

CHAPTER XII

# Mechanisms Within Practice

---

*“Knowledge does not keep any better than fish.”*

Alfred North Whitehead, “Universities and Their Function”

## 🐟 The Times Tables

In the last week of third grade, two girls sit the same times-table test: a hundred facts in five minutes. Nina gets seventy right. Ella gets all hundred. On Monday the class starts long division, and both girls go with it.

Nina spends the fall in trouble. Long division is a stack of multiplications, and she doesn't know a third of them. Every problem that needs  $7 \times 8$  stops her, and while she counts on her fingers she loses her place in the division. By November she has decided she's bad at math.

Ella's trouble comes later and more quietly. She knew all hundred facts in May. Nobody asked her for them over the summer. In September,  $6 \times 9$  takes her a second too long, then two, and she starts guessing.

The school asked one question, once: did you pass? Two others went unasked. How good must a skill be before you move on? Nina's seventy was

waved through. And when should a skill come back, so it doesn't fade? Ella's hundred never came back at all.

Practice has a mechanism for each question. **Mastery learning** is the depth gate: it sets how good a skill must be before the learner moves on. **Spaced repetition** is the timing schedule: it sets when the skill returns. Both carry large experimental literatures, which the model inherits.<sup>1</sup> Both work better when the engagement drivers of Chapters 5 to 11 are set well around them. And the two compete for the same minutes.

This chapter stays inside Practice. Depth and repetition on the Inquiry side are taken up in Chapter 13,<sup>2</sup> along with the developmental science behind the model.<sup>3</sup> How the two mechanisms change form across Egan's stages comes in Chapter 14.

## • The Depth Gate

### *Aptitude as a Rate*

Picture a typical unit on fractions. Everyone gets four weeks. At the end, a test sorts the room into As, Cs, and Fs, and the grades are read as a measure of ability: some kids are math people and some aren't. Time is fixed, and achievement varies.

In 1963 John Carroll turned that around. He proposed that the degree of learning depends on the time a learner spends relative to the time that learner needs. Aptitude, in his model, is a rate. Some learners need more time to reach the same place. One climber takes the cable car and another walks the switchbacks, and the summit is the same summit.

Benjamin Bloom (1968) turned Carroll's idea into a teaching method he called Learning for Mastery. Achievement became the constant, and time the variable.

### *Bloom's Cycle*

As Thomas Guskey summarizes it, the method runs in units of one to two weeks:

1. Teach the unit.
2. Give a formative assessment to find what each learner hasn't yet learned. It carries no grade. It works like the scan a mechanic plugs into your car to read the error codes.
3. Supply correctives to learners below the standard: more time, and a different approach. Saying the same thing again, louder, doesn't count.
4. Supply enrichment to learners who reached the standard, so the quick ones aren't left waiting.
5. Give a second, parallel assessment to confirm the correctives worked.

That second assessment is the gate. The learner moves on when the skill holds.

### *Keller's Version*

At about the same time, Fred Keller and his colleagues built the Personalized System of Instruction (PSI) for the new University of Brasília in the mid-1960s, and described it in an article titled "Good-bye, teacher..." (1968). A course was divided into units of written material. Learners moved at their own pace. Each unit ended in a test that could be retaken, in equivalent forms, until the learner showed mastery, and proctors, often advanced students who had finished the course, certified it. Lectures stayed, with a new job. They served motivation, and the content traveled through the written units and their tests. You went to the lecture to see why the physics mattered, and learned the physics at your own desk.

### *What the Evidence Shows*

It's an egalitarian idea, and an appealing one. Does it work? Chen-Lin Kulik, James Kulik, and Robert Bangert-Drowns (1990) analyzed 108 controlled evaluations of mastery programs.

- The average effect on examination performance was 0.52 standard deviations. Line up a hundred students by score: a student in the middle, at the 50th place, moves to roughly the 70th.
- Keller's approach averaged 0.48 and Bloom's 0.59. The difference wasn't significant.
- Lower-aptitude students gained more (0.61) than higher-aptitude students (0.40). The gate helps most the learners who would otherwise move on with gaps. Learners like Nina.
- Effects were larger on locally developed tests (0.57) than on standardized ones (0.29).
- Higher standards did better. Programs requiring 91–100 percent correct averaged 0.64, against 0.44 for 81–90 percent and 0.49 for 70–80 percent.
- Mastery programs took about 4 percent more instructional time, and self-paced courses had lower completion rates than conventional ones.

### *The Critique*

If the gate works, why isn't every school built around it? Robert Slavin (1987) reviewed group-based mastery learning in elementary and secondary schools, in studies lasting at least four weeks. On standardized achievement tests he found essentially no evidence of effects. On tests the experimenters made, the effects were positive but moderate, with little evidence that they lasted.

He named the problem the **coverage-versus-mastery dilemma**. A school year has a fixed number of hours. Time spent bringing every learner to the standard is time not spent moving the class forward. A conventional class might get through twenty chapters, most students half understanding them. A mastery class might get through twelve, well. The year-end test asks about all twenty.

Self-pacing escapes that trap and walks into another. Take the clock away and many learners drift, which is what the lower completion rates show.

### *Reading the Two Together*

The gate works best on what it tests. Where the test measures the unit's own skills, the effects are real. On broad standardized measures, they shrink. For the model, that's what a gate should do. A mastery gate is a claim about a specific skill in a specific setting, which is what Kurt Fischer says a skill is.<sup>4</sup> Transfer to other settings has to be earned separately.

Gate a learner on quadratic equations set out on a worksheet, and she masters worksheet quadratics. A month later the same algebra arrives inside three paragraphs about a rocket's trajectory, and she stalls. The gate certified the skill it tested. Picture the inspection station on a car assembly line that checks the paint: it passes a car with a perfect finish and no brakes, and by its own lights it's right.

## Building a Gate

A mastery gate has three parts.

**A standard.** How well: a percentage correct, a performance without errors, a speed. The evidence favors high standards.

**A measure.** A test that samples the skill itself, beyond recall of facts about it.

**A consequence.** Below the standard, correctives and a retake. Above it, advancement or enrichment.

### *How Gates Fail*

**Set too low.** The learner moves on with gaps, and the gaps compound. Nina's seventy percent went into long division with her. Long division goes into fractions, and fractions into algebra. A low standard does the most damage in foundational skills, the way a missing third of the concrete does the most damage in a foundation slab.

**Testing recall instead of skill.** The gate passes learners who can recognize an answer and can't perform. You may have spent months on a language app, scoring a hundred percent at picking the right verb form from four choices, and then frozen when a waiter asked you a quick question in that language. The app had gated multiple-choice recognition, and you'd passed it. Guskey notes that early mastery programs narrowed mastery to low-level skills, which Bloom never intended.

**Correctives that repeat the lesson.** The default response to a learner who didn't get it is the same explanation again, slower, or in bold. Bloom's correctives were a different approach: a picture where there was a paragraph, an object where there was a symbol, a classmate where there was a lecture. A train that stalled on one track will stall there again.

**Pacing collapse.** Group pacing forces the coverage dilemma Slavin named. Self-pacing without support loses learners, as the completion rates show. The group freezes, or the individuals float off.

**A gate without a return.** Mastered once is not mastered for good. Ella passed. A skill that clears the gate and never comes back decays, which is the problem the second mechanism solves.

### *Beyond the Gate: Style*

The gate asks whether a skill is solid. Whitehead pointed past it: "Style is the ultimate morality of mind." By style he meant the economy and exactness of work done well. Think of a carpenter looking at a dovetail joint. A gate could confirm that the joint holds. The carpenter is looking at something more: no wasted cut, the grain read correctly, a fit so close it looks inevitable. That's craft at its fullest.

Whitehead tied style to how a learner comes to hold a subject: "The love of a subject in itself and for itself ... is the love of style as manifested in that study."

The gate is where Practice certifies a skill. Style is where Practice is headed. It's also an answer to the child who asks why she has to keep drilling facts she already knows, and to the piano student sick of scales. Precision is what makes

style possible. You drill the scales until your fingers stop needing you, and then your mind is free for the music.

## • The Timing Schedule

### *The Forgetting Curve*

Hermann Ebbinghaus (1885) measured how fast learning fades, with himself as the only subject. He memorized lists of nonsense syllables, made-up strings like *wid* and *zof* chosen so nothing he already knew could help him, until he could recite them perfectly. Then he tested how much he'd kept after twenty minutes, an hour, a day, and longer. The result was the **forgetting curve**: a steep early drop, then a long, slow tail. In the same work he identified the **spacing effect**. Repetitions spread out over time held better than the same number of repetitions massed together.

The curve has aged well. A single-subject replication in 2015 (Jaap Murre and Joeri Dros) produced a forgetting curve close to his, 130 years later.

Whitehead's fish makes the same point with less arithmetic. He meant that knowledge goes stale unless it's kept in use. Spacing and retrieval are the Practice system's way of keeping it fresh. Worth and Inquiry keep it in use.

### *How Large and How General*

Nicholas Cepeda, Harold Pashler, Edward Vul, John Wixted, and Doug Rohrer (2006) synthesized 839 assessments of distributed practice, from 317 experiments in 184 articles. Spaced study reliably beat massed study. And the gap that produced the best retention grew as the retention interval grew: the longer the learning had to last, the longer the best gap.

### *The Gap Depends on the Goal*

Cepeda and colleagues (2008) then tested more than 1,350 people, with review gaps of up to three and a half months and final tests up to a year later. The best gap was about 20 percent of the test delay when the delay was a few

weeks, falling to about 5 percent when the delay was a year. Lengthening the gap helped up to that point and hurt beyond it.

- To hold a skill for five weeks, return after about a week.
- To hold it for a year, return after about two and a half weeks.

Come back too soon and the memory is still fresh, so retrieving it takes little effort and does little. Come back too late and there's little left to retrieve, and the review turns into relearning.

So every schedule starts from one question: how long does this have to last?

## 📖 Retrieval and Fluency

### *What to Do on Return*

Timing says when to return. Retrieval says what to do once you're there.

Henry Roediger and Jeffrey Karpicke (2006) had students either restudy prose passages or take recall tests on them.

- Five minutes later, restudying looked better: 81 percent recalled against 75.
- A week later, testing won: 56 percent against 42.
- Students who studied a passage once and were tested three times recalled 61 percent a week later. Students who studied it four times recalled 40 percent, after reading it about four times as often.

Reading the instructions for a push-up four times won't make you stronger. Doing three push-ups will.

John Dunlosky, Katherine Rawson, Elizabeth Marsh, Mitchell Nathan, and Daniel Willingham (2013) rated ten common study techniques. Only two earned high utility: practice testing and distributed practice. Rereading and highlighting, the techniques students use most, rated low.

### *Fluency Fools*

In the moment, massed rereading feels better and scores better, which is why learners choose it. On the fourth pass through a chapter the sentences go down easily, and the ease feels like knowing. You recognize every line. Recognizing a line and producing it from a blank page are two different acts, and the test, the meeting, and the waiter will ask for the second. A page painted yellow with highlighter is a record of reading.

This is the gap between performance and learning from Chapter 5,<sup>5</sup> and it's why the schedule has to be set from outside the learner's feeling of progress, at least at first. Left to that feeling, learners cram, and cramming feels like it's working.

### *What the Evidence Covers*

The largest synthesis covered verbal recall: facts, words, and passages. Evidence for procedures, such as mathematics, and for motor skills is smaller. Extending the schedule to them is reasonable, and less settled.

### *Schedulers*

Working out 20 percent of five weeks for vocabulary word 42, and 5 percent of a year for physics formula 6, would defeat anyone. The schedule has moved steadily from rule of thumb to model, and the arithmetic has moved off the learner's desk.

**Leitner boxes.** Sebastian Leitner's *So lernt man lernen* (1972) put flashcards in a row of boxes. A card answered correctly moves to the next box, reviewed less often. A card missed goes back to the first box, reviewed daily, however far along it had traveled. The hardest cards get the most time with no calculation at all. (A few shoeboxes and a pack of index cards will do.)

**SuperMemo.** Piotr Wozniak began a paper-based spacing schedule in 1985, while learning English vocabulary, and wrote the first SuperMemo program in 1987. Its early algorithm, SM-2, set each card's next interval from how easily the learner recalled it. SM-2 became the basis of Anki and many other flashcard programs.

**Anki and FSRS.** Anki's original scheduler is a version of SM-2. Version 23.10 added FSRS as an alternative. It models each card's difficulty, its stability, and its retrievability, the probability of recall at a given moment. The learner sets a desired retention, say 90 percent, and the scheduler times each review to land as recall falls to that level. Anki reports that FSRS needs fewer reviews than SM-2 for the same retention.

**Duolingo's half-life regression.** Burr Settles and Brendan Meeder (2016) estimated the half-life of each word in each learner's memory from that learner's history, the way a physicist describes a decaying isotope. The model cut prediction error by more than 45 percent against baselines, and in an operational test it raised daily student engagement by 12 percent.

Every scheduler answers the question Cepeda's research asks: how long can the gap be before the skill falls too far? Desired retention, the number a learner types into FSRS, is the design choice the research says matters most, stated as a setting.

## The Two Together

### *A Division of Labor*

The gate works before advancement: don't move on until the skill holds. The schedule works after it: bring the skill back before it fades. A gate without a schedule produces Ella, a skill that passes once and decays. A schedule without a gate spaces out reviews of skills that were never solid. That's Nina with a deck of flashcards.

The gate works before advancement; the schedule works after it

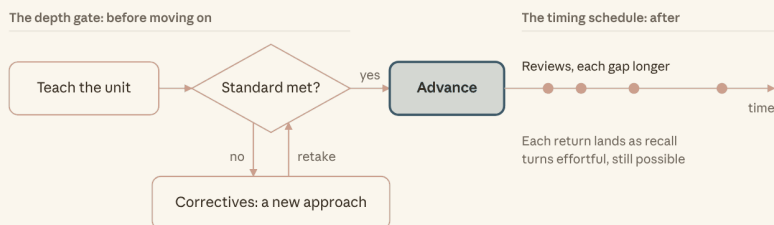


Figure 12.1 · the gate and the schedule (schematic)

A learner below the standard loops through correctives until the skill holds. After the gate, the skill returns at widening gaps.

### *Where They Pull Apart*

Both mechanisms spend the same minutes. Every review that falls due competes with new material waiting behind a gate, and the review load grows with everything the learner has mastered. Anyone who has kept up a flashcard deck knows the arc: five reviews a day, then fifty, then two hundred. Each skill pushed through the gate adds a small recurring charge to every future day. Left alone, the charges eat the budget, and the learner spends every session paying interest on old learning with nothing left over for new.

This is Slavin's coverage dilemma again, now inside one learner's day.

### *A Worked Example: Math Academy*

Math Academy runs both mechanisms in one system, and its designers have described how (Justin Skycak, 2023).

**The gate.** Learners reach each topic through its chain of prerequisites, and must master it before the topics built on it open.

**The schedule.** Once mastered, a topic enters spaced review, with timing calibrated to how fast this learner learns this topic.

**Implicit repetition.** Solving an advanced problem practices the simpler skills inside it. A multistep equation exercises the arithmetic and the fraction rules folded into it. Under Fractional Implicit Repetition (FIRE), a correct

answer on an advanced topic counts as partial review of its prerequisites. A failure on a prerequisite flows forward and lowers confidence in the topics that depend on it.

**Compression.** The system chooses tasks whose implicit reviews clear other due reviews, “like dominoes,” so much of the review load is carried by new learning instead of by separate drills.

### *What the Example Shows*

Once skills are nested, practicing the whole reviews the parts. The review load shrinks, and review arrives inside work that moves forward. This is the behavior graph of Chapter 16 run as a scheduler: skills as nodes, composition as edges, credit and penalty traveling along them.<sup>6</sup>

The same idea reaches well beyond mathematics. A web developer doesn't need a flashcard to remember how to write a link; the site she builds every week asks her for one every few days. A project that uses a skill regularly reviews it with no card in sight, and that's where worth comes in.

### *A Caution*

The description comes from the company's own designers. It explains the design clearly, and it isn't an independent evaluation of results. The Alpha School breakout in Chapter 2 places Math Academy near the pure-Practice end of the stack, legitimate Practice on its own terms.

## The Drivers Around the Mechanisms

The drivers of Chapters 5 to 11 don't replace the two mechanisms. They decide whether the learner stays inside them long enough for them to work, and whether what they produce is worth having.

## *Fit*

Spacing and retrieval are the standard examples of desirable difficulties: they lower performance now and raise learning later. A good schedule is fit applied to time. Each review lands when recall is effortful and still possible.

A mastery gate keeps the learner at the edge of reach. It stops advancement onto material that depends on skills the learner doesn't yet hold, which is where Nina spent her fall.

Both mechanisms fail the same way when fit is wrong. Reviews that come too soon are too easy. A gate held too high for too long becomes repetition without progress. Reading fit tells you which. Speed without attention says the review came too early. Freezing says it came too late, or the gate sits too far off.

## *Fidelity*

A gate is only as good as its measure. A mastery test on a worksheet certifies worksheet skill, and the coconut sellers of Chapter 6 show how little a skill may travel between settings: arithmetic flawless at the stall, shaky on paper.

Functional fidelity is the standard for the gate. The test should make the world answer back on the part being learned: change counted into a real hand, a passage played for a listener, a recipe actually scaled and baked.

Reviews can raise fidelity as the schedule stretches, with early reviews quick and abstract and later ones in the setting where the skill will be used. The girl at the farm stand in Chapter 6 got exactly that: round prices first, then messy ones, then the cash box on her own.

## *Worth*

Drill that leads toward value work holds attention better than drill for its own sake, because the learner can see the use waiting for the skill.

A gate can be set by use. The skill is mastered when the work that depends on it can be done and used. The letter goes out and gets an answer. The shelf holds the books. The script runs and the spreadsheet updates itself.

Real work can carry the schedule, too. A project that needs a skill every few days spaces its reviews without a flashcard in sight.

### *Ownership*

Practice ends with an owned skill, and the gate is the moment of handover. Crossing it should be marked, and the learner should know what they now own: here's what you can do, and it's yours.

The gate protects most the learners who would otherwise move on with gaps. Lower-aptitude students gained the most from mastery programs.

Self-paced courses lost more learners. Freedom over pace without structure becomes drift, which is Chapter 8's lesson: follow the pull, supply the structure.

As competence grows, hand over the schedule itself. A learner who understands why spacing works can set their own reviews. The adult owns the goal, and the learner comes to own the method.

### *Outside Drivers*

Chapter 11 proposed that structural drivers suit Practice. Spaced-repetition software shows why: its streaks, progress bars, and review counts are structural drivers wrapped around a schedule.

Mastery gates produce the clearest progress signal Practice has, a unit crossed. That's progress pointed at the object.

Streaks keep a daily schedule running, and people come to pursue them as goals in themselves. The risk is reviews rushed to keep the count. It's 11:58 p.m., your eyes are half closed, and you tap through forty cards to move the number from 364 to 365. The scheduler believes the taps. A guessed card marked easy gets pushed weeks out, and the model of your memory drifts away from your memory. That's capture that never becomes engagement, and it quietly breaks the schedule too.

Rewards for reviews completed buy the quantity of practice. Its quality has to come from somewhere else.

Chapter 11 recommended mastery standards wherever ranking serves no purpose. A mastery gate is that recommendation in full. It's task-referenced: the standard is the task, and nobody else's score enters into it. That's the kind of competition Chapter 10 favors.

The test for each driver: does it keep the method honest, with reviews done by retrieval and gates passed by skill? Or does it reward the appearance of both, reviews clicked through and tests gamed?

## • Back to the Times Tables

Run third grade again with both mechanisms in place. Nina meets a gate at seventy percent and waits on long division. She gets a different approach (beans laid out in rows on the table, a classmate quizzing her at recess) and a second test a week later. Ella passes, and her facts come back over the summer: a few days later, then a couple of weeks, then once a month, folded into card games and the price of three ice creams at the pool. In September both girls start long division, and neither one counts on her fingers.

The schedulers will keep getting better. A model that knows the half-life of every word in your head could, in principle, follow you past the flashcards into everything you read and time a return for all of it. That might free the mind for style, with the basics held for you. It might also fill every spare hour with maintenance and leave no room for the slow, inefficient business of finding something new, which is Inquiry's work. The mechanism can't settle which. The people who set it up can.

## • Open Questions

- How should a gate be measured for young children, who can't sit a formal test?
- Do expanding intervals beat equal ones, and does the answer differ for skills and for facts?

- How far does the spacing evidence, built mostly on verbal recall, carry to procedures and motor skills?
- When the two mechanisms compete for the same minutes, a review falling due while new material waits behind a gate, which should win?
- Can a mastery gate test transfer to a new setting, beyond the unit's own skill?
- At what age, and by what signs, can a learner take over their own schedule?

If you've run a gate or a schedule with a child, a team, or yourself, I'd like to hear where it held and where it broke.

📌 Notes

1. See Chapter 2, “Three Matrices.”
2. See Chapter 13.
3. See Chapter 13, “Why Each Line Must Be Raised on Purpose.”
4. See Chapter 13, “Why Each Line Must Be Raised on Purpose.”
5. See Chapter 5, “Desirable Difficulties.”
6. See Chapter 16, “The Knowledge Graph and the Behavior Graph.”

📌 Sources

| SOURCE                                                                                    | USE                                             |
|-------------------------------------------------------------------------------------------|-------------------------------------------------|
| John Carroll, “A Model of School Learning,” <i>Teachers College Record</i> (1963)         | Time spent versus time needed                   |
| Benjamin Bloom, “Learning for Mastery,” <i>Evaluation Comment</i> 1 (1968)                | The mastery cycle                               |
| Thomas Guskey, “Mastery Learning”                                                         | Bloom's cycle summarized; early misapplications |
| Fred Keller, “Good-bye, teacher...,” <i>Journal of Applied Behavior Analysis</i> 1 (1968) | The Personalized System of Instruction          |

| SOURCE                                                                                                                                                  | USE                                               |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------|
| Chen-Lin Kulik, James Kulik, and Robert Bangert-Drowns, <i>Review of Educational Research</i> 60 (1990)                                                 | Meta-analysis of 108 mastery programs             |
| Robert Slavin, “Mastery Learning Reconsidered,” <i>Review of Educational Research</i> 57 (1987)                                                         | The critique; the coverage-versus-mastery dilemma |
| Hermann Ebbinghaus, <i>Über das Gedächtnis</i> (1885)                                                                                                   | The forgetting curve and the spacing effect       |
| Jaap Murre and Joeri Dros, <i>PLOS ONE</i> (2015)                                                                                                       | Replication of the forgetting curve               |
| Nicholas Cepeda, Harold Pashler, Edward Vul, John Wixted, and Doug Rohrer, <i>Psychological Bulletin</i> 132 (2006)                                     | Synthesis of 317 spacing experiments              |
| Nicholas Cepeda, Edward Vul, Doug Rohrer, John Wixted, and Harold Pashler, <i>Psychological Science</i> 19 (2008)                                       | The best gap as a share of the retention interval |
| Henry Roediger and Jeffrey Karpicke, <i>Psychological Science</i> 17 (2006)                                                                             | Retrieval practice versus restudy                 |
| John Dunlosky, Katherine Rawson, Elizabeth Marsh, Mitchell Nathan, and Daniel Willingham, <i>Psychological Science in the Public Interest</i> 14 (2013) | Ten study techniques rated                        |
| Sebastian Leitner, <i>So lernt man lernen</i> (1972)                                                                                                    | Leitner boxes                                     |
| Piotr Wozniak, SuperMemo                                                                                                                                | Early computerized scheduling                     |
| Anki FAQ: scheduling algorithms                                                                                                                         | SM-2 and FSRS                                     |
| Burr Settles and Brendan Meeder, <i>Proceedings of ACL</i> (2016)                                                                                       | Half-life regression at Duolingo                  |
| Justin Skycak, “Optimized, Individualized Spaced Repetition in Hierarchical Knowledge Structures” (2023)                                                | Math Academy’s Fractional Implicit Repetition     |
| Alfred North Whitehead, “The Aims of Education,” in <i>The Aims of Education and Other Essays</i> (1929)                                                | Style beyond the mastery gate                     |
| Whitehead, “Universities and Their Function,” in <i>The Aims of Education and Other Essays</i> (1929)                                                   | Knowledge that does not keep; epigraph            |

*From Inquiry and Practice, by Anatoly Lezhen. Working draft, October 2026.*