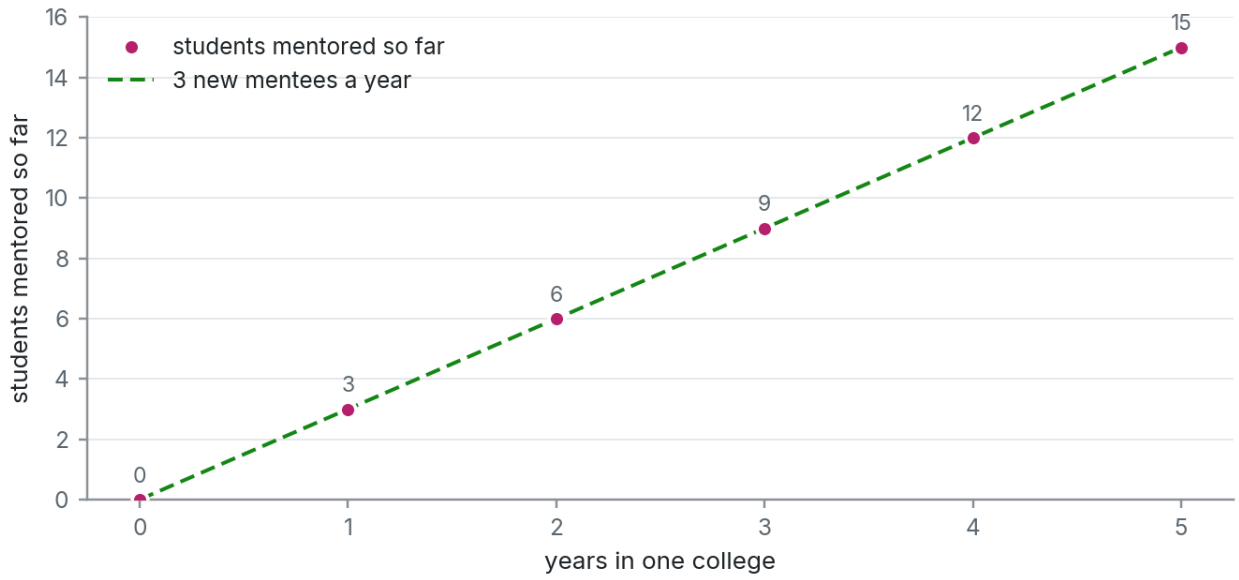
So is leading a gift you have or do not have? Kotter says no: it is not the province of a chosen few. Companies find people with the potential and develop it through early challenge. He adds that not everyone becomes a strong leader.

Many programmes teach leadership to doctors, and every study in one large review reported success. Most of that success is what the doctors said about themselves. It was mostly measured straight after the course. Leadership is taught, and doctors say they learn from it. Whether it changes what they do is barely tested.

You will be asked to build three different things. To tell them apart, ask one question: what happens to it the day you go?



How the 45 studies in Frich and colleagues' review were designed. 22 measured after the programme only and 23 before and after; 22 plus 23 = 45. The 4 with a comparison group are among those 45, so 41 had nothing to compare against.



Mentees under the CBME Curriculum 2024: three new first-year students each year. Stay five years in one college and 3 × 5 = 15 students have been yours. The five years are the example's; the three a year are the curriculum's.

Illustration 1. Here is the trap. A senior colleague tells your department meeting: "Leadership can be taught. A review of 45 studies found that every leadership programme for doctors worked."

Open the review, Frich and colleagues, "Leadership development programs for physicians: a systematic review", *Journal of General Internal Medicine* 2015, at https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/. Search the page for "Positive outcomes".

> *Positive outcomes were reported in all studies, although the majority of studies relied on learner satisfaction scores and self-assessed knowledge or behavioral change.*

So your colleague is right that every study reported a positive outcome. The section on liking a session sorted these same outcomes by level, and most sat at the bottom two. Now ask a different question: what was each study compared with? Find Table 2 and read its rows under "Evaluation design".

how the study was designed	studies (of 45)
measured after the programme only	22
measured before and after	23
had a comparison group	4

The first two rows cover every study once.

```
22 plus 23 = 45
```

The four with a comparison group are among those 45, not extra to them. That leaves 41 studies with nothing to compare against.

```
45 minus 4 = 41
```

A study measured after the programme only cannot show a change. A study measured before and after shows a change, but not that the programme caused it: the section on testing a teaching method set out why. Only the four could compare trained physicians with untrained ones.

Now search for "publication bias". The authors expect that studies finding no effect were less likely to be published. So "every study worked" is partly a fact about what gets printed.

Here is what you can say instead. Leadership programmes for doctors are common. Doctors report learning from them. Whether the programmes change what doctors do at work has been tested in very few studies.

Figures quoted in this illustration are from [317].

WHERE THIS PICTURE BREAKS. The counts describe published studies of leadership programmes for physicians, found in one database up to November 2013. A newer review may find more comparison-group studies: search for one before you quote these numbers.

Weak evidence that training works is not evidence that it does not. The claim is untested more than it is refuted.

Illustration 2. Picture your first year as an Assistant Professor of Community Medicine. By its end, three things are in your hands.

The first is your mentees. The Competency Based Medical Education (CBME) Curriculum 2024 gives each mentor three first-year students during the Foundation Course. They stay with you

until they finish internship, and three more are added each year. Stay five years in one college and the count is easy.

```
3 times 5 = 15
```

Fifteen students have been yours.

The second is a survey team. For one month you run four interns and two accredited social health activists, weighing and measuring children in two villages.

The third is the department's monthly journal club, which you started.

Now ask each one the same question. What happens to it the day you leave?

- The mentees carry what you taught them. If some of them teach it to their own juniors, it goes on spreading without you. That is a teaching lineage. It is measured in people, over years, and it grows only through teaching and time with each person.
- The survey stops, or passes to whoever runs it next. That is a team you run. It exists for one job. It needs direction, roles and work handed out, and the next section gives you the words for those.
- The journal club either stops, or keeps meeting because it has written rules, a rota of convenors and a named successor. Only in the second case is it an institution that outlives you. Building one means making yourself unnecessary to it, on purpose.

Kotter, writing about companies, puts the last of these highest: "Institutionalizing a leadership-centered culture is the ultimate act of leadership." That is his judgement, not a finding. It does tell you which of the three is hardest to fake.

Figures quoted in this illustration are from [346].

WHERE THIS PICTURE BREAKS. The survey team, the journal club and the five years are made up for this example; the mentor numbers are the curriculum's.

The three are not always separate. A team you run well can turn into a lineage, and an institution needs a team to run it in any given year. The question "what happens when I leave?" sorts what you have built. It does not tell you how to build it.

MUST KNOW POINTS FOR YOU.

- Being the best at the work does not make you able to lead others in it. When you are chosen to run something because you know the most, expect to have to learn a second skill.
- Do not say leadership programmes are shown to change practice. In Frich's review, 2 of 45 studies measured behaviour at work objectively, and 4 had a comparison group (the Results text says five; see section 13).
- You lead before you have a title. Take the small chances now: run the journal club, lead the camp, own one change on the ward.

- Ask of anything you run: what happens to it the day I leave? If the answer is “it stops”, you have built a team, however long it has run. If you want it to last, write its rules and name a successor while you are still there.
- Your first lineage is handed to you. From your first post as Assistant Professor you can be made a mentor. You get three first-year students a year, and they stay with you until they finish internship.
- Join one professional body and one civil society organisation, and be useful to both. This course asks it of you at the next rung. A body gives you peers who check your work. An organisation outside government gives your evidence a use. The next book names the bodies.
- The claim that leading can be learned rests on argument and on programme reports, not on trials. Judge a course as you would judge a teaching session: ask what was measured, at what level, and against what comparison.

EXERCISE 1 (retrieval). Name the three things a leader is asked to build, as this section defines them. Give the one question that tells them apart, and what the answer is for each.

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). The Dean says at a committee meeting: “Dr Rao is the best physician we have. She has published more on obesity than anyone in the state. Make her head of the new obesity clinic and it will run itself.” Find the step in the argument that does not follow, and say what you would want to know before agreeing.

EXERCISE 3 (teaching). For a journalist, three minutes, no jargon. The journalist asks: “Can you teach doctors to be leaders, or are leaders born?” Give one quotable sentence, and be honest about how strong the evidence is.

16 · Words for how a team works

DEFINITION. Kotter (1990) sets management and leadership apart as two different and complementary systems of action. Management copes with complexity. Leadership copes with change.

Management has three activities. Planning and budgeting is setting targets, laying down detailed steps to reach them, and allocating resources to those steps. Organizing and staffing is creating a structure of jobs, filling the jobs with qualified people, communicating the plan to them, delegating responsibility for it, and setting up systems to monitor it. Controlling and problem solving is monitoring results against the plan, identifying where they depart from it, and planning and organizing to fix that.

Leadership has three matching activities. Setting a direction is developing a vision of the future, often the distant future, with strategies for making the changes it needs. Aligning people is communicating that direction to everyone who can help carry it out or can block it, so that they understand it and are committed to it. Motivating and inspiring is keeping people moving in that direction, despite obstacles, by appealing to their needs, values and emotions.

Kotter holds that an organisation needs both, each balancing the other.

The words below are this book's working definitions, not Kotter's, and not findings. The next book in this subject gives the evidence on each. In this book, authority is the right to decide or to direct that comes with a post. Power is the capacity to get things done through other people, from any source: authority, expertise, relationships, reputation, or control of something others need.

In this book, delegation is handing a task, and the authority it needs, to another person, while you stay answerable for the result. How much direction and checking you give is matched to that person's competence and commitment for that task, not to their seniority in general.

In this book, role clarity is a written statement, for each decision or piece of work, of who decides, who does the work, and who is asked first. The statement can add one more line: who tells whom when the work goes wrong. That line shows how feedback travels in the team.

In this book, conflict about the work is disagreement over what to do, how to do it, or on what evidence. Conflict about the person is friction over who someone is: blame, dislike, a remark about their character or motives. [350, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377]

IN PLAIN TERMS. Kotter's short way of saying it: management copes with complexity, leadership copes with change.

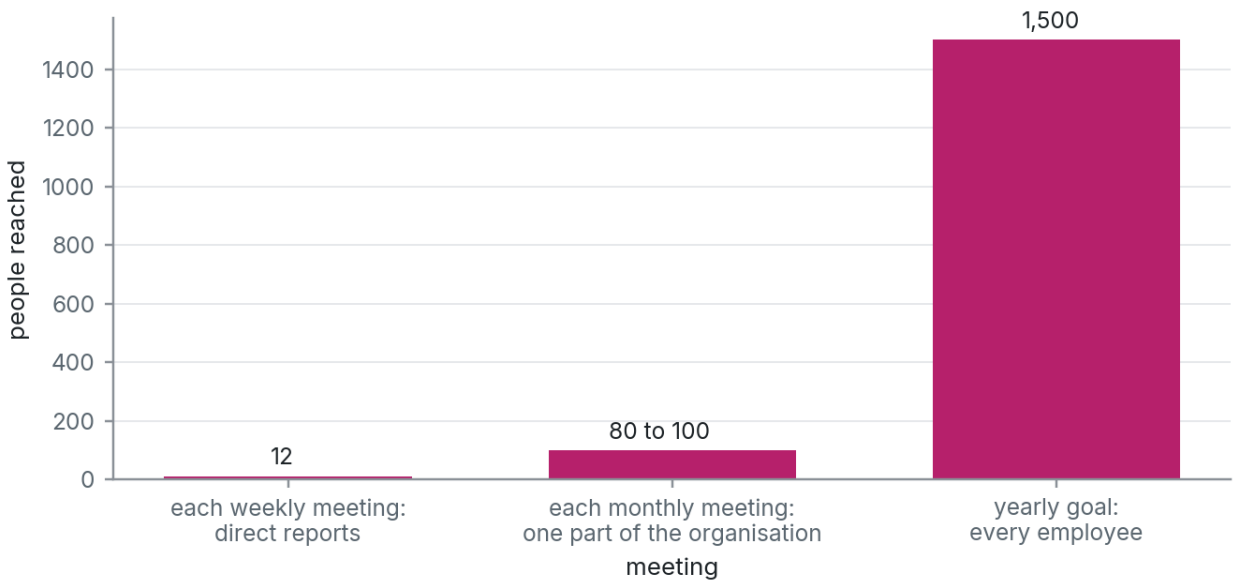
People follow a direction only if they believe the person giving it. Kotter calls that credibility. It rests on your record, and on whether what you do matches what you say.

In this book, authority comes with the post. The senior resident everyone trusts may have more power over the ward than the unit head.

The same intern may need close watching on one job and none on another.

Role clarity means writing down, for each decision, who makes it, who does the work, and who is asked first. When a team argues over whose job something was, look for the missing line on that sheet first.

Last, two kinds of disagreement. They are named apart because they are handled apart. How to handle each is for the next book in this subject.



How many people Crandall's aligning reached, in Kotter's account: his 12 direct reports at each weekly meeting; 80 to 100 people at each monthly meeting (drawn at the upper end, 100); and a goal of all 1,500 employees in a small meeting each year. The first two bars count one meeting, the third a whole year. Kotter's sentence: aligning talks to many more people than organizing does.

Illustration 1. Here is a week in the life of a head of a Community Medicine department. The department has decided that every MBBS student's family-study visits will now include weighing each adult and measuring their waist. The eight things below are made up for this example. Before you read the right-hand columns, cover them and name each activity yourself.

what the head of department did	Kotter's name	job
Wrote one page on what the department should teach about weight in five years	setting a direction	leading
Put the new visits into the term's timetable, and asked for scales in the budget	planning and budgeting	managing
Gave each tutor a group of students and a village	organizing and staffing	managing
Met the medical officer of the primary health centre the students visit	aligning people	leading
Asked the Medicine department to take the students' referrals	aligning people	leading

Asked the tutors how they would run the next round, and used their answer	motivating and inspiring	leading
Checked each month how many families each group had measured	controlling and problem solving	managing
Found why one group skipped the waist measurement, and changed the form	controlling and problem solving	managing

Notice three things.

The same person did all eight. Leading and managing are two jobs, not two people.

Two of the leading rows are about people the head of department has no authority over: the medical officer and another department. Kotter's point about aligning is exactly this. Anyone who can help or block the change is someone you must talk to.

Now imagine the plan failed. The timetable was right, the scales arrived, and the groups were counted every month. But the medical officer was never met, and the families were not told why students were at their door. That is a well-managed change that nobody led. The names let you say so in one sentence, instead of "it just did not work".

WHERE THIS PICTURE BREAKS. Kotter's six names sort activities; they do not rank them. A department that only leads, and never checks results against its plan, fails as surely as one that only manages.

Some activities sit on the line. Asking the tutors for their views is motivating if it gives them a real say, and only a courtesy if you have already decided. The label depends on what you do with the answer.

Illustration 2. Now watch aligning at the scale Kotter wrote about. His article describes Bob Crandall, who ran engineering and manufacturing in Eastman Kodak's copier business in the mid-1980s, and who set out to change how it worked.

Crandall met his own 12 direct reports every week. Once a month he and all of them met 80 to 100 people from one part of the organisation, to discuss anything they wanted. He then set a goal: to get all of his 1,500 employees into at least one small meeting on a specific topic each year.

Kotter counts every one of these as a way of aligning people. Notice how far they reach beyond the 12 who reported to Crandall: 80 to 100 people a month, and then all 1,500. Kotter's sentence for it: "Aligning invariably involves talking to many more individuals than organizing does."

Scale it down to your own department. Organizing the family-study visits means talking to your tutors. Aligning them means talking to the students, the medical officer, the accredited social health activists in the village, and the Medicine department too.

Figures quoted in this illustration are from [378].

WHERE THIS PICTURE BREAKS. This is Kotter's account of one company's change, written to illustrate his argument. It reports that quality and costs improved afterwards. With nothing to compare, that does not show the meetings caused it. Its numbers belong to an organisation of

1,500 people in the 1980s. A department of fifteen does not need a monthly forum; the principle, not the format, is what carries over.

Illustration 3. You run the department's monthly journal club. Write its role sheet before the first meeting, not after the first muddle.

what	who decides	who does it	who is asked first
choosing the paper	the month's convenor	the presenter proposes three	the head of department
presenting it	the month's convenor	the presenter	the convenor, a week out
booking the room, telling staff	the month's convenor	the department office	nobody

Now delegate one task. The presenter for March is an intern who has never read a trial critically. The presenter for April is a senior resident who has run three journal clubs. The task is the same. The delegation is not. With the intern, you go through the paper together once and hear a practice run. With the senior resident, you agree the date and look at the slides the day before. Next month the intern may present a topic they know well, and need almost nothing from you.

Then, at the April meeting, two things happen. Two residents argue over whether the trial's comparison group was fair. That is conflict about the work, to be settled on the evidence. Later, one of them says the other "only picks papers to look clever". That is conflict about the person. It has left the paper behind. Name which one you are in before you decide what to do.

WHERE THIS PICTURE BREAKS. The journal club and its people are made up for this example.

A role sheet makes the roles clear. It does not make them fair, and it does not make anyone keep to them. And a disagreement can start about the work and turn into one about the person halfway through a sentence. The names help you notice the turn; they do not stop it.

MUST KNOW POINTS FOR YOU.

- Leading and managing are two sets of activities. Kotter argues that an organisation needs both done, by one person or by different people on one team. When a change stalls, ask which half was missing.
- A detailed plan is not a direction. Before you write the timetable for a change, write in a few sentences where it is going and why.
- Aligning means talking to everyone who can help or block the change, not only the people who report to you. List them before you start, and meet the ones you have no authority over first.
- When people do not do what you ask, find out whether you lack authority or lack power. An order from the post does not reach a department you do not head. Trust built over time can reach people no post does. The fix for each is different.
- Delegate to the person and the task, not to the rank. Before you hand over a task, ask how well this person has done this task before, and how much they want it.
- Write down who decides, who does the work and who is asked first, before the work starts, so that when something is not done, you can see whose line was missing.

- Do not answer a disagreement about the work as if it were a personal attack, or the reverse. This book's stance: settle a disagreement about the work on the evidence. Whether such disagreement helps a team is for the next book.
- These are words, not a diagnosis. Using the words does not show that a team will do better.

EXERCISE 1 (retrieval). Write Kotter's three management activities and his three leadership activities, each management activity beside the leadership activity that matches it.

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (interpretation). These five sentences come from the notes of a meeting after a school weight-screening camp run by a Community Medicine department. For each, name the word from this section that fits best, and say why.

1. "The Principal put me in charge of the camp, so the Paediatrics department had to send two residents. They did not come."
2. "Nobody knew who was meant to book the school hall."
3. "Anita and Ravi disagreed about whether to screen with waist measurement or with body mass index."
4. "Ravi said Anita always sides with whoever is senior."
5. "We gave the new intern the same one-line instruction as the senior resident, and her forms were all incomplete."

EXERCISE 3 (design). Write the role sheet for the teaching session you are planning for this book's build. Cover at least four decisions or pieces of work: for example, the objective, the time slot, the tests, and the teaching itself. For each, name who decides, who does it, and who is asked first.

EXERCISE 4 (teaching). For a first-year resident, ten minutes with a whiteboard. "What is the difference between leading and managing, and why would a unit need both?"

References

1. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Significance. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

2. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Abstract. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

3. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Results and Discussion, the paragraph on analysing at the student or group level. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

4. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Results and Discussion, the paragraph on Table 3, model 1. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

5. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Results and Discussion, the paragraph on Table 3, model 2. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

6. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Materials and Methods, the paragraph describing the control and experimental groups. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

7. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Materials and Methods, the paragraph defining traditional lecturing. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

8. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Materials and Methods, the sentence after the definition of traditional lecturing. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

9. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Results and Discussion, the paragraph proposing three causal factors. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

10. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Results and Discussion, the paragraph on novices and metacognition. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

11. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Conclusion, the sentence on student evaluations of teaching. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

12. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Results and Discussion, the paragraph on the semester-long intervention, its second sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

13. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Results and Discussion, the paragraph on the semester-long intervention. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

14. Roediger, Henry L, III and Karpicke, Jeffrey D. Test-enhanced learning: taking memory tests improves long-term retention. 2006. figures quoted in the text. https://doi.org/10.1111/j.1467-9280.2006.01693.x (accessed 2026-09-24).

15. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. § 1.4 Experimental Design and Ethics, the paragraph on lurking variables. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

16. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. § 1.4 Experimental Design and Ethics, the paragraph on lurking variables, its fifth sentence. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

17. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. § 1.4 Experimental Design and Ethics, the paragraph on the power of suggestion. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

18. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. § 1.4 Experimental Design and Ethics, the worked problem on birth order, its solution. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

19. Higgins, Julian PT and Li, Tianjing and Deeks, Jonathan J. Chapter 6: Choosing effect measures and computing estimates of effect. 2024. § 6.5.1.1 The mean difference (or difference in means), first paragraph. https://training.cochrane.org/handbook/current/chapter-06 (accessed 2026-09-24).

20. Higgins, Julian PT and Li, Tianjing and Deeks, Jonathan J. Chapter 6: Choosing effect measures and computing estimates of effect. 2024. § 6.5.1.2 The standardized mean difference, first paragraph. https://training.cochrane.org/handbook/current/chapter-06 (accessed 2026-09-24).

21. Higgins, Julian PT and Li, Tianjing and Deeks, Jonathan J. Chapter 6: Choosing effect measures and computing estimates of effect. 2024. § 6.5.1.2 The standardized mean difference, second paragraph. https://training.cochrane.org/handbook/current/chapter-06 (accessed 2026-09-24).

22. Higgins, Julian PT and Li, Tianjing and Deeks, Jonathan J. Chapter 6: Choosing effect measures and computing estimates of effect. 2024. § 6.5.1.2 The standardized mean difference, the paragraph on Hedges' g and Glass' delta. https://training.cochrane.org/handbook/current/chapter-06 (accessed 2026-09-24).

23. Higgins, Julian PT and Li, Tianjing and Deeks, Jonathan J. Chapter 6: Choosing effect measures and computing estimates of effect. 2024. § 6.5.1.2 The standardized mean difference, the paragraph on the term effect size. https://training.cochrane.org/handbook/current/chapter-06 (accessed 2026-09-24).

24. Higgins, Julian PT and Li, Tianjing and Deeks, Jonathan J. Chapter 6: Choosing effect measures and computing estimates of effect. 2024. § 6.5.1.2 The standardized mean difference, third paragraph. https://training.cochrane.org/handbook/current/chapter-06 (accessed 2026-09-24).

25. Higgins, Julian PT and Li, Tianjing and Deeks, Jonathan J. Chapter 6: Choosing effect measures and computing estimates of effect. 2024. § 6.5.1.2 The standardized mean difference, MECIR Box 6.5.a, C61. https://training.cochrane.org/handbook/current/chapter-06 (accessed 2026-09-24).

26. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Materials and Methods, Data Analysis, the paragraph on how effect sizes were calculated. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

27. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Results, the opening paragraph on the effect sizes derived. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

28. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Abstract. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

29. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Materials and Methods, Coding, the paragraph after the coding criteria. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

30. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Results, first paragraph. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

31. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Materials and Methods, Criteria for Admission, criterion i. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

32. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Introduction, fourth paragraph, the first outcome variable. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

33. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Materials and Methods, Data Analysis, the paragraph on the two types of study. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

34. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Table 1, row “Randomized assignment or crossover design”. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

35. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Table 1, row “Quasirandom—no statistical difference on metrics of academic ability/preparedness”. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

36. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Results, the paragraph on methodological quality. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

37. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering,

and mathematics. 2014. Materials and Methods, Data Analysis, the paragraph on the random effects model. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

38. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Discussion, third paragraph. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

39. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Discussion, the second of two translations of 0.47 SDs. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

40. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Materials and Methods, the paragraph on controls, item 4. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

41. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Materials and Methods, the paragraph on controls, item 5. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

42. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Materials and Methods, the paragraph describing the two kinds of class. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

43. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Materials and Methods, the consensus definition of active learning. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

44. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. Results, the paragraph on the heterogeneity analysis by assessment type. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

45. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Table 1, spring semester, row “Average of first 2 midterms (%)”. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

46. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Table 1, fall semester, row “Average of first 2 midterms (%)”. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

47. Roediger, Henry L, III and Karpicke, Jeffrey D. Test-enhanced learning: taking memory tests improves long-term retention. 2006. Experiment 2, Results and Discussion, Final Tests, the 5-minute result (p. 252). https://doi.org/10.1111/j.1467-9280.2006.01693.x (accessed 2026-09-24).

48. Roediger, Henry L, III and Karpicke, Jeffrey D. Test-enhanced learning: taking memory tests improves long-term retention. 2006. Experiment 2, Results and Discussion, Final Tests, the 1-week result (p. 252). https://doi.org/10.1111/j.1467-9280.2006.01693.x (accessed 2026-09-24).

49. Roediger, Henry L, III and Karpicke, Jeffrey D. Test-enhanced learning: taking memory tests improves long-term retention. 2006. Experiment 2, Results and Discussion, the post hoc analyses (p. 253), 5-minute test. https://doi.org/10.1111/j.1467-9280.2006.01693.x (accessed 2026-09-24).

50. Roediger, Henry L, III and Karpicke, Jeffrey D. Test-enhanced learning: taking memory tests improves long-term retention. 2006. Experiment 2, Results and Discussion, the post hoc analyses (p. 253), 1-week test. https://doi.org/10.1111/j.1467-9280.2006.01693.x (accessed 2026-09-24).

51. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

52. Freeman, Scott and Eddy, Sarah L and McDonough, Miles and Smith, Michelle K and Okoroafor, Nnadozie and Jordt, Hannah and Wenderoth, Mary Pat. Active learning increases student performance in science, engineering, and mathematics. 2014. figures quoted in the text. https://www.pnas.org/doi/10.1073/pnas.1319030111 (accessed 2026-09-24).

53. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Introduction, first sentence. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

54. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Aims, the paragraph giving the review's operational definition of retrieval practice. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

55. Roediger, Henry L, III and Karpicke, Jeffrey D. Test-enhanced learning: taking memory tests improves long-term retention. 2006. Introduction, first paragraph, naming the testing effect. https://doi.org/10.1111/j.1467-9280.2006.01693.x (accessed 2026-09-24).

56. Roediger, Henry L, III and Karpicke, Jeffrey D. Test-enhanced learning: taking memory tests improves long-term retention. 2006. Experiment 1, Results and Discussion, Final Test, post hoc analyses (5-minute test). https://doi.org/10.1111/j.1467-9280.2006.01693.x (accessed 2026-09-24).

57. Roediger, Henry L, III and Karpicke, Jeffrey D. Test-enhanced learning: taking memory tests improves long-term retention. 2006. Experiment 1, Results and Discussion, Final Test, post hoc analyses (1-week test). https://doi.org/10.1111/j.1467-9280.2006.01693.x (accessed 2026-09-24).

58. Roediger, Henry L, III and Karpicke, Jeffrey D. Test-enhanced learning: taking memory tests improves long-term retention. 2006. General Discussion, on the absence of feedback. https://doi.org/10.1111/j.1467-9280.2006.01693.x (accessed 2026-09-24).

59. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Abstract. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

60. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Method, Data Coding, the effect-size paragraph. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

61. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Considerations for Applied Research in Classrooms, Definition of "Classroom" Research, the three screening guidelines. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

62. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Results, paragraph on Fig. 2. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

63. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Recommendations for Classroom Implementation of Retrieval Practice. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

64. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Results, paragraph on Fig. 2, the negative effects. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

65. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Results, Moderating Variables, education level. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

66. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Table 2, Location rows. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

67. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 8.4 Issues for implementation, on feedback. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

68. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Retrieval practice, paragraph on teacher blogs. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

69. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Results, Feedback Provision and Timing. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

70. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Future Directions, first recommendation, on the delay before the final test. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

71. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Results, Retrieval Practice Format and Final Test Format, rephrased questions. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

72. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Considerations for Applied Research in Classrooms, the Sennhenn-Kirchner example. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

73. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Methods, Literature Search, first sentence. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

74. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Future Directions, final recommendation, on non-WEIRD countries. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

75. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Retrieval practice, final paragraph, on test anxiety (Hinze and Rapp). https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

76. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Retrieval practice, paragraph on retrieval success. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

77. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Retrieval practice, paragraph on implementing it in class. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

78. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Aims, on why the review does not pool the effect sizes. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

79. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. figures quoted in the text. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

80. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Table 1, row 9, the description of distributed practice. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

81. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9 Distributed practice, opening paragraph, definition of the distributed-practice effect. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

82. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Spaced practice, first paragraph, on equal study time. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

83. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Summary, the foreshadowed recommendations. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

84. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9 Distributed practice, opening paragraph, on cramming in the short term. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

85. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9.2d Criterion tasks, on immediate against delayed tests. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

86. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Current Study, number of subjects and conditions. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

87. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Method, Stimuli and materials. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

88. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Method, Stimuli and materials, the definitions of gap and retention interval. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

89. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Current Study, the second learning session. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

90. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9.2a Learning conditions, on the range of gaps in Cepeda et al. (2008). https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

91. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Results, first paragraph, the shape of the gap effect. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

92. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Results, first paragraph, the optimal gaps for recall. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

93. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Abstract, the optimum gap as a share of the test delay. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

94. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Discussion, the fourth constraint on the retention surface. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

95. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Method, Design and procedure, the gap-to-retention-interval ratios tested. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

96. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Results, first paragraph, the size of the effect. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

97. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Educational Implications, on gaps that are too long against too short. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

98. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9.2a Learning conditions, Dunlosky's summary of Cepeda et al. (2008). https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

99. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9.2a Learning conditions, the worked examples after the 10-20% summary. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

100. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9.2a Learning conditions, on distributed testing against distributed study. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

101. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9.2c Materials, on complex material. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

102. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9.4 Issues for implementation, the procrastination scallop. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

103. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and edu-

cational psychology. 2013. 9.4 Issues for implementation, on students' ratings after massed study. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

104. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9.1, the Bahrick (1979) experiment, practice performance with 30-day gaps. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

105. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Spaced practice, third paragraph, on reviewing in class. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

106. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Results, first paragraph, the optimal gaps for recognition. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

107. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Results, first paragraph, the size of the effect on recognition. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

108. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Results, second paragraph, the improvement at each retention interval. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

109. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Educational Implications, on retention over several years. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

110. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Method, Subjects, who took part. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

111. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9.2a Learning conditions, describing Cepeda et al. (2008). https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

112. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Spaced practice, fourth paragraph, on confidence and desirable difficulty. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

113. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9.1, on theories of the distributed-practice effect. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

114. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 9.1, the deficient-processing account. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

115. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. Discussion, Theoretical Implications, on deficient processing theory. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

116. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. figures quoted in the text. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

117. Cepeda, Nicholas J and Vul, Edward and Rohrer, Doug and Wixted, John T and Pashler, Harold. Spacing effects in learning: a temporal ridgeline of optimal retention. 2008. figures quoted in the text. https://files.eric.ed.gov/fulltext/ED505660.pdf (accessed 2026-09-24).

118. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Table 1, rows 4 and 7, the descriptions of highlighting/underlining and rereading. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

119. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Summary, first paragraph, the number of techniques reviewed. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

120. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 4 Highlighting and underlining, opening paragraph, on treating the two as one technique. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

121. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Reviewing the Learning Techniques, on the three overall ratings. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

122. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Reviewing the Learning Techniques, on what a low rating can mean. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

123. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Closing Remarks, Relative utility of the learning techniques, on what the ratings assess. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

124. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Summary, the techniques rated low. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

125. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Summary, the techniques rated high. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

126. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 4.5 Highlighting and underlining, Overall assessment. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

127. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 4.5 Highlighting and underlining, Overall assessment, the next sentence. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

128. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 4.2d Criterion tasks, the Peterson (1992) study. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

129. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 7.2a Learning conditions, on dosage. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

130. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 7.5 Rereading, Overall assessment, on durability. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

131. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 7.4 Rereading, Issues for implementation, on head-to-head comparisons. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

132. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 7.5 Rereading, Overall assessment, closing sentence. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

133. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 7.3 Rereading, Effects in representative educational contexts. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

134. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Results, on comparison conditions in the 50 classroom experiments. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

135. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Summary, the recommendation for rereading and highlighting. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

136. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 7 Rereading, opening paragraph, the Karpicke et al. (2009) survey. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

137. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 7 Rereading, opening paragraph, the same survey, the next clause. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

138. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 4.1 Highlighting, the Fowler and Barker (1974) experiment, the first detailed finding. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

139. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 4.1 Highlighting, the Fowler and Barker (1974) experiment, the third detailed finding. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

140. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 4.1 Highlighting, on overmarking. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

141. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 4.1 Highlighting, on how much students mark. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

142. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 4.2b Student characteristics, on overmarking. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

143. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Table 4, the Highlighting row. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

144. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Table 4, the Rereading row. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

145. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Table 4, the Practice testing row. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

146. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Table 4, note, the Q rating. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

147. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Table 4, note, the I rating. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

148. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Table 4, note, the P and N ratings. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

149. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Closing Remarks, on cells with two ratings. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

150. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Closing Remarks, the N rating for highlighting. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

151. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 7.5 Rereading, Overall assessment, materials. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

152. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 7.4 Rereading, Issues for implementation. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

153. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 7.2b Rereading, Student characteristics. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

154. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Closing Remarks, Implications for research, on classroom evaluations. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

155. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Conclusions and Recommendations, Beliefs Versus Evidence as a Foundation for Educational Practices and Policies, first paragraph. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

156. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Discussion, the two likely factors behind the negative correlation. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

157. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 4.5 Effects in representative educational contexts (the PDF's own numbering), the Cashen and Leicht (1970) classroom study. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

158. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Abstract, first paragraph, the definition of learning styles. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

159. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Abstract, first paragraph, on how learning style is assessed. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

160. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. What Evidence Is Necessary, Existence of Study Preferences. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

161. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. What Evidence Is Necessary, The Learning-Styles Hypothesis, first paragraph. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

162. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Abstract, first paragraph, the meshing hypothesis. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

163. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Interactions as the Key Test of the Learning-Styles Hypothesis, the first criterion. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

164. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Interactions as the Key Test, the second criterion. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

165. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Interactions as the Key Test, the third criterion. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

166. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Interactions as the Key Test, the fourth criterion. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

167. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Interactions as the Key Test, the paragraph after the criteria, the crossover. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

168. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Interactions as the Key Test, on Figure 1B. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

169. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Interactions as the Key Test, on Figure 1C. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

170. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Interactions as the Key Test, the most liberal criterion. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

171. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. The Learning-Styles Hypothesis, on the hypothesis being true without meshing. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

172. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Interactions as the Key Test, on Figures 1D to 1I. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

173. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Figure 2 caption. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

174. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. The Learning-Styles Hypothesis, the rearrangement of Figures 1D and 1G. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

175. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Overview of Learning Styles, on the size of the literature. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

176. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Evaluation of Learning-Styles Literature, Style-by-Treatment Interactions, the one study found. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

177. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Evaluation of Learning-Styles Literature, the Sternberg et al. (1999) study. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

178. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Evaluation of Learning-Styles Literature, the Sternberg et al. (1999) study, first feature. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

179. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Evaluation of Learning-Styles Literature, the Sternberg et al. (1999) study, second and third features. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

180. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Learning-Styles Studies With Appropriate Methods and Negative Results, closing paragraph, the verdict on those studies. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

181. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Learning-Styles Studies With Appropriate Methods and Negative Results, Cook et al. (2009), the prediction tested. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

182. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Learning-Styles Studies With Appropriate Methods and Negative Results, Constantinidou and Baker (2002). https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

183. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Learning-Styles Studies With Appropriate Methods and Negative Results, Cook et al. (2009). https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

184. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Learning-Styles Studies With Appropriate Methods and Negative Results, closing paragraph. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

185. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Learning-Styles Studies With Appropriate Methods and Negative Results, closing paragraph, the caution. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

186. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Abstract, fifth paragraph. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

187. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Conclusions and Recommendations, Points of Clarification. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

188. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Interactions of Individual Differences and Instructional Methods, on prior knowledge. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

189. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Conclusions and Recommendations, Costs and Benefits of Educational Interventions. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

190. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. Everybody's Potential to Learn, on testing. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

191. Newton, Philip M and Miah, Mahallad. Evidence-based higher education – is the learning styles `myth' important?. PMID 28396647, PMC5366351. CC BY. Quoted from the publisher's article page. Methods, Participants. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.00444/full (accessed 2026-09-24).

192. Newton, Philip M and Miah, Mahallad. Evidence-based higher education – is the learning styles `myth' important?. PMID 28396647, PMC5366351. CC BY. Quoted from the publisher's article page. Discussion, first paragraph. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.00444/full (accessed 2026-09-24).

193. Newton, Philip M and Miah, Mahallad. Evidence-based higher education – is the learning styles `myth' important?. PMID 28396647, PMC5366351. CC BY. Quoted from the publisher's article page. Discussion, first paragraph, opening sentences. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.00444/full (accessed 2026-09-24).

194. Newton, Philip M and Miah, Mahallad. Evidence-based higher education – is the learning styles `myth' important?. PMID 28396647, PMC5366351. CC BY. Quoted from the publisher's article page. Results, Continued Use of Learning Styles?, and Abstract. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.00444/full (accessed 2026-09-24).

195. Pashler, Harold and McDaniel, Mark and Rohrer, Doug and Bjork, Robert. Learning styles: concepts and evidence. 2008. figures quoted in the text. https://doi.org/10.1111/j.1539-6053.2009.01038.x (accessed 2026-09-24).

196. Newton, Philip M and Miah, Mahallad. Evidence-based higher education – is the learning styles `myth' important?. PMID 28396647, PMC5366351. CC BY. Quoted from the publisher's article page. figures quoted in the text. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2017.00444/full (accessed 2026-09-24).

197. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. Abstract, first and second sentences. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

198. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 2.1.1 Methods (Study 1), second paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

199. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 2.1.1 Methods (Study 1), first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

200. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 2.1.1 Methods (Study 1), Phase 1, the rating scale. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

201. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 2.1.1 Methods (Study 1), Phase 2. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

202. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 2.1.1 Methods (Study 1), Phase 3. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

203. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 2.1.1 Methods (Study 1), Phase 3, last sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

204. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 2.1.1 Methods (Study 1), first paragraph, the test items. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

205. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. Fig. 3 caption (Studies 1 and 2). https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

206. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 2.1.3 Discussion (Study 1), debriefing. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

207. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 2.2.1 Methods and results (Study 2). https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

208. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 2.6.3 Discussion (Study 6, the explicit warning). https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

209. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 3.4.3 Discussion (Study 10), closing summary of Studies 1-10. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

210. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. Table 8, rows for Studies 7 and 9. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

211. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. Table 8, row for Study 9. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

212. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 3.2.3 Discussion (Study 8, procedures). https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

213. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. 3.2.1 Methods (Study 8). https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

214. Fisher, Matthew and Keil, Frank C. The curse of expertise: when more knowledge leads to miscalibrated explanatory insight. 2015. Abstract, Experiment 2 sentence. https://cogdevlab.yale.edu/sites/default/files/files/Fisher-2015.pdf (accessed 2026-09-25).

215. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Table 1, rows 1 and 2. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

216. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. Summary box, paragraph on the recommendations. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

217. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 2.5 Self-explanation: Overall assessment. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

218. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 1.5 Elaborative interrogation: Overall assessment. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

219. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 1.2d Criterion tasks (elaborative interrogation), paragraph on delays. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

220. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 2.2d Criterion tasks (self-explanation), paragraph on durability. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

221. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 2.2d Criterion tasks (self-explanation), end of paragraph on durability. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

222. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 1.3 Effects in representative educational contexts (elaborative interrogation). https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

223. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 2.4 Issues for implementation (self-explanation), paragraph on time on task. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

224. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Retrieval practice, paragraph on how easy it is to use in class. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

225. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 2.4 Issues for implementation (self-explanation), paragraph on time on task, Matthews and Rittle-Johnson. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

226. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 2.4 Issues for implementation (self-explanation), first paragraph, Didierjean and Cauzinille-Marmeche. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

227. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Elaboration, paragraph on elaborative interrogation. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

228. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 1.3 Effects in representative educational contexts (elaborative interrogation), Smith et al. 2010. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

229. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 1.4 Issues for implementation (elaborative interrogation), paragraph on dosage. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

230. Rozenblit, Leonid and Keil, Frank. The misunderstood limits of folk science: an illusion of explanatory depth. 2002. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC3062901/ (accessed 2026-09-25).

231. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Abstract. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

232. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Introduction, first paragraph. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

233. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Short History, Human Cognitive Architecture Used in 1998, last paragraph. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

234. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Categories of Cognitive Load, first paragraph (the word 'characteristics'). https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

235. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Categories of Cognitive Load, first paragraph, closing sentences. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

236. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Categories of Cognitive Load, first paragraph, last sentence. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

237. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Categories of Cognitive Load, second paragraph. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

238. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Conclusions, third paragraph, the list of strengths. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

239. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Worked Example Effect. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

240. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Redundancy Effect. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

241. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Transient Information Effect. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

242. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Expertise Reversal Effect. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

243. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Guidance-Fading Effect. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

244. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Guidance-Fading Effect, the completion strategy. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

245. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Guidance-Fading Effect, on programmes of longer duration. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

246. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Element Interactivity Effect. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

247. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Variability Effect. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

248. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Self-Explanation Effect. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

249. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. Instructional Effects Reported in 1998. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

250. Sweller, John and van Merri"enboer, Jeroen J G and Paas, Fred. Cognitive architecture and instructional design: 20 years later. CC BY 4.0. Quoted from the publisher's article page; its Table 1 is not quoted. The Narrow Limits of Change Principle. https://link.springer.com/article/10.1007/s10648-019-09465-5 (accessed 2026-09-24).

251. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Interleaving, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

252. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Interleaving, first paragraph, Rohrer and Taylor (2007). https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

253. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Interleaving, first paragraph, last sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

254. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Interleaving, fourth paragraph, the caveats. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

255. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 1.2b Student characteristics (elaborative interrogation), Woloshyn, Pressley and Schneider 1992. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

256. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. 1.2b Student characteristics (elaborative interrogation), sentence before Woloshyn et al. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

257. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

258. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Defining "Learning Objective", first paragraph. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

259. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Defining "Learning Objective", second paragraph, first sentence. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

260. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Table 1, Practical tips for writing effective learning objectives, seventh bullet. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

261. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Defining "Learning Objective", third paragraph (the verbs to avoid, and one action per objective). https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

262. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Blooms' Taxonomy and Action Verbs, first paragraph. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

263. Adams, Nancy E. Bloom's taxonomy of cognitive learning objectives. PMID 26213509, PMC4511057. Copyright 2015, the author; free to read in PMC, no open licence stated. Changes in Bloom's taxonomy, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC4511057/ (accessed 2026-09-24).

264. Adams, Nancy E. Bloom's taxonomy of cognitive learning objectives. PMID 26213509, PMC4511057. Copyright 2015, the author; free to read in PMC, no open licence stated. Comments, third paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC4511057/ (accessed 2026-09-24).

265. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Table 2, Revised Bloom's Taxonomy of Learning Objectives, the Remember and Understand rows. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

266. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Table 2, the Understand row. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

267. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Table 2, the Analyze row. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

268. Adams, Nancy E. Bloom's taxonomy of cognitive learning objectives. PMID 26213509, PMC4511057. Copyright 2015, the author; free to read in PMC, no open licence stated. Body, the paragraph on comprehension, first sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC4511057/ (accessed 2026-09-24).

269. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Table 2, the Apply row. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

270. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Table 2, the Evaluate row. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

271. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Table 2, the Create row. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

272. Adams, Nancy E. Bloom's taxonomy of cognitive learning objectives. PMID 26213509, PMC4511057. Copyright 2015, the author; free to read in PMC, no open licence stated. Body, the paragraph on evaluation, last sentences. https://pmc.ncbi.nlm.nih.gov/articles/PMC4511057/ (accessed 2026-09-24).

273. Adams, Nancy E. Bloom's taxonomy of cognitive learning objectives. PMID 26213509, PMC4511057. Copyright 2015, the author; free to read in PMC, no open licence stated. Comments, fourth paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC4511057/ (accessed 2026-09-24).

274. Adams, Nancy E. Bloom's taxonomy of cognitive learning objectives. PMID 26213509, PMC4511057. Copyright 2015, the author; free to read in PMC, no open licence stated. Comments, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC4511057/ (accessed 2026-09-24).

275. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Conclusion, the re-written objectives, stem and objective 3. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

276. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Incorporating Bloom's Taxonomy and SMART, second paragraph. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

277. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. figures quoted in the text. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

278. Biggs, John. Enhancing teaching through constructive alignment. 1996. Abstract, the sentences defining constructive alignment. https://link.springer.com/article/10.1007/BF00138871 (accessed 2026-09-24).

279. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Do Learning Objectives Add Value to Curriculum and Assessment?, first paragraph. https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

280. Chatterjee, Debnath and Corral, Janet. How to write well-defined learning objectives. 2017. Do Learning Objectives Add Value to Curriculum and Assessment?, second paragraph (the intubation example). https://www.seahq.org/assets/docs/xix_4_chatterjee.pdf (accessed 2026-09-24).

281. Adams, Nancy E. Bloom's taxonomy of cognitive learning objectives. PMID 26213509, PMC4511057. Copyright 2015, the author; free to read in PMC, no open licence stated. Body, the paragraph on knowledge. https://pmc.ncbi.nlm.nih.gov/articles/PMC4511057/ (accessed 2026-09-24).

282. Adams, Nancy E. Bloom's taxonomy of cognitive learning objectives. PMID 26213509, PMC4511057. Copyright 2015, the author; free to read in PMC, no open licence stated. Body, the paragraph on knowledge, last sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC4511057/ (accessed 2026-09-24).

283. Dunlosky, John and Rawson, Katherine A and Marsh, Elizabeth J and Nathan, Mitchell J and Willingham, Daniel T. Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. 2013. §8 Practice testing, first paragraph. https://doi.org/10.1177/1529100612453266 (accessed 2026-09-24).

284. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Results, Retrieval Practice Format and Final Test Format, second paragraph. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

285. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. Curriculum volume, Levels of competency (Vol. I p. 31), rows K, KH and SH. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

286. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. Curriculum volume, Levels of competency, the row P. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

287. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. Curriculum volume, head of the first subject table (Anatomy, Vol.

I p. 34), column headings. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

288. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. Assessment, (b) Internal Assessment, first paragraph and the list of what it relates to. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

289. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. Assessment, (b) Certifiable Competencies Achieved, item 3. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

290. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Method, the guidelines defining classroom research, second and third bullets. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

291. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Method, the guidelines defining classroom research, third bullet. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

292. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Results, Retrieval Practice Timing and Delay Before the Final Test. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

293. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Table 3, first two rows. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

294. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. Results, Retrieval Practice Format and Final Test Format, first paragraph. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

295. Agarwal, Pooja K and Nunes, Ludmila D and Blunt, Janell R. Retrieval practice consistently benefits student learning: a systematic review of applied research in schools and classrooms. 2021. General Discussion, Recommendations for Classroom Implementation, second paragraph. https://pdf.poojaagarwal.com/Agarwal_etal_2021_EDPR.pdf (accessed 2026-09-24).

296. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Retrieval practice, the paragraph beginning “Practicing retrieval is a powerful way”. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

297. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Retrieval practice, the same paragraph, on Weinstein et al. (2016). https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

298. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Retrieval practice, the paragraph on retrieval success. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

299. Weinstein, Yana and Madan, Christopher R and Sumeracki, Megan A. Teaching the science of learning. PMID 29399621, PMC5780548. CC BY 4.0. Quoted from the PubMed Central open-access text. Retrieval practice, the paragraph on indirect benefits. https://pmc.ncbi.nlm.nih.gov/articles/PMC5780548/ (accessed 2026-09-24).

300. Roediger, Henry L, III and Karpicke, Jeffrey D. Test-enhanced learning: taking memory tests improves long-term retention. 2006. Experiment 1, Procedure. https://doi.org/10.1111/j.1467-9280.2006.01693.x (accessed 2026-09-24).

301. Hake, Richard R. Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. 1998. Abstract, the sentence defining the average normalized gain. https://doi.org/10.1119/1.18809 (accessed 2026-09-25).

302. Hake, Richard R. Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. 1998. Abstract, the sentence introducing the average normalized gain. https://doi.org/10.1119/1.18809 (accessed 2026-09-25).

303. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §2.5 Measures of the Center of the Data, first paragraph. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

304. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §2.7 Measures of the Spread of the Data, first paragraph. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

305. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §1.4 Experimental Design and Ethics, the paragraph on lurking variables. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

306. Hake, Richard R. Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. 1998. figures quoted in the text. https://doi.org/10.1119/1.18809 (accessed 2026-09-25).

307. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Methods, Data Analysis, the paragraph introducing Kirkpatrick's model. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

308. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Methods, Data Analysis, Table 1, the Level 1 row. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

309. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Methods, Data Analysis, the sentence on subjective and objective outcomes. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

310. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Methods, Data Analysis, Table 1, the Level 2A row. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

311. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Abstract, Results. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

312. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Abstract, Methods. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

313. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Results, Evaluation Design and Outcomes, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

314. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Results and Discussion, first paragraph, the primary FOL item. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

315. Illowsky, Barbara and Dean, Susan and OpenStax. Introductory Statistics 2e. 2023. §1.4 Experimental Design and Ethics, first paragraph after the introduction. https://openstax.org/details/books/introductory-statistics-2e (accessed 2026-09-23).

316. Deslauriers, Louis and McCarty, Logan S and Miller, Kelly and Callaghan, Kristina and Kestin, Greg. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. PMID 31484770, PMC6765278. CC BY-NC-ND 4.0. Quoted from the PubMed Central open-access text. Introduction, third paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC6765278/ (accessed 2026-09-24).

317. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

318. National Medical Commission. Graduate Medical Education Regulations, 2023. 2023. Notification No. U-14021-8-2023-UGMEB, 2 June 2023, heading. Gazette of India. https://nmc.org.in/storage/cms/rules-regulation-nmc/G83KmfBTvRKFfp99s6Vvjuuw3gJ7WM2ZP28Z3Zhk.pdf (accessed 2026-09-24).

319. National Medical Commission. Graduate Medical Education Regulations, 2023. 2023. Chapter V, regulation 20 (Curriculum). Gazette of India. https://nmc.org.in/storage/cms/rules-regulation-nmc/G83KmfBTvRKFfp99s6Vvjuuw3gJ7WM2ZP28Z3Zhk.pdf (accessed 2026-09-24).

320. National Medical Commission. Graduate Medical Education Regulations, 2023. 2023. Notification heading, the enabling powers. Gazette of India. https://nmc.org.in/storage/cms/rules-regulation-nmc/G83KmfBTvRKFfp99s6Vvjuuw3gJ7WM2ZP28Z3Zhk.pdf (accessed 2026-09-24).

321. National Medical Commission. Graduate Medical Education Regulations, 2023. 2023. Chapter V, regulation 21 (Training period and maximum duration). Gazette of India. https://nmc.org.in/storage/cms/rules-regulation-nmc/G83KmfBTvRKFfp99s6Vvjuuw3gJ7WM2ZP28Z3Zhk.pdf (accessed 2026-09-24).

322. National Medical Commission. Graduate Medical Education Regulations, 2023. 2023. Chapter V, regulation 21, proviso (maximum duration). Gazette of India. https://nmc.org.in/storage/cms/rules-regulation-nmc/G83KmfBTvRKFfp99s6Vvjuuw3gJ7WM2ZP28Z3Zhk.pdf (accessed 2026-09-24).

323. Mahajan, Rajiv and Gupta, Kapil. Periodically modified regulatory reforms for implementation of competency-driven undergraduate medical curriculum in India: a comparative analysis. Editorial. PMID 38912356, PMC11189270. CC BY-NC-SA 4.0. Introduction, the paragraph on GMER 2023 and the withdrawn guidelines. https://pmc.ncbi.nlm.nih.gov/articles/PMC11189270/ (accessed 2026-09-24).

324. Mahajan, Rajiv and Gupta, Kapil. Periodically modified regulatory reforms for implementation of competency-driven undergraduate medical curriculum in India: a comparative analysis. Editorial. PMID 38912356, PMC11189270. CC BY-NC-SA 4.0. Supplementary Examinations and Course Duration, final paragraph on GMER 2023. https://pmc.ncbi.nlm.nih.gov/articles/PMC11189270/ (accessed 2026-09-24).

325. Mahajan, Rajiv and Gupta, Kapil. Periodically modified regulatory reforms for implementation of competency-driven undergraduate medical curriculum in India: a comparative analysis. Editorial. PMID 38912356, PMC11189270. CC BY-NC-SA 4.0. Introduction, second paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC11189270/ (accessed 2026-09-24).

326. Mahajan, Rajiv and Gupta, Kapil. Periodically modified regulatory reforms for implementation of competency-driven undergraduate medical curriculum in India: a comparative analysis. Editorial. PMID 38912356, PMC11189270. CC BY-NC-SA 4.0. Roles of Indian Medical Graduate. https://pmc.ncbi.nlm.nih.gov/articles/PMC11189270/ (accessed 2026-09-24).

327. Mahajan, Rajiv and Gupta, Kapil. Periodically modified regulatory reforms for implementation of competency-driven undergraduate medical curriculum in India: a comparative analysis. Editorial. PMID 38912356, PMC11189270. CC BY-NC-SA 4.0. Introduction, second paragraph, the 2019 regulation. https://pmc.ncbi.nlm.nih.gov/articles/PMC11189270/ (accessed 2026-09-24).

328. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. Covering letter No. D-11011/500/2024-AcademicCell, 12 September 2024. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

329. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. Guidelines, section 5 (Goals for the Learner), role b. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

330. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. Guidelines, section 5 (Goals for the Learner), roles d, f and g. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

331. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. Curriculum volume, How to use the Manual, Section 1 and its 2.3 Goals for the Learner. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

332. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. p. 29, II. The period of 4 1/2 years, Phase I. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

333. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. Annexure 4, Foundation Course hours. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

334. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. p. 27, I. Training period and time distribution. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

335. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. Curriculum volume, Explanation of terms used in this manual. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

336. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. p. 34, Foundation Course, mentor-mentee programme. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

337. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. p. 34, Foundation Course, who may be a mentor. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

338. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. p. 34, Foundation Course, when mentees are allotted and how long they stay. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

339. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. p. 27, I. Training period and time distribution, the sentence after the one-third cap. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

340. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. p. 44, University Examinations, clause (c), Phase III Part 1. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

341. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. p. 28, I. Training period, maximum duration. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

342. National Medical Commission. Medical Institutions (Qualifications of Faculty) Regulations, 2025. 2025. Notification F. No. N-P051(12)/18/2024-PGMEB-NMC, 30 June 2025, preamble. Gazette of India. https://nmc.org.in/storage/cms/rules-regulation-nmc/QZSDeGu7WXS9MMI0dL5ZR1EbS48kLaFbvMTpww2B.pdf (accessed 2026-09-24).

343. National Medical Commission. Medical Institutions (Qualifications of Faculty) Regulations, 2025. 2025. Schedule, Table E, rows 1 (Professor) and 2 (Associate Professor), (b)(iii). Gazette of India. https://nmc.org.in/storage/cms/rules-regulation-nmc/QZSDeGu7WXS9MMI0dL5ZR1EbS48kLaFbvMTpww2B.pdf (accessed 2026-09-24).

344. National Medical Commission. Medical Institutions (Qualifications of Faculty) Regulations, 2025. 2025. Schedule, Table E, row 2 (Associate Professor), Note 2. Gazette of India. https://nmc.org.in/storage/cms/rules-regulation-nmc/QZSDeGu7WXS9MMI0dL5ZR1EbS48kLaFbvMTpww2B.pdf (accessed 2026-09-24).

345. National Medical Commission. Medical Institutions (Qualifications of Faculty) Regulations, 2025. 2025. Schedule, Table E, row 1 (Professor), (b)(ii). Gazette of India. https://nmc.org.in/storage/cms/rules-regulation-nmc/QZSDeGu7WXS9MMI0dL5ZR1EbS48kLaFbvMTpww2B.pdf (accessed 2026-09-24).

346. National Medical Commission, Undergraduate Medical Education Board. Guidelines for Competency Based Medical Education (CBME) Curriculum 2024. 2024. figures quoted in the text. National Medical Commission. https://nmc.org.in/storage/cms/rules-regulation-nmc/YTW423v3EsCFdtPv3I5qZkaNNhuXfub42UZITDRM.pdf (accessed 2026-09-24).

347. Mahajan, Rajiv and Gupta, Kapil. Periodically modified regulatory reforms for implementation of competency-driven undergraduate medical curriculum in India: a comparative analysis. Editorial. PMID 38912356, PMC11189270. CC BY-NC-SA 4.0. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC11189270/ (accessed 2026-09-24).

348. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Introduction, first paragraph, the sentence distinguishing leadership from management. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

349. Stoller, James K. Leadership essentials for CHEST medicine professionals: models, attributes, and styles. 2021. Abstract, first sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC7501065/ (accessed 2026-09-25).

350. Kotter, John P. What leaders really do. 1990. Opening section, the paragraph beginning ‘Leadership is different from management’ (reprint p. 3). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

351. Kotter, John P. What leaders really do. 1990. Opening section, the paragraphs after ‘Leadership is different from management’ (reprint p. 3). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

352. Kotter, John P. What leaders really do. 1990. Opening section, the paragraph beginning ‘Of course, not everyone’ (reprint p. 3). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

353. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0.

Quoted from the PMC open-access text. Abstract, Methods and Results. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

354. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Table 2, Evaluation outcomes, Behavior/expertise (objective) (Level 3B). https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

355. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Discussion, limitations paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

356. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Discussion, limitations paragraph, last point. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

357. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Discussion, concluding paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

358. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Discussion, the paragraph calling for pilot programmes. https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

359. Frich, Jan C and Brewster, Amanda L and Cherlin, Emily J and Bradley, Elizabeth H. Leadership development programs for physicians: a systematic review. PMID 25527339, PMC4395611. Open access under CC BY 4.0. Quoted from the PMC open-access text. Results, first paragraph (setting and duration). https://pmc.ncbi.nlm.nih.gov/articles/PMC4395611/ (accessed 2026-09-24).

360. Stoller, James K. Leadership essentials for CHEST medicine professionals: models, attributes, and styles. 2021. The Ubiquity of the Need for Leadership, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC7501065/ (accessed 2026-09-25).

361. Kotter, John P. What leaders really do. 1990. Creating a Culture of Leadership, second paragraph (reprint pp. 9-10). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

362. Kotter, John P. What leaders really do. 1990. Creating a Culture of Leadership, last two paragraphs (reprint p. 11). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

363. Kotter, John P. What leaders really do. 1990. Creating a Culture of Leadership, last sentence (reprint p. 11). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

364. Kotter, John P. What leaders really do. 1990. The Difference Between Management and Leadership, first paragraph (reprint p. 3). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

365. Kotter, John P. What leaders really do. 1990. The Difference Between Management and Leadership, second paragraph (reprint p. 3). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

366. Kotter, John P. What leaders really do. 1990. The Difference Between Management and Leadership, planning and budgeting against setting a direction (reprint p. 4). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

367. Kotter, John P. What leaders really do. 1990. The Difference Between Management and Leadership, organizing and staffing against aligning people (reprint p. 4). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

368. Kotter, John P. What leaders really do. 1990. Aligning People Versus Organizing and Staffing, second paragraph (reprint p. 7). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

369. Kotter, John P. What leaders really do. 1990. Aligning People Versus Organizing and Staffing, third paragraph (reprint p. 7). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

370. Kotter, John P. What leaders really do. 1990. Box, 'Aligning People: Chuck Trowbridge and Bob Crandall at Eastman Kodak' (reprint p. 4). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

371. Kotter, John P. What leaders really do. 1990. Setting a Direction Versus Planning and Budgeting, the paragraph after the SAS example (reprint p. 6). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

372. Kotter, John P. What leaders really do. 1990. The Difference Between Management and Leadership, controlling and problem solving (reprint pp. 4-5). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

373. Kotter, John P. What leaders really do. 1990. The Difference Between Management and Leadership, controlling and problem solving against motivating and inspiring (reprint p. 5). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

374. Kotter, John P. What leaders really do. 1990. Aligning People Versus Organizing and Staffing, the paragraph on credibility (reprint p. 7). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

375. Kotter, John P. What leaders really do. 1990. Aligning People Versus Organizing and Staffing, the paragraph on credibility, its end (reprint p. 8). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

376. Kotter, John P. What leaders really do. 1990. Opening section, the paragraph after 'Leadership is different from management' (reprint p. 3). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

377. Kotter, John P. What leaders really do. 1990. Aligning People Versus Organizing and Staffing, the paragraph on what organizing involves (reprint p. 7). https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

378. Kotter, John P. What leaders really do. 1990. figures quoted in the text. https://hbr.org/2001/12/what-leaders-really-do (accessed 2026-09-25).

Appendix · Worked answers

1 · Learning is what lasts, and how a session feels is not evidence of it

EXERCISE 1. The quiz mean measures performance: what the batch could do minutes after the teaching. The feedback mean measures feeling of learning: how much they think they learned. Neither measures learning.

Both can point the wrong way. In Roediger and Karpicke's second experiment, the group that did best at 5 minutes, and was surest it would remember, recalled least a week later.

Add a short test on the same objective, with new questions, a week or more later. Only its result can support "objective achieved".

2 · Testing a teaching method: the comparison, the delayed test and the size of the difference

EXERCISE 1. Split. Write the 60 names on slips and draw 30 at random for the question-based session. The other 30 get the usual talk. That is the comparison group. Before teaching, record something that could affect the result, such as years of service. Check that the two groups come out roughly level on it.

Kept the same. The content and the handout, the length of the session, the day of the week and the hall. Use the same trainer for both if possible. If you use two trainers, give each of them half of each group, so a trainer's skill does not travel with one method.

Measured. One new set of growth-chart questions, the same for everyone, marked by someone who does not know which group each worker was in. Do not reuse the questions from the session.

When. At a meeting two to four weeks later, not at the end of the session. A test at the end measures performance. The later test measures what lasted.

Reported first. The mean difference, question-based group minus talk group, in percentage points, with each group's mean and SD beside it. Then the SMD, with the SD you divided by named.

Thirty in each group is small. Say plainly that a gap of a few points could be chance, and do not report the SMD without the MD in marks.

EXERCISE 2. For a journalist: three minutes, one quotable sentence, and plain honesty about what is missing.

An effect size like this is usually the gap between two groups' average marks, measured in units of how much individual students' marks vary. So 0.8 means the app's group scored, on average, most of one such unit higher than some other group. The number cannot tell you who that other group was.

Ask the company four things. Compared with what: a good class, or no teaching at all? Who decided which students used the app: a lottery, or the students themselves? When were they tested: at the end of the session, or weeks later? And what was the gap in actual marks?

The quotable line: “An effect size is only as good as the comparison behind it. Ask what the app was compared with, and whether anyone was tested weeks later.”

If they cannot answer, say the claim has not been shown. Do not say the app does not work.

1. The MD is P’s mean minus Q’s mean.

```
72 minus 63 = 9
```

The SMD is the MD divided by the SD.

```
9 divided by 18 = 0.5
```

The MD is 9 and the SMD is 0.5.

2.

```
47 minus 53 = -6
-6 divided by 12 = -0.5
```

The MD is -6 and the SMD is -0.5. The minus sign says the new method’s group scored lower, by half a standard deviation.

3. The SMD is the MD divided by the SD. So the MD is the SMD times the SD.

```
0.25 times 20 = 5
```

The MD is 5. Check it by dividing back.

```
5 divided by 20 = 0.25
```

4. The SMD is the MD divided by the SD. So the SD is the MD divided by the SMD.

```
9 divided by 0.6 = 15
```

The SD was 15. Check it.

```
9 divided by 15 = 0.6
```

5. Keep the subtraction the same way round both times: STTT minus SSSS.

```
71 minus 83 = -12
61 minus 40 = 21
```

At 5 minutes the MD is -12 percentage points. At 1 week it is +21 percentage points.

The sign turns round. At 5 minutes the group that only read did better, by 12 points. A week later the group that was tested did better, by 21 points. Had you subtracted SSSS minus STTT at 5 minutes and STTT minus SSSS at 1 week, both numbers would be positive. Then the reversal would disappear from your report.

6. Spring first.

```
78.1 minus 77.9 = 0.2
0.2 divided by 13.4 = 0.0149
```

An MD of 0.2 percentage points and an SMD of about 0.01.

Now fall.

```
72.02 minus 70.5 = 1.52
1.52 divided by 13.0 = 0.117
```

An MD of 1.52 percentage points and an SMD of about 0.12.

Before any teaching, the groups sat 0.01 and 0.12 SDs apart. Chance made them close, and in fall a little less close. The authors tested these gaps and called the groups “indistinguishable”. Still, a later gap has to be judged against a starting gap of up to 0.12 SD. It was not zero.

7. The SD is the MD divided by d.

```
12 divided by 1.22 = 9.84
21 divided by 1.26 = 16.7
```

At 5 minutes, the implied SD is about 9.8 percentage points. At 1 week, it is about 16.7.

A week later the scores are far more spread out. Some students kept much of the passage and some kept little. The gap in marks grew from 12 to 21 points, but the spread grew almost as much. So the gap, measured in SDs, stayed near 1.2.

These SDs are implied, not reported. The paper may have computed d another way, so treat them as rough.

8. The first two lines are right. The MD is 8 percentage points.

The third line broke. It divides the MD by the comparison group’s mean, 58. The SMD divides by the SD, 16.

```
8 divided by 16 = 0.5
```

The SMD is 0.5, not 0.14.

Why it looks reasonable: dividing a change by the value it started from is exactly how you work out a per cent change. Here it gives the gap as a share of the comparison mean, about 14%. That is a real number, but it is not an SMD. The SMD asks how big the gap is against how much individual learners differ.

9. The arithmetic in lines 2 and 3 is right. The subtraction is the wrong way round for the claim in line 4.

The MD in this book is the new method’s group minus the comparison group.

```
58 minus 64 = -6
-6 divided by 12 = -0.5
```

The SMD is -0.5. The new method’s group scored half an SD lower, not higher.

Why it looks reasonable: taking the bigger number minus the smaller one is a habit, and it always gives a positive answer. A positive answer then reads as a win for whatever you were hoping for. Fix the order before you subtract, and write it beside the answer.

10. Lines 1 to 4 are right. The MD is -1.4 mistakes and the SMD is -1.

The last line broke. On this score, fewer is better. A negative MD means the question-based group made fewer mistakes: 1.4 fewer on average. That is one SD better, not worse.

The Cochrane Handbook warns that some scales have higher scores for a better outcome and some have lower. Before you read a sign, find out which way the score runs. If you are putting several studies side by side, turn round the ones that run the other way, by multiplying by -1.

```
-1 times -1 = 1
```

Reported on a "higher is better" footing, the SMD is +1 in favour of the question-based session.

Why it looks reasonable: nearly every test you set scores right answers, so "negative means worse" becomes automatic. Counting mistakes turns that round, and nothing in the arithmetic warns you.

11. Turn 0.47 SD into marks for a test with this spread, by multiplying the SMD by the SD.

```
0.47 times 13.4 = 6.298
```

About 6.3 percentage points. Now as a share of the mean score.

```
6.3 divided by 77.9 = 0.081
```

About 8% of the mean.

So "6%" could mean two different things. As percentage points, 6.3 is close to 6. As a per cent of the score, it would be about 8% here. The paper's text does not say which it means. It bases its 6% on three courses, analysed in its supplementary material.

What this does not establish:

- That your course would gain 6 points. The conversion assumes your SD, and your SD is not theirs.
- That the gain lasts. The timing of the exams varied across Freeman's studies, and none of this is a delayed test of your own.
- That every teacher would get it. The teachers in Freeman's studies were volunteers. The authors call it an open question whether the gain would hold if all teachers had to do it.
- That "Freeman proved it" for your setting. The studies were undergraduate science courses, not an Indian medical college.

12. Check the arithmetic first. The change, post-test minus pre-test:

```
68 minus 52 = 16
16 divided by 16 = 1
```

The rise is 16 percentage points, and 16 divided by the pre-test SD is 1. The division is right.

What it does not establish:

- That the case-based sessions caused the rise. There is one group and no comparison group. Any teaching raises a post-test score. If the post-test repeated the pre-test's questions, practice on them raises it too.
- That this is comparable with Freeman's 0.47. Freeman's figure is a gap between two groups taught differently, measured against the spread of scores. This 1.0 is one group's change over time. The two numbers answer different questions, so "more than double" compares nothing.
- That anything lasted. If the post-test came at the end of the posting, it measures performance, not learning.

What would establish it? Split the next batch by lot into case-based and usual sessions. Give both groups the same new test, two weeks or more later. Then report the MD in marks, with the SMD beside it.

3 · Retrieval practice: pulling it out beats putting it in again

EXERCISE 1. A good plan names four things, and each can be checked against the evidence in this section.

What they retrieve: something specific from the hour, such as "the four steps of reading the panel", not "anything you remember".

Alone: each learner writes, or answers in private. A chorus answer, or a pair answering together, lets some learners watch rather than retrieve. Agarwal's review kept only experiments where each student retrieved individually.

Low stakes: the score is not recorded, or counts for almost nothing, and you say so before they start.

Feedback: the correct answer comes after every attempt. Without it, the classroom effects were mostly small.

The second retrieval sits on a later day, not straight after the first. The test of whether it lasted comes days after the session, not in its last five minutes. At five minutes, rereading can look better than recall.

A plan that tests understanding uses at least some questions that are worded differently from the quiz.

EXERCISE 2. Answer first. Yes: trying to remember something helps you keep it better than reading it again, and this has been tested many times.

Give one concrete result. In one experiment, students who tried to recall a passage remembered 56% of it a week later; students who reread it remembered 42%.

Give the classroom evidence with its limit. In a review of 50 classroom experiments, almost all of the 49 results favoured the recall practice. None was done in India.

A quotable sentence: "A quiz does not only measure learning. Given without pressure, with the answers afterwards, it is one of the best-tested ways of causing it."

Calibrate. The evidence is strong that it works; how big the gain is in an Indian classroom has not been measured in this body of work. Do not say "tests raise marks by" any figure.

4 · Spacing: the same practice, spread out, and how far apart

EXERCISE 1. A good plan names the two end dates first. Everything else is worked out from them.

It splits the days between them into a gap and a retention interval. For an end date a few weeks away, the gap is about 20% to 30% of the retention interval. So the whole is 1.2 to 1.3 times the retention interval. Dividing gives a window of days, not a single day, and the plan picks the later end when it can.

At the review, each learner retrieves alone and without notes, and then sees the answer (section 3). Rereading the notes together is not the review this evidence supports.

Your own review follows the same arithmetic, counted back from the day you teach. The end date may be many months away. Then the plan says the share should be smaller, and that no figure here was measured for that distance.

EXERCISE 2. Start from her experience, because it is real. Revision just before an exam does feel like it works, and on a test straight after, cramming can do as well or better.

Then draw the two schedules on the board. One: topic, a long empty stretch, a revision week, the exam. Two: topic, a short review a week or two later, and the exam. Say what the evidence measured: with the same study time, the

spread-out schedule leaves more behind on a delayed test. In one large study, recall after the best gap was 64% higher than after no gap, from 10% to 111% depending on when the test came.

Explain why it feels wrong. A review soon after learning is easy, so it feels productive. A review after some forgetting is harder. Weinstein and colleagues call this a “desirable difficulty”: harder now, better later. Say that why it works is still debated. The schedule with the most forgetting between sessions did best a month later.

Then change the plan. Decide when each topic must be used. Put one review so that the gap before it is about 20% to 30% as long as the wait after it. Make it a closed-book attempt, with the answers checked after. Keep a revision week if she likes, as a second review, not the only one.

1. Divide, then multiply by 100.

```
11 divided by 35 = 0.3142857143
0.3142857143 times 100 = 31.42857143
21 divided by 350 = 0.06
0.06 times 100 = 6
23 divided by 350 = 0.0657142857
0.0657142857 times 100 = 6.57142857
```

31%, 6% and 7%.

2. Turn each percentage into a decimal by dividing by 100, then multiply.

```
20 divided by 100 = 0.2
0.2 times 45 = 9
5 divided by 100 = 0.05
0.05 times 360 = 18
```

9 and 18.

3. The whole is the larger part plus 0.2 times the larger part, which is 1.2 times the larger part. Divide the whole by 1.2 to get the larger part.

```
36 divided by 1.2 = 30
0.2 times 30 = 6
30 plus 6 = 36
```

The parts are 30 and 6.

4. Divide each gap by its retention interval, then multiply by 100.

```
1 divided by 7 = 0.1428571429
7 divided by 35 = 0.2
7 divided by 70 = 0.1
21 divided by 350 = 0.06
```

14%, 20%, 10% and 6%. The share does not fall at every step: it rises from 14% to 20% between 7 and 35 days, then falls to 10% and 6%. The gaps tested were coarse, so one step up is not a pattern.

For recognition, the best gaps at 35 and 70 days were 7 days, against 11 and 21 days for recall. The best gap depended on the kind of test as well as on its date. A choice-of-answers test and a write-it-out test are different targets.

5. The test date first. December has 31 days.

```
1 plus 60 = 61
61 minus 31 = 30
```

Day 60 is 30 January 2027.

The 60 days are the gap plus the retention interval. At 20%, the whole is 1.2 times the retention interval.

```
60 divided by 1.2 = 50
0.2 times 50 = 10
```

At 30%, the whole is 1.3 times the retention interval.

```
60 divided by 1.3 = 46.15
0.3 times 46.15 = 13.85
```

The review goes on day 10 to day 14 after teaching.

```
1 plus 10 = 11
1 plus 14 = 15
```

From 11 December to 15 December 2026. If you must pick one day, pick the later one: a gap too long costs less than one too short.

6. This time the gap is known and the retention interval is not. The gap is the share times the retention interval, so the retention interval is the gap divided by the share.

```
9 divided by 0.3 = 30
9 divided by 0.2 = 45
```

A larger share gives a shorter retention interval. The test should fall 30 to 45 days after the review. Add the 9 days of the gap.

```
9 plus 30 = 39
9 plus 45 = 54
```

39 to 54 days after the first teaching. A test much sooner than that makes the 9-day gap too long for it. As Cepeda and colleagues found, that is the cheaper mistake, but it is still a mistake.

7. The step that broke is the third. A 64% increase is a per cent change, not a gain of 64 percentage points (Book 0, A4). It means the new value is the old value plus 64% of the old value.

```
0.64 times 40 = 25.6
40 plus 25.6 = 65.6
```

65.6%, which is possible. The same figure can also be written as 1.64 times 40.

```
1.64 times 40 = 65.6
```

Why the wrong answer looks reasonable. Both numbers carry a percent sign, and adding two percentages feels natural. But one is a share of the facts recalled and the other is a change relative to a starting value. They cannot be added.

8. The step that broke is the last. The gap is the time from first study to the review, not from the review to the exam. Revising 6 days before the exam makes the gap 24 days and the retention interval 6 days.

```
30 minus 6 = 24
24 divided by 6 = 4
```

The gap is 4 times the retention interval, 400% of it, where 20% to 30% was the aim. This is the revision week before the exam, with a single review. There is a smaller slip in the third line too: the 20% is a share of the retention interval, not of the whole 30 days.

```
30 divided by 1.2 = 25
0.2 times 25 = 5
30 divided by 1.3 = 23.08
0.3 times 23.08 = 6.92
```

Revise on day 5 to day 7, which is 25 to 27 October.

Why the wrong answer looks reasonable. It uses the right percentage, and revising just before the exam is the habit most of us grew up with. The review feels productive because the exam is close. That is the good feeling this section warns against.

9. What to compute: 10% and 20% of 3 years, and then whether the data behind the 10 to 20% reach that far. Take a year as 365 days.

```
3 times 365 = 1095
0.1 times 1095 = 109.5
0.2 times 1095 = 219
```

109.5 to 219 days. At about 30 days to a month, that is about 3.7 to 7.3 months, so the colleague's arithmetic is roughly right.

```
109.5 divided by 30 = 3.65
219 divided by 30 = 7.3
```

Dunlosky and colleagues do print this rule. They also carry it to 5 years: "to remember something for 5 years, the learning episodes should be spaced 6 to 12 months apart". The study behind it, Cepeda and colleagues 2008, stopped at 350 days. There its fitted best gap was 23 days, 7% of the retention interval. The authors say the share must fall as the retention interval grows.

What this does not establish. It does not establish that 3.5 to 7 months is right for 3 years. A fixed 10 to 20% is being carried far past the data, and the data show the share falling. It does not give a better number either: 6% of 3 years would be the same kind of guess. Cepeda and colleagues themselves say only that a review "of at least several months" seems likely to pay off when something must last several years. And one review of trivia facts is not a clinical subject revised for an examination.

What to say to the colleague: keep the long gap, since too long costs less than too short. Earlier, shorter reviews are reasonable too, though no figure here measures them. And do not quote 10 to 20% as if it were measured for 3 years.

10. What to compute: where 64% came from, and how much it varied inside the same study.

The 64% is Cepeda and colleagues' figure for recall of trivia facts. It compares the best gap with a gap of zero, with study time the same. The paper also gives it for each retention interval: 10, 59, 111 and 77% for 7, 35, 70 and 350 days.

```
111 divided by 10 = 11.1
```

Inside one study, the increase ran from 10% to 111%, an eleven-fold range, depending on when the test came. On the multiple-choice test the per cent increase was 26%, yet the paper's d there was larger, 1.5 against 1.1 (section 2). A per cent change and a difference in standard deviations can rank the same results differently.

What this does not establish. It does not establish any change in marks. 64% is a per cent change in recall, not in examination marks, and not percentage points. The comparison is against no gap at all, not against however the student studies now. So it is not what a student would gain from switching timetables. The material was trivia facts learned online, not a medical syllabus. And nothing here says this timetable spaces anything well.

What to say: spacing is well supported, and worth using. The 64% is one study's best case against cramming in one sitting, and it says nothing about marks.

5 · What does not work: rereading and highlighting

EXERCISE 1. Answer first. Mostly, yes, as students usually do it. A large 2013 review of the research rated highlighting low: in most studies it did little beyond reading the text.

Give one concrete result. In one experiment, students who highlighted articles did no better on a test a week later than students who only read them. They did better on what they had marked, and slightly worse on what they had not.

Give the fair exception, so the answer stays true. It may help a reader who already knows what matters in a text. Teachers underlining a few key lines has helped in a classroom study.

Give the better use of the time: close the book, write down what you remember, check it.

A quotable sentence: "Highlighting shows where a student has been on the page, not what they can recall with the book shut."

Calibrate. The evidence is consistent that highlighting adds little. It comes mostly from laboratory studies, and few were run in real classrooms. Do not say highlighting harms learning: that rests on a single study the reviewers say needs repeating.

6 · Learning styles: a popular claim that fails its own test

EXERCISE 1. Classify first. Every student in the batch takes the questionnaire. Keep the two style groups, visual and verbal, and write down beforehand how a student is placed in each.

Allocate within each group. A lottery sends half of each style group to method one, say a session built on diagrams. The other half get method two, the same content as written text. Same teacher, same time, same content.

Test everyone the same way. One test, written before the teaching, taken by all students after the same delay, a week or more (section 2). If the groups sit different tests, nothing can be concluded about styles.

Read the result across the rows. Take the MD, diagrams minus text, within each style group. Support for matching needs opposite signs, with each group best on its own matched method. The same sign in both groups means one method is best for everyone, however different the two MDs. A group that scores higher on both methods tells you nothing either way.

Say what a single result is worth. One batch can show a crossover by chance. Ask for a statistical test of the crossover and a repeat in another batch before any student is sorted.

1. Compare the two columns within each row.

```
row P: 70 is higher than 60, so Y
row Q: 65 is higher than 50, so Y
```

Both rows are highest in column Y. The best column does not change from row to row, so there is no crossover.

2. Subtract within each row, the same way round.

```
row P: 72 minus 64 = 8
row Q: 58 minus 66 = -8
```

One difference is positive and the other negative. X is best in row P and Y is best in row Q. There is a crossover.

3. In row P, X minus Y = -6, so Y is 6 more than X.

```
Y = 55 plus 6 = 61
check: 55 minus 61 = -6
```

In row Q, X minus Y = 4, so X is 4 more than Y.

```
X = 66 plus 4 = 70
check: 70 minus 66 = 4
```

Row P is best in Y (61 against 55). Row Q is best in X (70 against 66). The signs are opposite, so there is a crossover.

4. Take the MD, diagram minus text, within each group, in percentage points.

```
visual: 80 minus 72 = 8
verbal: 55 minus 63 = -8
```

The signs are opposite, so there is a crossover. Diagrams are best for the visual group and text for the verbal group: each group's best lesson is its matched one. In this made-up table, the pattern is the one meshing predicts.

The colleague is wrong. That one group scores lower on both lessons is a difference between the groups. The test asks something else: within each group, which lesson is better? A crossover can be present even when one group outscores the other on every method.

5. MD, diagram minus text, within each group, in percentage points.

```
visual: 75 minus 60 = 15
verbal: 64 minus 62 = 2
```

Both MDs are positive. Diagrams are best for both groups, by 15 points for the visual group and 2 points for the verbal group. The two MDs differ in size, not in sign, so there is no crossover. Meshing would send the verbal students to the text lesson, and on these numbers they do slightly worse there. Teach everyone with diagrams. A 2-point difference may also be chance; the table alone cannot say.

6. MD, picture cards minus spoken, within each group, in percentage points.

```
visual: 58 minus 66 = -8
auditory: 71 minus 63 = 8
```

The signs are opposite, so there is a crossover: the best session changes with the group. But look at which session wins. The visual group does best with the spoken session, and the auditory group with the picture cards. Each group does best with the method that does not match its label.

Suppose a result like this were real and repeated. It would support Pashler's broad version of learning styles: sorting by the questionnaire would tell you which session to give. It contradicts meshing, because matching the session to the label would give each group its worse session.

7. The first two lines broke it. Each compares the two style groups on one lesson: down a column. The test compares the two lessons within one group: across a row.

```
visual: 68 minus 57 = 11
verbal: 62 minus 59 = 3
```

Both MDs are positive. Diagrams are best for both groups, so there is no crossover. The verbal students do better with diagrams too.

Why the wrong answer looks reasonable. It does find each group "winning" on its own method, and drawn with the lessons along the bottom the bars do seem to cross. Pashler and colleagues warn of exactly this. But the winning is against a different set of people. A teacher choosing a lesson for a verbal student compares that student's two options, and on these numbers diagrams win.

8. The last step broke, and it broke because the table is half empty. Only two of the four cells are there: the two matched ones. There is no score for visual learners taught with text, and none for verbal learners taught with diagrams. So no MD can be taken within either group, and no crossover can be looked for.

The comparison the worked answer does make is against a pass mark. That shows both groups passed. It says nothing about whether either group would have done better or worse with the other lesson. If visual learners taught with text had also scored 71%, matching would have made no difference at all.

Why the wrong answer looks reasonable. Both numbers are high, and each group got the lesson that matched its style. Success under matching feels like evidence for matching. It is only evidence when the unmatched cells are there to compare with.

9. What to compute: the only difference the numbers allow, and then which cells of the 2 × 2 table it fills.

```
74 minus 66 = 8
```

8 percentage points. But this is not an MD within any group. It sets one subgroup under one session against the whole batch under another session. The whole batch includes the visual learners themselves, and everyone else.

What it does not establish. It does not show that visual learners do better with diagrams than with text, because their own text score is not given. It does not show a crossover, because there is no verbal group, and no cell for either group under the other session. The two sessions may also have differed in content, in test, or in time of year, and nobody was allocated to a session by lottery. What to ask for: both style groups, each split at random between the two sessions, and the same test for all.

10. What to compute: how many people the percentages stand for, so the claim is sized.

```
0.58 times 114 = 66.12
0.93 times 137 = 127.41
```

About 66 academics and about 127 schoolteachers. The percentages are rounded, so the counts are approximate.

The premise is true: these surveys did find those levels of agreement. The inference does not carry (Book 0, F4). A survey of belief measures what people think works. Whether matching teaching to style improves learning is a different question, and the crossover test answers it. Pashler and colleagues found one doubtful study that even potentially met the test. The studies that met it properly, one of them with 123 internal medicine residents, found no support.

What this does not establish. It does not establish that learning styles are proven useless in every version: the same review says many versions have never been tested. It does not tell you what teachers in India believe, because

neither survey asked them. And the 66 and 127 are counts of people agreeing with a sentence, not of people shown to teach better by it.

7 · Explaining is how you find out whether you understand

EXERCISE 1. Start with the mechanism. Feeling that you understand is a rating you give yourself, and people rate their understanding of how things work higher than it is. When they then write out a full explanation and rate themselves again, the second rating is lower. Being warned that the explanation would be asked for made the drop smaller, not absent. Being well educated in the subject does not stop it either.

Then the whiteboard. Do not ask only for the steps of a measurement; in one study, people judged their knowledge of everyday procedures about right. Ask why. Take the height measurement, step by step as the chapter gives it. For each step, ask why it is there and what happens to the reading if it is skipped. Let the resident talk, and mark where they stop or say “and then basically”.

Then the task for tonight. Close the book. Write out, from memory, the explanation they would give the batch, as a chain of steps with the reason for each. Mark every place they stall. Open the book only for those places. Writing it from memory is also retrieval practice, so the same hour checks the understanding and strengthens it.

Close with the limit. A smooth explanation can still be wrong, so the check against the book is part of the task, not an extra.

8 · Working memory is small: cognitive load, and effort that helps

EXERCISE 1. Three things add load the teacher could remove.

Talk is gone within seconds. Each worker has to hold the steps in working memory until the practice at the end, up to 40 minutes later. Put the steps on a chart that stays on the wall. Break the talk into short pieces, and after each piece have the workers do that step.

The slide gives the same information twice, as a labelled picture and as text that repeats it. Keep the picture and drop the repeating paragraph.

The workers are beginners at this task, and the plan sends them straight to a whole problem. Show one weighing worked through in full, step by step. Then have each worker do one while you finish the last steps with them. Then each does one alone. Last, cover the chart and have each worker write the steps from memory (section 3).

Say what the critique does not show: whether the session as planned would fail. That is found by testing what the workers can do afterwards, not by counting load.

9 · Writing a learning objective that can be assessed

EXERCISE 1. 1. “Know” is a state of mind: you cannot watch it. Who is too loose (“students”); there is no condition, no standard and no by-when. A rewrite:

> *At the end of this session, each student, given a one-paragraph history of a child with obesity, will list the possible causes the history points to, without notes, naming at least three that the marking key accepts, on the day and on a test one week later.*

The level depends on the history: if the student meets a new history, it is above remember. 2. “Appreciate” names an attitude, and nothing in the sentence can be watched. Ask what an ASHA who appreciated it would do: she would weigh the child. A rewrite:

> *At the end of this session, each ASHA, given a child and the scale used at her centre, will weigh the child, completing every step on the programme’s weighing checklist, including writing the weight in the register, observed by the trainer.*

Recording is a step on the checklist, not a second action. This is a hands skill, so Bloom’s thinking-skill list does not grade it. 3. Two actions, and one of them (“understand”) cannot be watched. Drop it: appraising shows the understanding. A rewrite:

> *At the end of the journal club, each resident, given the trial’s abstract and methods, will judge whether the comparison between its groups was fair, giving in writing the one reason that matters most, marked by the convener against a key.*

“Judge” and “appraise” name evaluate, the fifth level.

EXERCISE 2. There is no single right objective; judge yours against the six parts and the colleague test.

- Who is named, and says “each”.
- The action is one verb you can watch or mark.
- The content is stated.
- The condition says what the learner is given and what they may not use.
- The standard is a number or a list the answer is marked against.
- By when includes a check days after the session, not only at its end.

Suppose your colleague’s question could be answered by repeating a sentence from your session, and you meant more than remember. Then your action or your condition is not yet doing its job. The usual fix is a condition that hands the learner a case they have not seen.

EXERCISE 3. Draw two columns. On the left, write the six revised levels, remember to create. On the right, beside “understand”, write three actions: explain in your own words, summarise, estimate.

Say that the level is the target and the verb is the evidence. “Understand” names what you hope is happening in the learner’s head. You cannot see that, so you cannot mark it. What you can see is the learner explaining, summarising or estimating something. Those are the words that go in the objective.

Then show that a verb is not a level. Take “describe”. Ask her to describe the parts of an objective from memory: that is remember. Ask her to describe what is missing from a colleague’s objective she has never seen: that is at least understand. Same verb, different task, different level.

Close with the check she can use on her own: write the question that tests the objective. If she cannot, the objective is not finished.

10 · A test that matches the objective

EXERCISE 1. It does not match. The objective's action is to weigh a child. The quiz asks the ASHAs to recognise the steps on paper. An ASHA could pick every right answer and still miss a step with a real child and a real scale. This is Chatterjee and Corral's intubation example in another setting: a written exam for a hands-on objective.

A matched test: each ASHA weighs a child, with her own centre's scale, while the trainer watches. The trainer ticks each step of the programme's checklist and checks the register entry. The standard is every step done.

In the terms of the Indian curriculum, this is a skill to be shown, not a thing to know. It needs watching, as a competency at SH would.

A second form: a different child. Or change the scale set-up, for example a scale that must first be set to zero. Keep the same checklist. What must stay the same is the action and the steps marked.

EXERCISE 2. Judge your pair of forms with four checks.

- Every item asks for the objective's action, under its condition. None can be answered by repeating a sentence from the session.
- Each partner changes the case or the numbers and keeps whatever makes the item hard. For BMI that was where the adult sits in the table. For reading an abstract it might be the length of the abstract and where the key sentence sits.
- The marking follows the standard: right or wrong per item, or a key written in advance.
- Half the learners take form A first and half take form B first, the two halves the same size, chosen before the session. An alphabetical split is fine if you fix it in advance.

If you have fewer than five items, each one is worth more than 20 points. Say in your notes that one item decides a lot.

EXERCISE 3. The answer first: the quiz tested whether ASHAs can recognise the steps on paper. The training was meant to make them weigh children correctly. Those are different things, and the quiz measured only the first.

A written quiz can be passed by memory of the training. Weighing a child needs hands, a scale and a real child. The only test of it is to watch an ASHA do it and check each step.

What this costs. A trainer or supervisor watches each ASHA weigh one child, with a checklist, during a routine visit. That is minutes per ASHA, not a new programme.

What to ask for next time. Each training plan should name what the trainee will be able to do, and should test exactly that, at the end and again weeks later.

11 · Running one session around retrieval rather than transmission

EXERCISE 1. One resident at a time answers. The other seven listen, so for most of the questioning most of the room is not recalling anything. Change: ask every resident the same question, each writing an answer alone before anyone speaks.

Marks that count can make the questions feel high-stakes. section 3 showed one study where pressure during quizzing cut the later benefit. Change: keep these questions out of formal assessment, so that nobody is afraid to get them wrong.

Nobody is told the right answer. Change: after the written answers, give the answer, or the key points of a good one, and let each resident check their own.

Thirty minutes of presenting in one block. Change: break it into short parts, each followed by one written question.

No objective, no pre-test, no closing free recall and no delayed test. Change: write one objective, such as judging whether the trial's comparison was fair. Test it before and after on two different abstracts, and book a third abstract for a week later.

EXERCISE 2. Check your pack before the session against this list.

- The objective has who, one action you can watch or mark, the content, the condition, the standard and by when.
- Every test item asks for that action on a case not used in the teaching. The two forms pair item by item. Half the group takes each form first, decided in advance.
- The plan has the seven parts in order. Objective, pre-test, and short teaching parts, each with a written question and its answer. Then a question from earlier material, free recall, the post-test, and a booked delayed test.
- Every in-session question is answered by every learner, alone, in writing, notes away, and then checked against the answer you show.
- No mark from the in-session questions goes into internal or any other formal assessment.
- You have added up the minutes in which every learner answers from memory, and written the total on the plan.

After the session, the build is met when the session happened as planned and the three sets of scores are kept. Whether the learners gained is a separate question, answered from those scores.

EXERCISE 3. The answer first: try it in a few trainings first, and measure it. Do not switch everything on the evidence we have.

Why it is worth trying. In the classroom studies Agarwal and colleagues reviewed, having every learner answer questions from memory, then showing the answer, taught more than the same time spent on lectures without quizzes.

Why not switch everything at once. Those studies were in schools, colleges and medical schools. The review names no training for ASHA workers. The effect may hold for them. The review cannot tell you.

What to do. Pick a few trainings. Keep the content, and cut the talking into short parts. After each part, ask every ASHA one question she answers on her own. Then give the answer. Test what she can do, not what she can recite, at the end and again a month later. Compare with trainings run the old way.

What it costs. Trainer time to rewrite the sessions and one short test a month later. No new staff.

12 · Measuring what changed: pre-test, post-test, gain

EXERCISE 1. A good plan names everything before the session, so nothing is chosen after you see the scores.

The objective and its standard, written as the section on objectives showed. For example: at the end of the session, each student will work out a family's daily energy intake. They are given a one-day diet record and a food table. The standard is at least 8 of 10 items right.

The test. Ten items, each asking for the objective's action, not for a sentence remembered from the session. Two parallel forms, with half the group taking each first, as the section on tests showed. A third form for the delayed test if you can build one; otherwise each learner's first form again.

The dates. Pre-test at the start of the session. Post-test at the end. Delayed test on a date you book now, a week or more later.

The numbers. Mean pre-test and post-test with the SD of the raw gains. The class normalised gain from the class means. The share reaching the standard at post-test and at delayed test. The share of the gain kept at the delayed test. The item table.

Three learners at 100% on the pre-test. Their normalised gain does not exist. Report how many there were, apart from the rest. Work out the class normalised gain without them and say so. Next time, the test needs harder items that still match the objective.

EXERCISE 2. The abstract gives 0.48 as an average over 48 courses: "〈g〉 IE-ave". Each course has one class normalised gain, worked out from that course's class-average scores. So the 0.14 beside it reads as the standard deviation of those 48 course gains around 0.48. The abstract does not spell this out, so hold the reading loosely. On that reading the interactive courses varied a good deal among themselves: many sat within about 0.14 of 0.48, and some further out.

It tells you nothing about single students. A course with a class normalised gain of 0.48 can hold students with gains near 0 and near 1. The spread of students' gains inside a course is not in the abstract.

Nor does it tell you the methods caused the difference. The abstract calls the work "a survey of pre/post-test data". It does not say the courses were assigned at random.

1.

```
62 minus 35 = 27
100 minus 35 = 65
27 divided by 65 is about 0.415
```

A raw gain of 27 percentage points, and a normalised gain of about 0.42.

2. Learner P.

```
85 minus 70 = 15
100 minus 70 = 30
15 divided by 30 = 0.5
```

Learner Q.

```
60 minus 20 = 40
100 minus 20 = 80
40 divided by 80 = 0.5
```

Raw gains of 15 and 40 points. Both normalised gains are 0.5: each learner gained half the room they had.

3.

```
10 plus 20 plus 15 plus 25 plus 30 = 100
100 divided by 5 = 20
```

The mean gain is 20 points. The deviations from 20 are -10, 0, -5, 5 and 10. Square them and add.

```
100 plus 0 plus 25 plus 25 plus 100 = 250
250 divided by 4 = 62.5
```

Divide by 4, one less than the 5 learners, because they are a sample. The square root of 62.5 is about 7.9. The SD of the gains is about 7.9 percentage points.

4. The room to gain is 100 minus 40. A quarter of it is gained.

```
100 minus 40 = 60
0.25 times 60 = 15
40 plus 15 = 55
```

post = 55.

5. The raw gain is g times the room to gain.

```
post - pre = g × (100 - pre)
85 - pre = 0.75 × (100 - pre)
```

Multiply out the right-hand side: 0.75 times 100 is 75, and 0.75 times pre is 0.75 × pre.

```
85 - pre = 75 - 0.75 × pre
```

Add pre to both sides, and take 75 from both sides.

```
85 minus 75 = 10
10 = pre - 0.75 × pre = 0.25 × pre
10 divided by 0.25 = 40
```

pre = 40. Check it forwards.

```
85 minus 40 = 45
100 minus 40 = 60
45 divided by 60 = 0.75
```

6. The room to gain is 100 minus 45.

```
100 minus 45 = 55
```

At a gain of 0.23:

```
0.23 times 55 = 12.65
45 plus 12.65 = 57.65
```

At a gain of 0.48:

```
0.48 times 55 = 26.4
45 plus 26.4 = 71.4
71.4 minus 57.65 = 13.75
```

Mean post-tests of about 57.7% and 71.4%, about 13.8 percentage points apart. These are what the two averages would mean for this class. They are not a prediction: each is an average over physics courses, and courses varied around it.

7.

```
23.3 divided by 30 = 0.7767
0.7767 times 100 = 77.67
100 minus 77.67 = 22.33
```

A pre-test of about 77.7%, which leaves about 22.3 percentage points of room.

```
0.48 times 22.33 = 10.72
77.67 plus 10.72 = 88.39
```

A raw gain of about 10.7 points, to a post-test of about 88.4%. A class that starts this high reaches an average normalised gain with a small raw gain. The study did not report a post-test on this test, so this is a what-if, not its result.

8. Before: workers 5 and 11 are at 70% or more.

```
2 divided by 12 is about 0.167
```

After: workers 1, 2, 4, 5, 7, 9, 10 and 11 are at 70% or more.

```
8 divided by 12 is about 0.667
66.7 minus 16.7 = 50
```

About 16.7% before and 66.7% after, a change of 50 percentage points. Four workers, 3, 6, 8 and 12, still did not reach the standard.

9. Divide each count by 20.

```
item 1: 17 divided by 20 = 0.85, 19 divided by 20 = 0.95
item 2: 6 divided by 20 = 0.30, 16 divided by 20 = 0.80
item 3: 4 divided by 20 = 0.20, 15 divided by 20 = 0.75
item 4: 5 divided by 20 = 0.25, 6 divided by 20 = 0.30
item 5: 9 divided by 20 = 0.45, 18 divided by 20 = 0.90
```

The changes, in percentage points.

```
95 minus 85 = 10
80 minus 30 = 50
75 minus 20 = 55
30 minus 25 = 5
90 minus 45 = 45
```

Item 1 was right for 85% before the journal club. Cut the time spent on it, or drop it. Item 4 barely moved, from 25% to 30%. Read the item first: if it is clear and matches the objective, teach that part again next time. If it is unclear, rewrite it.

10.

```
74 minus 38 = 36
61 minus 38 = 23
100 minus 38 = 62
36 divided by 62 is about 0.581
23 divided by 62 is about 0.371
23 divided by 36 is about 0.639
```

At the post-test: a raw gain of 36 points and a normalised gain of about 0.58. A week later: 23 points and about 0.37. About 64% of the end-of-session gain was still there.

11. The first two lines are right: a raw gain of 30 points.

The third line broke. The room to gain is measured from where the learner started, the pre-test, not from where they ended.

```
100 minus 50 = 50
30 divided by 50 = 0.6
```

The normalised gain is 0.6. The last line should have been the warning. A normalised gain cannot be above 1, because no one can gain more than all the room there was.

Why it looks reasonable: 100 − post is a real quantity, the room still left at the end. It is the wrong one to divide by.

12. The four learners' own normalised gains are right.

```
100 minus 90 = 10, 100 minus 90 = 10, 10 divided by 10 = 1.0
30 minus 20 = 10, 100 minus 20 = 80, 10 divided by 80 = 0.125
40 minus 30 = 10, 100 minus 30 = 70, 10 divided by 70 = 0.143
50 minus 40 = 10, 100 minus 40 = 60, 10 divided by 60 = 0.167
1.0 plus 0.125 plus 0.143 plus 0.167 = 1.435
```

The step that broke is the comparison. Hake's 0.23 is a class normalised gain worked out from class averages. The 0.36 is the mean of four learners' own normalised gains. Work out the same kind as Hake's.

```
90 plus 20 plus 30 plus 40 = 180
180 divided by 4 = 45
100 plus 30 plus 40 plus 50 = 220
220 divided by 4 = 55
55 minus 45 = 10
100 minus 45 = 55
10 divided by 55 is about 0.18
```

The class normalised gain is about 0.18. Every learner gained exactly 10 points. The one who started at 90% turned those 10 points into a gain of 1.0, and that one learner lifted the mean to 0.36.

The last line fails a second time. One session of four learners is not a physics course. Even at 0.18 against 0.23, the comparison says little.

Why it looks reasonable: "the mean normalised gain" sounds like the same thing as "the average normalised gain". They are computed in different orders, and here they differ by a factor of two.

13. Compute the gap in the abstract's own units, and the ratio the colleague is claiming.

```
0.48 minus 0.23 = 0.25
0.25 divided by 0.14 = 1.79
0.48 divided by 0.23 = 2.09
```

The gap is 0.25, which is about 1.8 of the interactive courses' SD of 0.14. That is what "almost two standard deviations" means in the abstract. The average interactive course gained about 2.1 times the room the average traditional course did.

What it does not establish. First, on the likeliest reading, the SD is a spread among the 48 course gains, not among students; the abstract does not say. Nothing in it describes how any one student will do. Second, the figures come from "a survey of pre/post-test data". Nothing in the abstract says the courses were assigned their methods at random, so the survey alone cannot say the methods caused the gap. Third, "twice as much" is true of the normalised gain. In raw points it depends on where a class starts, as the problem on a class at 45% showed.

14. Compute the change both ways, then the share of the room that was gained.

```
90 minus 60 = 30
30 divided by 60 = 0.5
100 minus 60 = 40
30 divided by 40 = 0.75
```

The pass rate rose by 30 percentage points. That is a 50% rise on the starting 60%, so the report's arithmetic holds, but only as a per cent change. The pass rate closed three quarters of the gap to 100%. That is the same as three quarters of the earlier failures passing only if nobody who passed before failed after.

What it does not establish. The same test before and after, on the same day, can raise scores by practice alone. There is no comparison group to separate that from the teaching. The test at the end of the day shows what the workers could do then, not a week later. Last, a pass rate depends on where the pass mark was set. The report should say what the mark was, and whether it matches the objective's standard.

13 · Liking a session is not learning from it

EXERCISE 1. The item "I have learned a great deal" is the student's own view of their learning. That is Level 2A, self-assessed knowledge. The rest of the form, if it asks about the session and the teacher, is Level 1.

The conclusion is about the learning objectives: that students can now do what each objective says. That is Level 2B, and it needs a test that asks for each objective's action, scored against a key. Nothing in the report is that.

The step that fails is the inference, not the premise. The ratings may be accurately high. They are still about how students felt. The first section of this book showed a rating and a test result pointing opposite ways.

Next year, give each session a short test on its objective. Give it at the start, at the end and a week or more later, as the section on measuring change set out. Report the share reaching each objective's standard. Keep the form for what it measures.

EXERCISE 2. Two of the three places say four: the abstract and Table 2. The Results sentence says five, but it cites only four study numbers after the word "studies" (34, 42, 45 and 57). The most likely reading is four, and "five" a slip.

Check the arithmetic of the table too.

```
4 divided by 45 is about 0.089
```

About 9%, which matches the table's "(9)". Five would have been about 11%.

```
5 divided by 45 is about 0.111
```

Write "four of 45 studies (Frich and colleagues' abstract and Table 2)". If the exact number matters to your argument, add that the Results text says five. Do not silently pick one. Whichever it is, fewer than one study in ten had anything to compare against.

EXERCISE 3. A worked plan, for a session teaching MBBS students in community medicine to work out a family's daily energy intake from a diet record:

- Level 1: a short form at the end. Could they hear, was the pace right, what was least clear. Carried out.
- Level 2B: the ten-item test on the objective, scored against a key. Given at the start, at the end, and a week later on a parallel form. Carried out. This is the one this book asks for.

- Level 3: whether students use the method on their family-study visits. A tutor checks one diet calculation in each student's family-study record a month later against the same key. Carried out if the tutor agrees; otherwise listed as not done.
- Level 4: whether the families' diets change. Not carried out. One session is unlikely to move it alone, the time is too long, and there is no comparison group.

A good answer says in writing which levels it will not reach, so nobody later reads the Level 1 form as if it covered them.

EXERCISE 4. The answer first: the report does not say whether the training worked. It says the workers liked it. Fund it again only with a plan to measure what they can do.

Why. Training can be judged at four levels. Did people like it? Did they learn? Do they work differently? Did anything change for the children?

A rating answers only the first. People can rate one kind of teaching above another and score lower on a test at the end of the class. A randomised study of university physics students found exactly that.

What to ask for, in order of cost.

- A short practical test for each worker, at the start, the end and a month later, on the task the training teaches. Cheap, and it answers the second question.
- A supervisor's checklist at a sample of routine visits. It answers the third.
- The programme's own records on whether children's growth problems are found sooner, in districts that had the training against districts not yet reached. Slow, but it is the only answer to the fourth.

Tie the next round of funding to the first of these at least.

14 · Where teaching sits in Indian medical education

EXERCISE 1. Clinician; leader and member of the health care team and system; communicator; lifelong learner; professional; critical thinker; researcher.

Critical thinker and researcher are the two added after 2019. The first five are the 2019 list.

EXERCISE 2. The second half names the wrong paper. The 2022 rules on who may teach were replaced on 30 June 2025. The new ones are the Medical Institutions (Qualifications of Faculty) Regulations, 2025. Quote those.

The first half claims more than Table E says. Table E names the course for Professors and Associate Professors in broad specialty subjects taught to undergraduates. Its Assistant Professor row does not name it. Community Medicine is taught to undergraduates. So the course will apply when you become an Associate Professor.

Before saying more, check three things. The parts of the 2025 rules not read here. Your college's and your state's own rules. Any later change or written answer from NMC. The table does not say "cannot teach until". Do not claim it does, and do not claim the opposite either.

EXERCISE 3. Start with what the cap does. It forces at least two thirds of a subject's hours into small groups, practicals, clinics, self-directed learning and similar work.

The curriculum states the rule, and the paragraph that sets it cites no study. So the argument has to come from this book. section 10 says to choose activities that make learners do what the test will ask. This book suggests those formats leave more room for that than a large group does.

Then the limit. A small group can still be a lecture to twelve people. The cap counts hours by format, and says nothing about what happens in them. Whether learners can do the objective afterwards is shown only by testing it.

15 · Leading people is a craft of its own

EXERCISE 1. A teaching lineage, a team you run, and an institution that outlives you.

The question is: what happens to it the day you leave? The lineage goes on in the people you taught, and in the people they teach. The team's work stops or passes to its next leader. The institution keeps working without you, because it has written rules, roles and a successor.

EXERCISE 2. The premises may all be true. Dr Rao may be the best physician and the most published. The step that fails is from "best at the work" to "able to lead the people who do the work". Stoller argues that knowing the work and leading others in it are different skills.

"It will run itself" is a second slip. A new clinic is a change. Someone has to decide what it is for, and bring the nurses, dietitians and referring doctors round to it. Someone also has to draw up its rota and its budget.

Before agreeing, ask what Dr Rao has led before: a unit, a project, a team of any size. Ask what she wants to do, since the job will take time from the work she is best at. And ask who will run the clinic day to day if she decides what it is for. None of this says she is the wrong choice. It says her record as a physician does not settle it.

EXERCISE 3. The quotable sentence: "Leading is probably a skill doctors can learn, but we have mostly asked the doctors whether it worked, and hardly ever checked."

Then the detail. One of the best-known writers on leadership, John Kotter, says it does not depend on charisma. Leadership courses for doctors are common: one review found 45 published studies of them by 2013. Every one reported a benefit. But most of the benefit was what the doctors said about themselves, measured straight after the course. Only four studies compared trained doctors with untrained ones. Studies that found nothing were probably less likely to be published.

So: probably learnable, taught often, and tested badly. The evidence is weak, not negative.

16 · Words for how a team works

EXERCISE 1. Planning and budgeting, beside setting a direction.

Organizing and staffing, beside aligning people.

Controlling and problem solving, beside motivating and inspiring.

Management copes with complexity; leadership copes with change.

EXERCISE 2. 1. Authority against power. Being put in charge of the camp gave authority over the camp, not over another department. Getting Paediatrics to send residents needed their agreement, which is aligning, and power from outside the post. 2. Role clarity. No one had been written down as the person who decides, or does, the booking. 3. Conflict about the work: which measure to use, on what evidence. 4. Conflict about the person: a remark about Anita's motives, not about the screening. 5. Delegation not matched to the person. The task was the same, but the intern needed more direction and checking than the senior resident.

EXERCISE 3. One worked sheet, for a session for MBBS students in Community Medicine:

- The objective: you decide, you write it, the batch's tutor is asked first.
- The time slot: the head of department decides, the department office enters it in the timetable, you are asked first.

- The pre-test, post-test and delayed test: you decide and you write them. A colleague who teaches the topic is asked first, to check that they match the objective.
- The teaching: you decide and you do it. A colleague observes, if you want a second view.

A good sheet has a name in every "decides" cell. Where two names appear there, the decision will stall; pick one.

EXERCISE 4. Draw two columns. Write "copes with complexity" over the left and "copes with change" over the right: Kotter's short version.

Under the left, write planning and budgeting, organizing and staffing, controlling and problem solving. Under the right, beside each, write setting a direction, aligning people, motivating and inspiring.

Then take a real change on your unit, for example starting a weight and waist check for every admitted adult. Ask the resident which column each step belongs in. Writing the form and the rota is the left. Getting the nurses and the other units to see why it matters is the right.

Finish with Kotter's warning: strong leadership with weak management is no better than the reverse. A unit where the new check was explained well but nobody wrote the rota will not do it. Nor will one with a perfect rota that nobody believes in. Say that this is Kotter's argument, not a measured result.

Glossary

Every term of art this book teaches, in the plain words it gets where it is first used, and the section that teaches it.

Term	Plain words	Section
2 × 2 table	a table of two rows and two columns, read "two by two"	6
action (of a learning objective)	one verb whose result can be seen or marked	9
aligning people	Kotter: communicating the direction to everyone who can help carry it out or can block it, so that they understand it and are committed to it	16
at the objective's level (of an item)	the item cannot be answered by repeating something said or shown in the session	10
authority	the right to decide or to direct that comes with a post	16
Bloom's taxonomy	a sorted list of kinds of thinking skill; the 2001 revision has six levels, from lowest to highest: remember, understand, apply, analyse, evaluate and create	9
ceiling	(1) the level a saturating shape heads for and flattens against (C6); (2) of a test's scale, the top score, which no score passes (12)	12
civil society organisation	a group formed by citizens, outside government and outside business, to work for a public cause	15
class normalised gain, 〈g〉	Hake's class form of the normalised gain, worked out from class-average pre-test and post-test scores; not the mean of the learners' own normalised gains	12
cognitive load	the demand a learning task makes on working memory	8
comparison group	(1) in a study built to answer a relate question, the people without the exposure, or without the outcome (S55-R1-C03); (2) in a test of a teaching method, the group that gets the usual way (2)	2
competency based (medical education)	the course is built from what a graduate must be able to do, not from a list of topics to cover	14
condition (of a learning objective)	what the learner is given, or is not allowed, while doing it	9
conflict about the person	friction over who someone is: blame, dislike, a remark about their character or motives	16
conflict about the work	disagreement over what to do, how to do it, or on what evidence	16
controlling and problem solving	Kotter: monitoring results against the plan, identifying where they depart from it, and planning and organizing to fix that	16
credibility	Kotter: whether people believe the person giving a direction; it rests on your record, and on whether what you do matches what you say	16

Term	Plain words	Section
criterion task	Dunlosky and colleagues' name for the final test	5
crossover (interaction)	the method that gives one style group its best score is not the one that gives another group its best; in a 2 × 2 table, one row's MD is positive and the other's negative	6
delegation	handing a task, and the authority it needs, to another person, while you stay answerable for the result	16
effect size	a name papers often give an SMD	2
elaborative interrogation	a study technique: the learner works out why a stated fact is true	7
element (in working memory)	a piece of information held as one	8
evaluate (teaching)	to measure some outcome of it	13
extraneous load	cognitive load that comes from how the material is presented and what the learner is made to do; the teacher can change it	8
feedback (on a retrieval attempt)	showing the learner the correct answer after the attempt	3
feeling of learning	the learner's own judgement of how much they learned, usually given as a rating	1
fluency	the ease with which a text reads or a fact comes to mind; it can come from things unrelated to understanding	5
formative (of a test)	used only to show the learner and you where they stand	10
free recall	notes away, each learner writes down all they can about the topic, then checks it against the key points	11
gap (between study sessions)	the time between two study sessions of the same material	4
grand rounds	a teaching talk for a whole department	9
guideline (NMC)	a document NMC's undergraduate board sends out by letter and puts on its website; it binds because a regulation says students are trained as it says	14
highlighting	marking the parts of a text that seem important while reading it	5
illusion of explanatory depth	people feel they understand how things work in far more detail than they do; their rating of their own understanding falls after they write out an explanation	7
Indian Medical Graduate (IMG)	the doctor the MBBS course is meant to produce; the CBME Curriculum 2024 gives the IMG seven roles	14
institution that outlives you	a standing group, with written terms, roles and a named successor, that keeps working after you leave it	15
interleaving	mixing different problem types in one session instead of doing one type in a block	8

Term	Plain words	Section
internal assessment	the CBME Curriculum 2024's marks from day-to-day tests; a student needs at least 50% of them, and 40% in theory and in practical separately, to sit the university exam in that subject	10
intrinsic load	cognitive load that comes from the material itself and from how much the learner already knows; it changes only if what is taught or the learner's knowledge changes	8
item (of a test)	one question or task on a test, marked on its own	10
Kirkpatrick's levels	four kinds of outcome an evaluation of teaching can measure: reaction (Level 1), knowledge (Level 2), behavioural change (Level 3) and system results (Level 4)	13
leadership (to lead people)	getting a group to move towards a change and keeping them at the work it takes (15); in Kotter's terms, coping with change, by setting a direction, aligning people, and motivating and inspiring (16)	15, 16
learning (in S57)	a lasting change in what a person can do, measured by a test taken after a delay	1
learning objective	one sentence stating what a named learner will be able to do at the end of a piece of teaching, in a form someone else could watch or mark	9
learning styles	the idea that people differ in which mode of teaching or study works best for them	6
learning-styles hypothesis	the claim that teaching fitted to a learner's style gives a better learning outcome	6
long-term memory	where what has been learned is kept; it has no known limit, and anything learned comes back from it as one piece	8
low-stakes test	a test whose score counts for little or nothing	3
lurking variable	anything other than the method that affects the score	2
management	Kotter: coping with complexity, by planning and budgeting, organizing and staffing, and controlling and problem solving	16
massed practice	the same study occasions put back to back; often called cramming	4
mean difference (MD)	the mean score of one group minus the mean score of the other; its sign says which scored higher, so state which was subtracted from which	2
mean gain	the mean of a group's raw gains, equal to the mean post-test score minus the mean pre-test score	12
meshing hypothesis	the commonest version of the learning-styles hypothesis: teaching works best in the format that matches the learner's preference	6
motivating and inspiring	Kotter: keeping people moving in the direction, despite obstacles, by appealing to their needs, values and emotions	16

Term	Plain words	Section
normalised gain (g)	the raw gain divided by the room to gain, (post − pre) ÷ (100 − pre); a fraction with no unit, not defined for a pre-test of 100%	12
objective (of how an outcome is measured)	measured by a test, an observation or a record (Frich's B levels)	13
organizing and staffing	Kotter: creating a structure of jobs, filling them with qualified people, communicating the plan to them, delegating responsibility for it, and setting up systems to monitor it	16
parallel forms	two forms of a test in which each item on one has a partner on the other; only the case or the numbers differ	10
performance (in S57)	what a learner can do during a session or within minutes of it	1
planning and budgeting	Kotter: setting targets, laying down detailed steps to reach them, and allocating resources to those steps	16
post-test	the test taken at the end of the teaching	10
power	(1) repeated multiplication of one number by itself, as 10^3 is ten times ten times ten (A6); (2) over people, the capacity to get things done through other people, from any source, authority being only one (16)	16
pre-test	the test taken before the teaching	10
professional body	an association whose members share one profession or discipline	15
proportion reaching the standard	the number of learners whose score meets the objective's standard, divided by the number tested	12
random allocation	a chance process decides which group each learner goes into, not the teacher or the learners	2
raw gain	post-test score minus pre-test score, in percentage points	12
reaction (Kirkpatrick's Level 1)	whether the learners liked the session	13
rereading	studying a text again after a first reading: restudy done with a text	5
response variable	the thing the teaching is meant to change; each Kirkpatrick level names a different one	13
restudy	reading or hearing the same material again, for the same time	3
retention interval	the time between study and the test	1
retrieval practice (practice testing)	an attempt by a learner, during learning, to bring material already met back to mind from memory, by recall or recognition, without looking at it	3
role clarity	a written statement, for each decision or piece of work, of who decides, who does the work, and who is asked first	16
room to gain	100 − pre: the most a learner, or a class, could have gained	12

Term	Plain words	Section
self-explanation	a study technique: the learner explains how new information relates to what they already know, or explains the steps they take while solving a problem	7
session built around retrieval	a session in which the learners spend much of the time bringing material back from memory, each alone and in writing, and seeing the correct answer afterwards	11
session plan	a written table of a session's parts, in order, with the minutes for each	11
setting a direction	Kotter: developing a vision of the future, often the distant future, with strategies for making the changes it needs	16
spacing (distributed practice)	a schedule: the same material studied or retrieved on two or more occasions spread out in time	4
spacing effect	for the same total study time, spaced practice gives better long-term retention than massed practice	4
standard (of a learning objective)	how much, or how well, counts as done	9
standardised mean difference (SMD)	the MD divided by the SD: the MD as a number of standard deviations; which SD is used is part of the definition	2
subjective (of how an outcome is measured)	stated by the learner or someone near them, as an opinion or a rating (Frich's A levels)	13
summative (of a test)	counts towards a mark or a decision about the learner	10
supersession	the new instrument taking the old one's place	14
teaching lineage	the people you teach, and the people they go on to teach	15
team you run	a set of people whose work on one job you direct, for as long as the job lasts	15
testing effect	a retrieval attempt improves memory on a later test more than restudy does; it holds on a delayed test, and on a test minutes later restudy can come out ahead	3
transmission	teaching by telling: the teacher explains and the learners listen, take notes, and ask the odd question	1
utility (of a study technique)	Dunlosky and colleagues' overall rating, low, moderate or high: how widely a technique's benefit holds, and how it compares with the others	5
WEIRD	short for western, educated, industrialized, rich and democratic: a review's label for a set of countries	3
worked example	a problem shown with its full solution	8
working memory	where new information is handled first; it can hold and work on only a limited number of new elements at a time, and only for a short time	8

The series: all 196 books

In build order: Book 0 first, then every Introductory rung, then every Intermediate rung, and so on; within a level, a subject comes after the subjects it depends on. "done" means that book's version 1 is finished. This is Book 8.

GROUND FLOOR		
0	Ground floor	done
INTRODUCTORY		
1	Energy balance and dynamic weight-change modelling · Rung 1	done
2	Mathematics for modelling and inference · Rung 1	done
3	Qualitative and mixed methods · Rung 1	done
4	Food systems, value chains and Indian food policy · Rung 1	done
5	Policy process, framing and advocacy craft · Rung 1	done
6	R, reproducibility engineering and data management · Rung 1	
7	Question selection and research strategy · Rung 1	done
8	**Teaching, mentorship and coalition building · Rung 1 ◀ you are here**	done
9	Scientific writing, visualisation and public communication · Rung 1	
10	Epistemics: philosophy of science and calibration · Rung 1	
11	Adjacent specialist domains: recognise and route · Rung 1	
12	Probability and statistical inference · Rung 1	
13	Integrative metabolism, insulin resistance and energy expenditure regulation · Rung 1	
14	Commercial determinants and corporate political activity · Rung 1	
15	System dynamics and complexity concepts · Rung 1	
16	Indian regulatory architecture and the live policy agenda · Rung 1	
17	Comparative and international policy · Rung 1	
18	Python for simulation, scraping and tooling · Rung 1	
19	Regression modelling for health outcomes · Rung 1	
20	Potential outcomes, identification and target trial emulation · Rung 1	
21	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 1	
22	Appetite, reward and the determinants of intake · Rung 1	
23	Nutrient science for obesity and cardiometabolic risk · Rung 1	
24	Physical activity and sedentary behaviour science · Rung 1	
25	Sleep science and obstructive sleep apnoea · Rung 1	
26	Participatory systems mapping and group model building · Rung 1	
27	Agent-based and network modelling · Rung 1	
28	Clustered, longitudinal, hierarchical and small-area models · Rung 1	
29	Evidence synthesis and meta-analysis · Rung 1	
30	Prediction modelling and machine learning · Rung 1	
31	Measurement error theory and correction · Rung 1	
32	Causal diagrams, structural bias and sensitivity analysis · Rung 1	
33	Study and trial design, power and missing data · Rung 1	
34	Clinical assessment, staging and secondary causes · Rung 1	
35	Indian dietary patterns, guidelines, food composition and diet cost · Rung 1	
36	Circadian biology and chrononutrition · Rung 1	

37	Behavioural theory, choice architecture and intervention design · Rung 1
38	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 1
39	G-methods, time-varying treatment and mediation · Rung 1
40	Instrumental variables and Mendelian randomisation · Rung 1
41	Quasi-experimental policy evaluation · Rung 1
42	Genetics, developmental origins and life-course epidemiology · Rung 1
43	Social epidemiology, equity measurement and surveillance in India · Rung 1
44	Nutritional epidemiology and its pathologies · Rung 1
45	Lifestyle, dietary and behavioural treatment · Rung 1
46	Pharmacotherapy of obesity · Rung 1
47	Metabolic and bariatric surgery: selection and co-management · Rung 1
48	Eating behaviour, mental health and weight stigma · Rung 1
49	Food environment measurement, retail and digital delivery · Rung 1
50	Food demand, elasticity and the economics of externality · Rung 1
51	Implementation science and programme craft · Rung 1
52	Protocol, preregistration, reporting standards and research ethics · Rung 1
53	Life-stage obesity: paediatric, adolescent and pregnancy · Rung 1
54	Ultra-processed food and the effects of processing · Rung 1
55	Activity measurement and compositional 24-hour analysis · Rung 1
56	Fiscal instruments: taxation and subsidy design in India · Rung 1
57	Burden of disease, economic evaluation and decision-analytic modelling · Rung 1
58	The Indian health system and its financing mechanics · Rung 1
59	India's data infrastructure, access and complex survey analysis · Rung 1
60	Ethics of obesity intervention and industry engagement · Rung 1
61	Health financing, HTA and pharmaceutical access in India · Rung 1
INTERMEDIATE	
62	Energy balance and dynamic weight-change modelling · Rung 2
63	Mathematics for modelling and inference · Rung 2
64	Qualitative and mixed methods · Rung 2
65	Food systems, value chains and Indian food policy · Rung 2
66	Policy process, framing and advocacy craft · Rung 2
67	R, reproducibility engineering and data management · Rung 2
68	Question selection and research strategy · Rung 2
69	Teaching, mentorship and coalition building · Rung 2
70	Scientific writing, visualisation and public communication · Rung 2
71	Epistemics: philosophy of science and calibration · Rung 2
72	Probability and statistical inference · Rung 2
73	Integrative metabolism, insulin resistance and energy expenditure regulation · Rung 2
74	Commercial determinants and corporate political activity · Rung 2
75	System dynamics and complexity concepts · Rung 2
76	Indian regulatory architecture and the live policy agenda · Rung 2
77	Comparative and international policy · Rung 2
78	Python for simulation, scraping and tooling · Rung 2
79	Regression modelling for health outcomes · Rung 2
80	Potential outcomes, identification and target trial emulation · Rung 2
81	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 2

82	Appetite, reward and the determinants of intake · Rung 2
83	Nutrient science for obesity and cardiometabolic risk · Rung 2
84	Physical activity and sedentary behaviour science · Rung 2
85	Sleep science and obstructive sleep apnoea · Rung 2
86	Participatory systems mapping and group model building · Rung 2
87	Agent-based and network modelling · Rung 2
88	Clustered, longitudinal, hierarchical and small-area models · Rung 2
89	Evidence synthesis and meta-analysis · Rung 2
90	Prediction modelling and machine learning · Rung 2
91	Measurement error theory and correction · Rung 2
92	Causal diagrams, structural bias and sensitivity analysis · Rung 2
93	Study and trial design, power and missing data · Rung 2
94	Clinical assessment, staging and secondary causes · Rung 2
95	Indian dietary patterns, guidelines, food composition and diet cost · Rung 2
96	Circadian biology and chrononutrition · Rung 2
97	Behavioural theory, choice architecture and intervention design · Rung 2
98	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 2
99	G-methods, time-varying treatment and mediation · Rung 2
100	Instrumental variables and Mendelian randomisation · Rung 2
101	Quasi-experimental policy evaluation · Rung 2
102	Genetics, developmental origins and life-course epidemiology · Rung 2
103	Social epidemiology, equity measurement and surveillance in India · Rung 2
104	Nutritional epidemiology and its pathologies · Rung 2
105	Lifestyle, dietary and behavioural treatment · Rung 2
106	Pharmacotherapy of obesity · Rung 2
107	Metabolic and bariatric surgery: selection and co-management · Rung 2
108	Eating behaviour, mental health and weight stigma · Rung 2
109	Food environment measurement, retail and digital delivery · Rung 2
110	Food demand, elasticity and the economics of externality · Rung 2
111	Implementation science and programme craft · Rung 2
112	Protocol, preregistration, reporting standards and research ethics · Rung 2
113	Life-stage obesity: paediatric, adolescent and pregnancy · Rung 2
114	Ultra-processed food and the effects of processing · Rung 2
115	Activity measurement and compositional 24-hour analysis · Rung 2
116	Fiscal instruments: taxation and subsidy design in India · Rung 2
117	Burden of disease, economic evaluation and decision-analytic modelling · Rung 2
118	The Indian health system and its financing mechanics · Rung 2
119	India's data infrastructure, access and complex survey analysis · Rung 2
120	Ethics of obesity intervention and industry engagement · Rung 2
121	Health financing, HTA and pharmaceutical access in India · Rung 2
ADVANCED	
122	Energy balance and dynamic weight-change modelling · Rung 3
123	Mathematics for modelling and inference · Rung 3
124	Qualitative and mixed methods · Rung 3
125	Food systems, value chains and Indian food policy · Rung 3
126	Policy process, framing and advocacy craft · Rung 3

127	R, reproducibility engineering and data management · Rung 3
128	Question selection and research strategy · Rung 3
129	Teaching, mentorship and coalition building · Rung 3
130	Scientific writing, visualisation and public communication · Rung 3
131	Epistemics: philosophy of science and calibration · Rung 3
132	Probability and statistical inference · Rung 3
133	Integrative metabolism, insulin resistance and energy expenditure regulation · Rung 3
134	Commercial determinants and corporate political activity · Rung 3
135	System dynamics and complexity concepts · Rung 3
136	Indian regulatory architecture and the live policy agenda · Rung 3
137	Regression modelling for health outcomes · Rung 3
138	Potential outcomes, identification and target trial emulation · Rung 3
139	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 3
140	Appetite, reward and the determinants of intake · Rung 3
141	Nutrient science for obesity and cardiometabolic risk · Rung 3
142	Physical activity and sedentary behaviour science · Rung 3
143	Sleep science and obstructive sleep apnoea · Rung 3
144	Participatory systems mapping and group model building · Rung 3
145	Clustered, longitudinal, hierarchical and small-area models · Rung 3
146	Evidence synthesis and meta-analysis · Rung 3
147	Measurement error theory and correction · Rung 3
148	Causal diagrams, structural bias and sensitivity analysis · Rung 3
149	Study and trial design, power and missing data · Rung 3
150	Clinical assessment, staging and secondary causes · Rung 3
151	Indian dietary patterns, guidelines, food composition and diet cost · Rung 3
152	Circadian biology and chrononutrition · Rung 3
153	Behavioural theory, choice architecture and intervention design · Rung 3
154	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 3
155	G-methods, time-varying treatment and mediation · Rung 3
156	Instrumental variables and Mendelian randomisation · Rung 3
157	Quasi-experimental policy evaluation · Rung 3
158	Genetics, developmental origins and life-course epidemiology · Rung 3
159	Social epidemiology, equity measurement and surveillance in India · Rung 3
160	Nutritional epidemiology and its pathologies · Rung 3
161	Lifestyle, dietary and behavioural treatment · Rung 3
162	Pharmacotherapy of obesity · Rung 3
163	Metabolic and bariatric surgery: selection and co-management · Rung 3
164	Eating behaviour, mental health and weight stigma · Rung 3
165	Food environment measurement, retail and digital delivery · Rung 3
166	Food demand, elasticity and the economics of externality · Rung 3
167	Implementation science and programme craft · Rung 3
168	Protocol, preregistration, reporting standards and research ethics · Rung 3
169	Life-stage obesity: paediatric, adolescent and pregnancy · Rung 3
170	Ultra-processed food and the effects of processing · Rung 3
171	Activity measurement and compositional 24-hour analysis · Rung 3
172	Fiscal instruments: taxation and subsidy design in India · Rung 3
173	Burden of disease, economic evaluation and decision-analytic modelling · Rung 3

174	The Indian health system and its financing mechanics · Rung 3
175	India's data infrastructure, access and complex survey analysis · Rung 3
176	Ethics of obesity intervention and industry engagement · Rung 3

EXPERT

177	Energy balance and dynamic weight-change modelling · Rung 4
178	R, reproducibility engineering and data management · Rung 4
179	Question selection and research strategy · Rung 4
180	Probability and statistical inference · Rung 4
181	Indian regulatory architecture and the live policy agenda · Rung 4
182	Regression modelling for health outcomes · Rung 4
183	Potential outcomes, identification and target trial emulation · Rung 4
184	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 4
185	Appetite, reward and the determinants of intake · Rung 4
186	Measurement error theory and correction · Rung 4
187	Causal diagrams, structural bias and sensitivity analysis · Rung 4
188	Circadian biology and chrononutrition · Rung 4
189	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 4
190	Genetics, developmental origins and life-course epidemiology · Rung 4
191	Nutritional epidemiology and its pathologies · Rung 4
192	Eating behaviour, mental health and weight stigma · Rung 4
193	Activity measurement and compositional 24-hour analysis · Rung 4
194	India's data infrastructure, access and complex survey analysis · Rung 4
195	Ethics of obesity intervention and industry engagement · Rung 4

OBESITY EXPERTISE · BOOK 8

Teaching, mentorship and coalition building · Rung 1

A short book on teaching and on the first step of leading. It covers what learning research has found about retrieval, spacing and re-reading, why "learning styles" do not hold up, how to write an objective that can be assessed, how to run a session around retrieval rather than lecture, how to measure whether it worked, and why leading people is a craft of its own.

Book 8 of the series' 196 books, free to read at drharshmaheshwari.com.



Dr. Harsh Maheshwari
MD Community Medicine Resident, AIIMS Raipur
contact@drharshmaheshwari.com
ORCID 0000-0001-8848-1918
drharshmaheshwari.com

Version 1.0 · 25 September 2026

CC BY-NC-SA 4.0 · free to share, not to sell