

Critical Thinking for Life

Michael Gannotti



Figure 1: Cover

Critical Thinking for Life

Michael Gannotti

Teaching-manual edition 30 August 2026. Chicago notes run in one series.
Chapter-opening images are original still-lives, not portraits of persons or children.
Front matter is unillustrated.

Contents

- Welcome
- How to Use This Book This Week
- If You Only Remember Five Things
- The Practice Hour
- Chapter 1 — What we mean
- Chapter 2 — You need something to think about
- Chapter 3 — Maxims are not enough
- Chapter 4 — How to practice this week
- Chapter 5 — The test is not the thing
- Chapter 6 — Leave the page
- Chapter 7 — Stakes
- Chapter 8 — When a machine finishes
- Chapter 9 — Records a stranger can read
- Chapter 10 — Programs, placement, and extra help

Welcome

This book exists because the thinking at your table is the whole program. There is no department down the hall and no specialist waiting after lunch. There is you, a student, and this week's subject. That is enough, if you know what to do with the hour.

I wrote this for a capable, busy, willing parent. You may be teaching two ages at once. You may be fitting a history chapter, a science lab, and a news story between a job, a toddler, and a grocery run. You may have loved school debates, or you may remember “critical thinking” as a poster on a classroom wall. Many adults feel rusty when they sit down to teach judgment. That feeling is common. It is not a verdict. We will move on from it. This book will make you fluent enough in this week's idea to notice a wrong turn and ask a good question — without taking over the pencil.

You do not need to be a philosopher. You do need a sentence you can actually use. Robert Ennis's is the one this book keeps: critical thinking is reasonable, reflective thinking about what to believe or do.¹ Reasonable means there are reasons, not only a feeling. Reflective means you pause long enough to look at those reasons. And the last two words matter. A life is not only appraisal of other people's arguments. It is sharing, consenting, hiring, voting, refusing. The thinking includes action.

That sentence names the activity. It does not, by itself, tell you how to teach it. Daniel Willingham's constraint does. People have treated critical thinking as a

¹Robert H. Ennis, “A Logical Basis for Measuring Critical Thinking Skills,” *Educational Leadership* 43, no. 2 (1985): 44–48. Restated in Robert H. Ennis, “Critical Thinking: A Streamlined Conception,” later outlines, and the 2011 recollection in *Inquiry*. The public sentence: “Critical thinking is reasonable reflective thinking focused on deciding what to believe or do.” Access date for URLs in these notes: 30 August 2026.

skill like riding a bicycle — once learned, usable in any situation. Thinking is not that sort of skill. The processes of thinking are mixed with the content of thought. You can teach maxims. Without knowledge and practice in a real subject, the student cannot put the maxims to work.² Even a three-year-old can think carefully in a domain they know. Even a trained scientist can fail outside theirs. That is encouraging, not discouraging. It means this week’s work sits inside history, science, mathematics, or the news on the table — not in a content-free hour labeled “thinking.”

About 3.4 percent of U.S. students ages 5–17 were homeschooled in 2022–23 — roughly 1.765 million children.³ That figure is context, not a ranking. You are one of those tables. The hour in front of you is the work.

What a good thinking hour looks like

You sit down already knowing today’s idea, and already knowing the subject it lives in. The student warms up on facts they can already get right in that subject. You model one judgment out loud, short. Then they try, and you wait. You ask one good question — not “did you think critically?” but “what would you need to see to change your mind?” Practice mixes the new move with last week’s. The hour ends with one or two problems they do alone, book closed, no hints. You stop talking sooner than feels polite. The student holds the pencil. You hold the key. That is the hour. The next long piece of this front matter, *The Practice Hour*, will teach it in full. Later chapters will not reinvent it.

²Daniel T. Willingham, “Critical Thinking: Why Is It So Hard to Teach?,” *American Educator* 31, no. 2 (Summer 2007): 8–19. Authorized reprint at Reading Rockets. Thinking “is not that sort of skill”; “the processes of thinking are intertwined with the content of thought (that is, domain knowledge).” Domain-specific critical thinking can be taught; there is no proven way to teach a content-free general skill directly (restated in Willingham, NSW Department of Education paper, 2019).

³Ellena Sempeles and Jiashan Cui, *Parent and Family Involvement in Education: 2023*, NCES 2024-113 (Washington, DC: National Center for Education Statistics, September 2024), Table A-6: 3.4 percent homeschooled, approximately 1,765,000 students, ages 5–17 with a K–12 grade equivalent, 2022–23. <https://nces.ed.gov/pubs2024/2024113.pdf>. Latest federal count this book uses; not a 2026 national total.

What you will actually get

Each teaching chapter does eight jobs, always in the same order, so you are never hunting for the move.

You will learn why this week's idea is worth the struggle — what it unlocks later. You will understand the idea yourself, in plain language, with one everyday picture and one precise picture, and with the wrong answers you should be able to hear. You will get a session you can run this week, including exact wording. You will get practice that builds learning, not a pile of the same slogan. Your student will get a short section of their own. If it isn't clicking, you will get three diagnostics and a next move for each, with no shame in the room. Tools, including AI, stay optional and adult-side. And you will get a plain checklist for "done enough," so you can place by skill rather than by birthday.

This week you can learn Ennis's sentence and the session shape well enough to hear a wrong answer. Today the student can try one real judgment without you holding the pencil.

What this book will not do

This book will not hand you 180 days of worksheets. A year of photocopies is not a teaching method, and I will not pretend it is.

It will not sell you a curriculum. Later, a short resources chapter names common programs by fit — parent load, style, how they place a student — so you can choose. A poster of "elements and standards," a commercial test-prep course, and a philosophy-for-children kit are tools some families use. None of them is this book, and this book is not a catalog in disguise. Richard Paul's old warning is worth keeping and leaving there: training on "neutral" cases can make students more slick rather than more fair.⁴ This book is not a franchise of that brand.

It will not promise a score. There is no guaranteed percentile, no guaranteed test number, and no guaranteed college letter inside these pages. A short critical-thinking inventory is one operationalization of one conception. It is not the mean-

⁴Richard W. Paul, "Teaching Critical Thinking in the 'Strong' Sense: A Focus on Self-Deception, World Views, and a Dialectical Mode of Analysis," *Informal Logic Newsletter* 4, no. 2 (1981): 2–7. The 1981 warning that training on "neutral" cases can make students "more sophistic rather than less so." This book keeps the warning. It does not adopt the later classroom franchise.

ing of the work.⁵ What this book can promise is a path: knowledge plus evaluation habits, practiced inside real subjects, at the skill the student actually has, until they can do the next judgment unaided.

It will not open as if Socrates invented the English phrase. He did not say “critical thinking.” Bacon did not. The English educational goal-name is modern; the practices — asking for reasons in public, checking a claim against grounds — are not.⁶ John Dewey already described reflective thought as active, persistent, careful consideration of a belief in light of the grounds that support it and the further conclusions it tends toward.⁷ That is ancestry enough. A 2,500-year brand story is not a method.

And this is not a book that lectures your child about you, or you about your character. The student is a person, not a percentile. When this book speaks to them, it speaks with respect. No baby-talk. No research sermon.

The promise

If you only remember one sentence, remember this: you cannot think about what you do not know, and knowing is not the same as a poster of maxims.

The work is to take this student through real subjects — history, science, mathematics, the news, money, health — and to practice, inside those subjects, the habits of deciding what to believe or do. You will sometimes slow down. You will sometimes skip ahead. Both are teaching. Birthday is not placement.

⁵Peter A. Facione, *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction* (the Delphi report), ERIC ED315423 (Millbrae, CA: California Academic Press, 1990). Expert consensus for assessment and instruction, 46 panelists, six rounds, 11 February 1988–25 September 1989. Not “the APA’s official definition.” A commercial test built from that construct is one operationalization, not the meaning.

⁶David Hitchcock, “Critical Thinking,” *Stanford Encyclopedia of Philosophy*, first published 21 July 2018, substantive revision 12 October 2022. Dewey (1910) used the English phrase as an educational goal-name, more commonly calling the same goal reflective thinking. This book does not claim that anyone coined the phrase. Socrates, Bacon, and Kant did not say “critical thinking.”

⁷John Dewey, *How We Think* (Boston: D. C. Heath, 1910), ch. 1: “Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the further conclusions to which it tends, constitutes reflective thought.” Same book: thinking “is specific, not a machine-like, ready-made apparatus to be turned indifferently and at will upon all subjects, as a lantern may throw its light as it happens upon horses, streets, gardens, trees, or river.”

You can do this. You do not have to know next year's idea today. You have to know this week's idea well enough to sit still while they struggle, then ask one good question. Start here. Read *How to Use This Book This Week*, then the one-page list of five things, then *The Practice Hour*. After that, open the chapter that matches the skill in front of you. You will know more after one chapter than you know this morning. Your child will have something to try today. We can do this.

How to Use This Book This Week

Start at the skill in front of you, not on page one because a catalog, a birthday, or a well-meant relative said so.

This book is a path, not a calendar. The teaching chapters run: what we mean by the phrase; why you need something to think about; why maxims are not enough; how to practice this week inside a real subject; why a test score is not the thing; how to leave the page in civic and media life; stakes in clinic, work, and ordinary decisions; then one chapter on machines that finish the sentence. Records and resources come last. Chapter labels here are sequence anchors. They tell you where an idea usually sits. They do not tell you where *this* student sits.

Pick the chapter by skill, not by birthday

Open the chapter you think is right. Skip to **What “done enough” looks like** at the end of the *previous* chapter, or to **If it isn’t clicking** in the one you opened. If the student can already do those checks unaided, you are too early. If the checks from two chapters back are still failing, drop back. “Pushing through media literacy” with a hole in the subject’s facts is how judgment becomes a fog of slogans.

If they cannot yet say, in their own words, that the work is reasonable reflective thinking about what to believe or do — if “critical thinking” still means “be skeptical” or “I already have it” — stay in Chapter 1, even if they are fourteen. The sentence is the gate. Skipping it because a child is “good at arguing” is the expensive shortcut.

If they can already name the decision, point to the knowledge it needs, and catch an empty maxim, and they are still doing Chapter 1 worksheets because the cover says “introduction,” skip ahead. What a student of a given age can learn is mostly what

they have already had a chance to learn, not a lock on a birthday.⁸ Homeschool tables already stretch. A twelve-year-old may be ready for sourcing in history. A sixteen-year-old may still need the knowledge constraint said plainly. Both can be right. Neither is a law about your kitchen.

A publisher's "grade 8 critical thinking" book is a scope, not a legal grade. A fourteen-year-old can sit in three different "grades" in three programs on the same morning. Use this book's checklists, or a placement check from the program you own, and then teach.

Two children may need two chapters. That is ordinary. The session shape in *The Practice Hour* still holds; the idea on the table changes.

How you use the parent half

Most of each teaching chapter is for you. Read it *before* the lesson, not over the student's shoulder.

Why this matters tells you what this idea unlocks. You will teach with more patience when you can see the later door.

For the parent: understand it yourself gives one everyday picture, one precise picture, and three to five wrong answers you should be able to hear. Sit with those. If "consider both sides" still looks like a complete lesson, that is the maxim idea, and it is worth five quiet minutes now.

How to teach it this week assumes the session shape from *The Practice Hour*. It will not rebuild the hour. It will give you the first problem, the wording for *this* idea, and how to fade your help.

Practice that actually builds learning is what goes on the page after the model: a short blocked set of the new move, then mixed problems so the student has to choose, plus a little retrieval of facts they already know in the subject. Kitchen

⁸National Mathematics Advisory Panel, *Foundations for Success: The Final Report of the National Mathematics Advisory Panel* (Washington, DC: U.S. Department of Education, 2008), Finding 15: what is developmentally appropriate is "largely contingent on prior opportunities to learn"; age-lock claims have "consistently been shown to be wrong" when prerequisite knowledge is present. ERIC ED500486, <https://files.eric.ed.gov/fulltext/ED500486.pdf>. Access date for URLs in these notes: 30 August 2026. The finding is about learning, not only about mathematics. This book uses it the same way: birthday is not placement.

talk, money, and the news can motivate or give a second picture. They do not replace the written judgment, the named knowledge, or the unaided problem.

Tools, including AI is optional, short, and for the adult. The rules live once in *The Practice Hour*.

What “done enough” looks like is how you leave. You are looking for unaided work, not a perfect Tuesday.

You do not have to read the whole chapter tonight. You do have to read the parent half of *this week’s* idea before you sit down with the student.

The five-minute parent warm-up

Make this a habit.

Five minutes. Student not yet in the chair. Phone face down.

1. Read today’s idea until you can say it in one sentence.
2. Work two problems yourself, on paper, without the key until you finish. Use a claim from this week’s actual subject — a history paragraph, a science result, a news headline, a word problem.
3. Glance at the “wrong answers you should be able to hear.” Name the one you would have given at fourteen.
4. Write one sentence you will actually say. Not a speech. Example: “Critical thinking is deciding what to believe or do, with reasons.” Or: “You cannot think carefully about a topic you do not know.”
5. Close the book to the student page. You are ready.

That five minutes is how you stay out of their working memory. If you are learning the idea *while* they are stuck, you will talk too much or take the pencil. Prepare first. Then sit still. You need today’s sentence, not a seminar.

How the student uses “For the student”

Every teaching chapter includes a short section written to the student, not about them. One or two pages. Warmer. Direct. What the idea is, a tiny worked example, two tries, an “explain it back” prompt, and one challenge problem.

Hand them that section when it is time for their attempt — after your short model, not instead of it. They may read it silently or out loud. You stay in the room. You do not add a lecture about high school.

The two tries are theirs. “Explain it back” checks that they have a picture, not only a copied slogan. The challenge is optional that day; it is there so a ready student does not sit in busywork.

The student page is not something to send off with an unsupervised chatbot. The attempt is still theirs. You still hold the key.

When to skip ahead

Skip ahead when this chapter’s “done enough” checklist is already true *unaided*. Lingering past that is not thoroughness. It is boredom wearing a virtue.

Slow down when the same wrong answer repeats after a clear model and a real attempt. **If it isn’t clicking** will give you three likely causes and a next move: a missing picture of the activity, a maxim with nothing to think about, or a subject hole dressed up as a thinking problem. Missing facts are a knowledge problem, not a character problem. Ten minutes of struggle on a well-posed new idea is ordinary work.

Diagnose the sentence, the knowledge constraint, and the empty maxim before you buy a standalone “critical thinking” workbook. Accelerate a student who has the prerequisites. Hold still for a student who does not. Both are the same kindness.

If you need a human tutor, that is a normal high-school plan, not a failure of the kitchen.

For this week: pick the chapter by skill. Do the five-minute warm-up. Run the hour as *The Practice Hour* describes it. Let the student page be theirs. Stop talking sooner than you want to. That is how you use the book.

If You Only Remember Five Things

Keep this page. The chapters will add wording, pictures, and practice. They will not replace these.

1. The work is deciding what to believe or do. Critical thinking is reasonable, reflective thinking aimed at a decision. Reasonable: there are reasons. Reflective: you pause long enough to look at them. Believe *or do*: a life includes sharing, consenting, refusing, voting, hiring — not only marking someone else’s paragraph. Ennis’s sentence is the one this book keeps. A personality compliment, a mission stamp, and a short test score are not the same object.⁹

2. It is not a bicycle-skill. Riding a bicycle, once learned, works whenever a bicycle is present. Thinking is not that sort of skill. The processes of thinking are mixed with the content of thought. Even a three-year-old can think carefully in a domain they know. Even a trained scientist can fail outside theirs. Domain-specific judgment can be taught. A content-free general skill, taught directly, has no proven method.¹⁰

3. Knowledge first. Maxims fail without something to think about. “Consider both sides” is useful advice. Without background knowledge and practice, the student cannot implement it. You can teach the maxim. You cannot supply, by the maxim, the facts the maxim needs. This week’s practice sits inside a real subject — history, science, mathematics, the news on the table — not in a poster hour with

⁹Robert H. Ennis, “A Logical Basis for Measuring Critical Thinking Skills,” *Educational Leadership* 43, no. 2 (1985): 44–48. “Critical thinking is reasonable reflective thinking focused on deciding what to believe or do.” Access date for URLs in these notes: 30 August 2026.

¹⁰Daniel T. Willingham, “Critical Thinking: Why Is It So Hard to Teach?,” *American Educator* 31, no. 2 (Summer 2007): 8–19. Restated in Willingham, 2019: “We are not even sure the general skills exist, but we are quite sure there is no proven way to teach them directly.”

nothing to think about.¹¹

4. Struggle before rescue. Don't grab the pencil. Ask. Wait. Hint. Then model. Ten minutes of honest work on a problem they have the pieces for is the lesson. Taking the pencil feels kind. It teaches them to watch. If they are missing facts, that is not struggle — that is a hole, and you fill the hole. If they have the pieces, sit on your hands.

5. The student attempts first. You hold the key. AI is optional for the adult. The student does the problem. You keep the answers. A tool may explain today's idea *to you*, write extra practice with the key hidden, draft a hint script, or help you diagnose work already done. It may not sit in the chair during the attempt, and it may not paste the worksheet. The full rules are in *The Practice Hour*, once. Later chapters will point back rather than preach.

If this week needs a compass, this is it. Five things. Then sit down and teach.

¹¹Willingham, 2007: "You can teach students maxims about how they ought to think, but without background knowledge and practice, they probably will not be able to implement the advice they memorize." Edward M. Glaser, *An Experiment in the Development of Critical Thinking* (Teachers College, Columbia, 1941), 175, via Hitchcock: the attitude of wanting evidence transfers better than skill in applying methods, which is limited by pertinent knowledge.

The Practice Hour

The hour has a shape. Learn it once. Later chapters will give you today's idea, today's first problem, and today's sentence. They will not rebuild this hour. When a chapter says "run the session," it means this.

You do not need a school bell. You need a beginning, a short model, a real attempt, one good question, mixed practice, and an unaided check. Younger children may finish in twenty-five minutes. Older students may need forty-five. The shape does not change. The idea on the table does. The subject does not change either: this hour lives inside history, science, mathematics, or the news you already teach. It is not a separate "thinking" period with nothing to think about.

Sit down having already done the five-minute parent warm-up from *How to Use This Book This Week*. You know today's sentence. The student has the pencil. You have the key.

1. Warm-up

Three to five minutes. Facts and moves they can already get right in this week's subject. Mixed, not a page of the same item. This is retrieval, not a test of character.

Say:

"We're going to start with things you already know."

"Tell me two facts from yesterday's chapter."

"What does this word mean in this unit? Take your time."

If they already get these right untimed, you may use a brief timer. Say:

"We'll time this only because you already get these right. Ready?"

If they do not already get them right, skip the timer. Time only what they can already produce. Fluency is quick, flexible, accurate recall of the subject's facts — not a race and not optional. You are building retrieval, not an identity.

Keep this short. The warm-up is not the lesson. When you feel yourself teaching a new idea here, stop. That idea belongs in the model.

2. Short model

Five to eight minutes. One fully worked judgment, out loud. Then a second example with holes they fill. Then you stop.

New, dense material — first time naming “believe or do,” first time catching an empty maxim, first time sourcing a document — is where a short model earns its keep. A novice staring at a blank page spends working memory searching, not learning. Show the move. Then fade.¹²

Say:

“Watch this one. I’ll say every step out loud.”

“The work is deciding what to believe or do, with reasons. That’s why I am going to name the decision first.”

“What knowledge would I need that I do not have yet?”

Then, on the second example:

“Your turn to tell me what I write next. I’ll hold the pencil this time. You do the thinking.”

“I’m leaving the last step blank. You finish it.”

Include one incorrect example when the idea is new, and ask:

“What did they think ‘critical thinking’ meant?”

¹²John Sweller and Graham Cooper, “The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra,” *Cognition and Instruction* 2, no. 1 (1985): 59–89: for novices, a worked solution outperforms unguided search. Alexander Renkl and Richard K. Atkinson developed guidance fading (full example, then completion problems, then conventional problems). Once the learner can execute the procedure, extra full examples become redundant. The method travels; the original paper is algebra.

That question is worth more than another correct copy.

When they can already execute the move, stop showing full examples. Extra modeling after they can do it is noise. Fade: full example, completion problem, conventional problem. Then they work.

The model is short on purpose. If you are still talking at minute twelve, you are giving a lecture. Close the model. Hand them the first problem.

3. Student attempt

This is the center of the hour. Eight to twelve minutes. Maybe longer for a document or a multi-step case. The first problem is theirs.

Hand them “For the student” if the chapter has it, or the first problem from **How to teach it this week**. Then you talk less than you want to.

Say:

“This one is yours. I’ll be quiet.”

Then be quiet.

If they stall, use this order: ask, wait, hint, then model. Not the reverse.

Ask:

“Show me what you tried.”

“What do you already know about this?”

“Is this a decision about what to believe, or about what to do?”

Wait. Count a slow ten in your head. The silence is the work. If you fill it, you took the problem back.

Hint, one hint:

“Name the knowledge this would need.”

“Look only at this claim first.”

“You already know a fact from this week’s subject that would help. Which one?”

Then, if they are still stuck after a real try:

“I’m going to show you this one step. Then you take it from here.”

Show the step. Return the pencil.

If they are missing a fact — they cannot place the event, they cannot say what the experiment measured — that is a hole, not useful struggle. Fill the hole with a short retrieval or a picture, then return to the problem. Struggle on a well-posed new idea, with the pieces in place, is the lesson. Struggle on missing prerequisites is just being lost.

Keep your hands off their pencil. Taking it feels kind. It teaches them to watch. If you feel the urge, put your hands in your lap and say the ask-sentence again.

4. One good question

Two or three minutes. Not “did you think critically?” They will say yes. You will learn nothing.

A good question makes them generate the idea again, in words or in a second representation.

Say:

“How did you know?”

“What would you need to see to change your mind?”

“What would happen if this fact were different?”

“Explain it so I could do the next one.”

“What knowledge is this decision sitting on?”

One question. Maybe a follow-up. Then stop. This is not a seminar. You are checking that they have a picture, not collecting a speech.

If they cannot explain it back, they copied a slogan. Return to the subject — the paragraph, the data, the document — and give one more attempt. Copying is what a stuck novice does when the model was too fast or the attempt was too short. Slow the model tomorrow. Keep today’s pencil theirs.

5. Mixed practice

Ten to fifteen minutes. First a short blocked set of the new move — four to six problems that use today’s idea — so they learn to execute. Then mix. Yesterday’s type. Last month’s type. A problem from the subject they already study. The new move sitting next to an old one so they have to choose.

Blocked practice teaches execution. Mixed practice teaches *when* to use the move. A page of twenty identical “consider both sides” prompts feels like mastery at 9 p.m. It often is not still there in two weeks.¹³

Say:

“These first few are today’s move. Get the feel of it.”

Then:

“These are mixed on purpose. First decide what kind of problem it is. Then choose.”

“You already know this type from last week. Find it among the new ones.”

Word problems, news stories, and lab write-ups are types — a claim about a cause, a comparison, a source, a number — not magic words. “Study says” does not mean believe. Teach the structure. Vary the cover story so they match the type, not the noun.

Kitchen, money, and the news can motivate, or give a second picture of a judgment they are already practicing in a subject. They do not replace the subject, the written reasons, or this page. A student who can argue at dinner and cannot source a paragraph has performed, not learned.

Keep practice short enough that attention holds. More pages are not more learning if the last ten are sloppy. Stop while they can still think.

¹³Doug Rohrer, Robert F. Dedrick, and Kaleena Burgess, “The Benefit of Interleaved Mathematics Practice Is Not Limited to Superficially Similar Kinds of Problems,” *Psychonomic Bulletin & Review* 21 (2014): 1323–30. Doug Rohrer, Robert F. Dedrick, Marissa K. Hartwig, and Chi-Ngai Cheung, “A Randomized Controlled Trial of Interleaved Mathematics Practice,” *Journal of Educational Psychology* 112, no. 1 (2020): 40–52. Grade-7 classroom trials. This hour borrows the *practice* of mixing types after a short blocked introduction, not those coefficients as a thinking-hour finding.

6. Unaided exit ticket

Three to five minutes. One or two problems. Book closed. No hints. No “For the student.” No chatbot. No looking at the worked example.

This is the check. Tonight’s helped practice, with you in the chair and a tool in reach, is not this check.¹⁴

Say:

“Two problems. Book closed. I won’t help. That’s the point.”

“Wrong is useful. We’ll look after.”

“You don’t need a perfect score. I need to see what you can do alone.”

Then you are quiet again. When they finish, you look. If both are right, you are done enough for today. If one is wrong, you diagnose — which wrong answer from the parent section is this? — and you note tomorrow’s first move. You do not redo the whole hour. You do not punish. You find the picture that was missing.

An exit ticket that always needs you is not an exit ticket. Fade until it doesn’t.

When to stop talking

After you ask, wait. If you fill the silence, you took the problem back.

During the attempt, your job is almost nothing. Presence, not narration. If you hear yourself explaining while they write, you have moved back into the model. Stop. Let them write.

Instead of “this is the hard part” or “I was never good at this,” say today’s sentence: “Let’s name the decision.” The first sentence fills their working memory with your feeling. The second sentence teaches.

Praise a clear picture, a second method, a catch of their own error — not speed.

“You noticed you didn’t have the fact yet. That’s the move I want.”

¹⁴Nicholas C. Soderstrom and Robert A. Bjork, “Learning Versus Performance: An Integrative Review,” *Perspectives on Psychological Science* 10, no. 2 (2015): 176–99: current performance is “often an unreliable index” of learning. The unaided exit ticket is the check this hour trusts.

Stop talking when the next sentence would be a rescue. Hands in your lap. Count ten. Then one hint, or one step, and the pencil goes back.

If you remember only one move from this page, remember that one.

The AI rules, once

Later chapters will point here in one sentence. They will not reprint this box.

Student first. You hold the key. AI is optional for the adult.

The student attempts the problem before any tool is in the room. You keep the answer key — on paper, in a closed tab, in your head after the warm-up. A chatbot does not sit beside them during the attempt.

Use a tool, if you use one at all, for four jobs: explain today's idea *to you* before the lesson; make extra practice with the answers hidden from the student; write a hint script you can read from; diagnose work they already produced.

Never paste the worksheet, ask “what's the answer,” photograph the page for a key, or leave an unsupervised chatbot in their hands during the attempt. Bastani and colleagues found that an unguarded chatbot raised practice scores and then cut the unaided exam.¹⁵

Facts are not a language-model job. Use a tool to generate a short oral list or a practice deck. The student retrieves the fact from the subject.

You can teach this entire book with no AI. Many families will. The hour still has the same shape.

If a product claims it “does 100 percent of the teaching,” treat that as advertising. You still need to hear a wrong answer. You still hold the key.

¹⁵Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman, “Generative AI without guardrails can harm learning: Evidence from high school mathematics,” *Proceedings of the National Academy of Sciences* 122, no. 26 (2025): e2422633122. In a field experiment, a ChatGPT-like tool raised assisted practice grades 48 percent relative to control, then cut unaided exam grades 17 percent. A hint tutor that withheld full solutions raised practice without that exam drop. Not a homeschool study; the transferable caution is the crutch. Access date for URLs in these notes: 30 August 2026.

A Tuesday, said plainly

Here is an illustration, not a reported family. A parent has spent five minutes with Ennis's sentence and written "reasonable reflective thinking about what to believe or do." The student warms up on two facts from this week's history chapter. The parent models one judgment out loud on a claim from that chapter — names the decision, names the knowledge it needs, leaves the last reason blank. The student finishes the reason, slowly. The parent asks, "What would you need to see to change your mind?" Mixed practice: two claims from the chapter, one older science fact, one comparison. Exit ticket: "Here is a new sentence from the same chapter. Decide what to believe. I won't help." Right or wrong, the hour had a shape. Tomorrow's first move is already visible.

You can run that hour. You do not need a degree. You need today's sentence, a pencil you refuse to take, and the willingness to stop talking.

Later chapters will say: run the session as in *The Practice Hour*. That sentence is this page. The idea will change. The hour will not.

Chapter 1

What we mean



Figure 2: A fallen picture frame and a poster of silhouetted figures on a wooden floor before a dark podium and curtain.

Why this matters

The words are already in the house. A school brochure said them. A relative used them as a compliment. A workbook cover promised them. “We teach critical thinking” can mean almost anything: we assigned a debate, we are not *merely* about facts, we bought a test. Those uses are not identical, and none of them, by itself, tells you what to do on a Tuesday.

This chapter gives you a sentence you can actually use. Robert Ennis wrote it as a working definition in 1985 and kept it: critical thinking is reasonable, reflective thinking focused on deciding what to believe or do.¹⁶ That is the activity. Reasonable: there are reasons, not only a mood. Reflective: you pause long enough to look at those reasons, and at where they lead. Focused on deciding: the work has a point. And the last two words do the life work. A life is not only marking someone else’s paragraph. It is sharing a story, consenting to a treatment, refusing a deal, voting, hiring, putting the phone down. Ennis includes action. This book keeps that inclusion.

The sentence names the work. It does not, by itself, tell you how to teach it. Daniel Willingham’s constraint does, and you need it in the same sitting as the definition, or the definition becomes a poster. People have treated critical thinking as a skill like riding a bicycle — once learned, usable in any situation. Thinking is not that sort of skill. The processes of thinking are mixed with the content of thought.¹⁷ You can teach maxims. Without knowledge and practice in a real subject, the student cannot put the maxims to work. That is not a reason to wait. It is a reason to teach the sentence *inside* this week’s history, science, mathematics, or news, not in a content-free hour labeled “thinking.”

What this idea unlocks is everything later in the book. If “critical thinking” still

¹⁶Robert H. Ennis, “A Logical Basis for Measuring Critical Thinking Skills,” *Educational Leadership* 43, no. 2 (1985): 44–48. Restated in later outlines and in Ennis, “Critical Thinking: A Streamlined Conception.” The public sentence: “Critical thinking is reasonable reflective thinking focused on deciding what to believe or do.” Access date for URLs in these notes: 30 August 2026.

¹⁷Daniel T. Willingham, “Critical Thinking: Why Is It So Hard to Teach?,” *American Educator* 31, no. 2 (Summer 2007): 8–19. Authorized reprint at Reading Rockets. “People who have sought to teach critical thinking have assumed that it is a skill, like riding a bicycle, and that, like other skills, once you learn it, you can apply it in any situation. Research from cognitive science shows that thinking is not that sort of skill. The processes of thinking are intertwined with the content of thought (that is, domain knowledge).”

means a personality, a compliment, or a score, Chapters 2 through 4 have nowhere to sit. If it means Ennis's activity, you can hear a wrong turn: empty skepticism, a slogan with no subject, a maxim with nothing to think about. Hearing that this week is worth the struggle.

Grade labels and ages are sequence anchors, not birthdays. A ten-year-old or a sixteen-year-old may sit here. A student who already argues well at dinner, and still cannot say what the phrase means, may sit here too. What a student of a given age can learn is mostly what they have already had a chance to learn.¹⁸ Move a student who has the sentence. Hold the work for a student who does not.

You do not need to be a philosopher. You do need to understand today's idea well enough to hear "I already think critically" as a compliment, not as evidence, and "consider both sides" as a maxim that still needs something to think about.

For the parent: understand it yourself

Many adults feel rusty. That is ordinary. Five minutes of this section, then the warm-up at the end, is enough for tomorrow.

Everyday picture. A headline at breakfast: a storm is coming, or a product is "clinically proven," or a friend says everyone at school already knows. Two decisions hide in that sentence. What do you believe? And what do you do — cancel the picnic, buy the bottle, share the post? Ennis's sentence is the picture of that kitchen. The thinking is reasonable if you can name a reason. It is reflective if you pause long enough to look at the reason, and at what follows if the reason is weak. It is focused on a decision, not on sounding sharp.

A second kitchen picture, still ordinary. The student wants to skip a chapter because "I already get it." That is a decision about what to do. The reasons might be good. They might be fluency — the chapter felt easy last night, which is not the same as holding it unaided this morning.¹⁹ Ask for the reasons. Wait. That is

¹⁸National Mathematics Advisory Panel, *Foundations for Success* (Washington, DC: U.S. Department of Education, 2008), Finding 15, ERIC ED500486. Age-lock claims have "consistently been shown to be wrong" when prerequisite knowledge is present. Used here as a placement principle, not as a mathematics finding only.

¹⁹Nicholas C. Soderstrom and Robert A. Bjork, "Learning Versus Performance: An Integrative Review," *Perspectives on Psychological Science* 10, no. 2 (2015): 176–99. Current performance is "often an unreliable index" of learning.

already the activity, sitting inside a Tuesday you were going to have anyway.

Precise picture. Write the sentence out: *reasonable, reflective thinking about what to believe or do*. Circle *reasonable*. Reasons can be checked. A feeling can be real and still not be a reason. Circle *reflective*. John Dewey called reflective thought the active, persistent, careful consideration of a belief in light of the grounds that support it and the further conclusions it tends toward.²⁰ Suspended judgment is part of that work, and it is a little painful. That pain is not a design flaw. Circle *believe or do*. Appraisal without action is only half a life. Action without appraisal is a reflex. The activity is both.

Willingham's bicycle is the second half of the precise picture, and it belongs next to Ennis, not later. Riding a bicycle, once learned, works whenever a bicycle is present. Thinking is not that sort of skill. The processes of thinking are mixed with the content of thought — that is, with domain knowledge.²¹ Even a three-year-old can think carefully in a domain they know. Even a trained scientist can fail outside theirs. Edward Glaser, in 1941, already split the same way: the attitude of wanting evidence transfers better than skill in applying methods, which is limited by pertinent knowledge.²² John McPeck's short version: thinking is always about something; "I teach thinking" with no *about what* is empty.²³ Ennis never denied that knowledge is necessary. His sentence is the aim. Willingham is the constraint

²⁰John Dewey, *How We Think* (Boston: D. C. Heath, 1910), ch. 1. Reflective thought: "Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the further conclusions to which it tends." Same book: thinking "is specific, not a machine-like, ready-made apparatus to be turned indifferently and at will upon all subjects, as a lantern may throw its light as it happens upon horses, streets, gardens, trees, or river." Dewey also used the English phrase *critical thinking* as an educational goal-name (Hitchcock, *Stanford Encyclopedia of Philosophy*, 2018/2022). This book does not claim that anyone coined the phrase. Socrates, Bacon, and Kant did not say "critical thinking."

²¹Willingham, 2007; restated in Willingham, NSW Department of Education paper, 2019: domain-specific critical thinking can be taught; generic add-ons have limited success; "We are not even sure the general skills exist, but we are quite sure there is no proven way to teach them directly." Even three-year-olds can think critically in a domain they know; even trained scientists fail outside theirs.

²²Edward M. Glaser, *An Experiment in the Development of Critical Thinking*, Teachers College Contributions to Education no. 843 (New York: Teachers College, Columbia University, 1941), 175, as quoted by David Hitchcock. Disposition of wanting evidence transfers better than skill in applying logical methods, which is "limited by" pertinent knowledge.

²³John E. McPeck, *Critical Thinking and Education* (New York: St. Martin's Press, 1981), 8: thinking is always about X; "I teach thinking" simpliciter is empty. This book keeps the *about what?* demand. It does not adopt a ban on any general thread.

on how you teach toward it.

A third precise mark, so the sentence does not get mistaken for a test. Peter Facione's 1990 Delphi report is expert consensus *for purposes of educational assessment and instruction* — forty-six panelists, six rounds. It is not a discovery of a mental organ, and it is not "the APA's official definition."²⁴ A commercial inventory built from that construct measures one operationalization. A score is not the activity. Chapter 5 of this book will spend that point. This week you only need to hear it: when someone says "we tested critical thinking," ask which instrument, on which tasks, and whether the student can still decide unaided when the prompt changes.

Bloom's taxonomy is a vocabulary schools borrowed. Analysis, synthesis, evaluation became the rungs a brochure could point at. Bloom is not a theory of this activity.²⁵ A debate can be practice of reasons in public. A debate can also be combat with a rubric attached. The mission sentence does not distinguish. Distinguishing is your job at the table, and Ennis's sentence is how you do it.

Wrong answers you should be able to hear

1. *"Critical thinking means being skeptical of everything."* Skepticism without reasons is a mood. Ennis's activity is reasonable: it can accept a claim when the grounds are good, and it can act. Empty doubt is not reflection. If the student uses "I'm just being a critical thinker" as a way to refuse every source, they have the compliment, not the sentence.
2. *"I already think critically. It's how I am."* That is a personality claim. Ennis's sentence is an activity you practice, in a subject, this week. Glaser's split helps you hear it: wanting evidence is closer to a transferable attitude; applying methods is limited by knowledge. A student who is curious and still cannot source a paragraph has a disposition without the skill. Praise the curiosity. Teach the skill inside the paragraph.

²⁴Peter A. Facione, *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*, ERIC ED315423 (1990). Forty-six panelists; six rounds, 11 February 1988–25 September 1989. About 61 percent included dispositions in the meaning versus about 30 percent strict proceduralism; 52 percent rejected building an ethical component into the meaning. Not "the APA's official definition."

²⁵Benjamin S. Bloom et al., *Taxonomy of Educational Objectives*, Handbook I: Cognitive Domain (New York: Longmans, Green, 1956), 38, as quoted by Hitchcock: labeled "critical thinking" by some, "reflective thinking" by Dewey, "problem solving" by others; the taxonomy used "intellectual abilities and skills." A vocabulary, not a theory of Ennis's activity.

3. “*We did the critical thinking test, so we’re done.*” A score is one operationalization of one conception. It is not the meaning. Liu, Bridgeman, and Adler showed that conclusions about college learning can flip depending on whether students are induced to try.²⁶ Tonight’s helped practice is not the unaided check. If the house treats a practice score as the skill, you will move on too soon.
4. “*Once you learn it, you can use it anywhere — like riding a bike.*” That is the assumption Willingham answered. Thinking is not that sort of skill. A student who can evaluate a science experiment and then treats a history document as if the same moves were automatic has assumed the bicycle. Name the subject. Name the knowledge. Then practice the decision in *this* domain.
5. “*We are the critical thinkers. They are not.*” That is a political compliment, and its inverse insult. Richard Paul called the skilled defense of one’s own side the weak sense of the work, and warned that training on “neutral” cases can make students more slick rather than more fair.²⁷ Ennis’s sentence is for deciding what to believe or do. Ranking people is a different speech act. If you hear the compliment, return to the claim on the table.

A sixth you will also hear: “It’s just higher-order thinking.” Bloom’s top rungs, used as a synonym. Higher-order is a compliment. It is not Ennis’s activity, and it is not a method. Ask what the student is deciding, and on what knowledge.

Five-minute parent warm-up

Do this before the lesson, on a scrap of paper, no student in the room.

Minute 1. Write Ennis’s sentence. Say it out loud in ordinary words: “The work is deciding what to believe or do, with reasons, after a pause.”

Minute 2. Take one claim from this week’s actual subject — a history sentence, a science result, a headline. Write two columns: *believe?* and *do?* Put one reason under each.

²⁶Ou Lydia Liu, Brent Bridgeman, and Rachel M. Adler, “Measuring Learning Outcomes in Higher Education: Motivation Matters,” *Educational Researcher* 41, no. 9 (2012): 352–62. Conclusions about college learning can flip depending on whether students are induced to try.

²⁷Richard W. Paul, “Teaching Critical Thinking in the ‘Strong’ Sense,” *Informal Logic Newsletter* 4, no. 2 (1981): 2–7. Training on “neutral” cases can make students “more sophistic rather than less so.” This book keeps the 1981 warning. It is not a classroom franchise of later materials.

Minute 3. True or false, said aloud: “Skepticism is the same as this work.” “A test score is the same as this work.” “Once you have it, you have it everywhere.” You are practicing the hearing, not racing.

Minute 4. Look at a student sentence such as “I already think critically.” Decide, without answering them yet, which wrong answer that is. You are not taking a test. You are installing the error so you can hear it tomorrow.

Minute 5. Write one sentence you will actually say: “Critical thinking is deciding what to believe or do, with reasons.” Put the pencil down. That sentence is the lesson.

If you can do those five minutes, you are ready to sit down. The student generates. You hear.

How to teach it this week

A good thinking hour this week has a shape. *The Practice Hour* is the full version, including the AI rules. Here is the shape scaled to this idea. Run it inside a subject you already teach. The subject supplies something to think about. This chapter supplies the sentence.

Warm-up (3–5 minutes, unaided). Two facts the student already knows from this week’s chapter. One ordinary decision already made in the house — “we believed the forecast enough to bring coats” — named in one line. Paper. No device.

Short model (5 minutes). One new idea, one picture, one written sentence. Ennis on the page. You talk for a few minutes. Then you stop.

Student attempt (10–15 minutes). Three to five items. The student writes. You wait. Struggle before rescue: ask, wait, hint, then model. Leave the pencil in the student’s hand.

One good question. Not “did you think critically?” A question that names the idea: “What are you deciding — believe, or do?” “What reason do you have?” “What would you need to see to change your mind?”

Mixed practice (5–10 minutes). Yesterday’s subject fact next to today’s decision. A claim from science next to a claim from history. If the book is almost entirely blocked — most “critical thinking” workbooks are — you pull one older item onto the page.

Exit ticket (5 minutes, unaided). Two items isomorphic to what you taught. Phone in another room. You grade against a key you hold.

That shape is a practice you impose on whatever book is on the table. It is not a 180-day plan. Philip Abrami and colleagues found that improvement “cannot be a matter of implicit expectation.”²⁸ Hoping the activity leaks from good teaching is the weakest design they measured. Making the objective explicit, this week, is the first move.

Exact wording you can say

On the sentence:

“Critical thinking is reasonable, reflective thinking about what to believe or do.”

“Reasonable means there are reasons we can check.”

“Reflective means we pause long enough to look at the reasons, and at what follows.”

On the two decisions:

“Is this a decision about what to believe, or about what to do — or both?”

“If this claim is true, what would we *do*?”

On the bicycle, said once, plainly:

“This is not like riding a bike. You cannot take the same move, unchanged, into a subject you do not know. We practice it here, in this chapter.”

On a stuck student:

“Show me what you tried.”

“What do you already know from this week’s subject that would help?”

“Name one thing you would need to know and do not know yet.”

When you are about to take over:

²⁸Philip C. Abrami et al., “Instructional Interventions Affecting Critical Thinking Skills and Dispositions: A Stage 1 Meta-Analysis,” *Review of Educational Research* 78, no. 4 (2008): 1102–34. Improvement “cannot be a matter of implicit expectation.” 117 studies, 20,698 participants, 161 effects; mean $g^+ = 0.341$. Immersion — critical thinking as a hoped-for by-product — was the weakest cell.

“Your pencil. I’ll ask a question.”

Then wait. Ten seconds of silence is teaching.

First problem for the student

Write this on a small card or the top of the page, exactly, using a claim from *this week’s* subject. Here is an illustration, not a reported family. If this week is a history chapter on a treaty:

The book says the treaty “ended the dispute.” Decide: what would you need to believe that? If you believed it, what would you do next — close the book, or look for one more source?

Say: “The work is deciding what to believe or do, with reasons. What goes in each box?”

Wait. If the student writes “be skeptical,” they heard a mood. If they write a reason from the chapter plus a next step, they heard the sentence. If they stare, show a fully worked different item — a weather claim, said as “I believe the forecast enough to bring a coat, because the map showed rain, and bringing a coat is the *do*” — then return to the treaty. That is fading: full example, then a completion, then the item.²⁹

Later the same week, the diagnostic item:

Someone says, “I already think critically, so I don’t need this chapter.”

Hear “personality.” Hear “compliment.” Hear “bicycle.” After the student has attempted, you may show a worked *incorrect* example that wrote “be skeptical of everything” and ask what that person thought the phrase meant. Generation first, then the named error. The student names the error. You do not narrate it as a verdict on talent.

How to fade help

Day of a new move: you model one fully. The student explains a second. A third is missing only the last written step. A fourth is conventional. When the student

²⁹John Sweller and Graham A. Cooper, “The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra,” *Cognition and Instruction* 2, no. 1 (1985): 59–89. Alexander Renkl and Robert K. Atkinson, “Structuring the Transition from Example Study to Problem Solving in Cognitive Skill Acquisition,” *Educational Psychologist* 38 (2003): 15–22. Method scaled to a judgment item.

is accurate, stop showing full examples. Bring the subject facts back when a new unit arrives. Fading is for accuracy, not for a calendar date.

When to stop talking

After the question. After the hint. After the model of a *different* item. If you hear yourself explaining the student's current problem in real time while their pencil is idle, you have started doing the work. Stop. Point to the sentence. Point to the two columns, *believe* and *do*. Wait.

True and false sentences about the phrase are not a game at the end of the page. "Skepticism is the same as this work" is a sentence about meaning. A student who says it is true has the mood. A student who says it is false because reasons can support a claim has the activity. Your job is to hear which.

A short pass on what the words mean, then a handful of claims from this week's subject, then a return to the sentence, then more claims including a true/false and a "what would you do." Withhold neither the idea nor the unaided problem for weeks. Understanding and practice grow together.

Practice that actually builds learning

Understanding and practice grow together. Wait weeks to practice a new move and the idea stays a speech. Drill a new slogan with no subject and no "why" and you get a chant.

Blocked, for a new move. The day you introduce Ennis's sentence, give a short set that is only that: three claims from this week's chapter. For each: what to believe, what to do, one reason. The day you introduce the bicycle constraint, a short set of "same move, new subject" items — a science claim next to a history claim — so they feel the knowledge gap, not as a trick, as the point.

Mixed, for when to use it. Later the same week, the new move reappears between unlike problems: a retrieval of yesterday's fact, a "believe or do" on a new sentence, a comparison. Mixing types is the practice of choosing. A purchased workbook that is almost entirely blocked can still be interleaved by you pulling yesterday's and last month's types onto the same page.

Brief retrieval of facts already known. Two or three minutes, mixed, after the student can produce those facts correctly untimed. Cover the answers. You listen.

The facts are the subject's facts. Critical thinking has nothing to operate on without them. Overlearning a definition sheet until tonight's score is 100 percent has limited additional benefit compared with meeting the sentence again next week, on a new claim.³⁰

One incorrect example to diagnose. After the student has attempted a claim, show a worked solution that wrote only "be skeptical." Ask: "What did this person think critical thinking meant?" After a personality claim, show "I already have it" and ask which part of Ennis's sentence is missing. The student names the error. You do not announce a talent verdict.

Kitchen, money, and the news. Talk about decisions in ordinary life. A price, a forecast, a share button. That talk can motivate, or give a second picture of a judgment they are already practicing in a subject. It does not replace the subject. A student who can argue about a commercial and cannot use a reason from this week's science chapter has performed, not learned. Kitchen talk is a second picture. The unaided problem in the subject is the first.

A weekly shape. Day of a new idea, a short blocked set. Same week: mixed retrieval of older facts from the subject. Later this week and next: the new move between unlike problems. Monthly: an unannounced mixed check without notes. You impose that shape on the book you already have.

Improvement cannot be a hoped-for by-product. Abrami's 2008 paper is the source of that sentence.³¹ Stamping "think critically" on the syllabus and hoping is immersion-by-committee. This week you name the activity, you practice it on real claims, and you check unaided. That is already more than a poster.

For the student

You are learning what the words actually mean, so you can use them on purpose.

³⁰Doug Rohrer and Kelli Taylor, "The Effects of Overlearning and Distributed Practice on the Retention of Mathematics Knowledge," *Applied Cognitive Psychology* 20 (2006): 1209–24. Comparison used here for the definition card, not as a thinking-hour experiment.

³¹Philip C. Abrami et al., "Instructional Interventions Affecting Critical Thinking Skills and Dispositions: A Stage 1 Meta-Analysis," *Review of Educational Research* 78, no. 4 (2008): 1102–34. Improvement "cannot be a matter of implicit expectation." 117 studies, 20,698 participants, 161 effects; mean $g^+ = 0.341$. Immersion — critical thinking as a hoped-for by-product — was the weakest cell.

Critical thinking is reasonable, reflective thinking about what to **believe** or **do**.

Reasonable means you have reasons, not only a feeling. Reflective means you pause long enough to look at those reasons, and at what follows if they are weak. Believe *or do* means a decision. Sometimes the decision is “is this true?” Sometimes it is “should I share this, buy this, skip this chapter, ask a second question?” Often it is both.

It is not a personality you either have or lack. It is work you do, this week, on something you actually know.

It is not the same as being skeptical of everything. If the reasons are good, you can accept a claim. If the reasons are weak, you can wait, or look again, or act cautiously.

It is not like riding a bicycle. A bicycle, once learned, works on any road that has a bicycle on it. Thinking is mixed with what you know. You can think carefully about a topic you know. You need knowledge before the same care will work on a topic you do not know. That is not an insult. It is the reason we practice inside real subjects, not on empty slogans.

Tiny worked example

A forecast says rain this afternoon.

Believe? I believe rain is likely, because the map shows a front, and the last three forecasts this week were right.

Do? I bring a coat.

The thinking is reasonable (there are reasons). It is reflective (I looked at the map, not only at a feeling). It decides what to believe *and* what to do.

Two tries

Use a claim from *your* chapter this week. If you need a stand-in, use this one:

1. A label says a drink is “clinically proven to boost focus.” What would you need to believe that? If you believed it, what would you do?
2. Your history book says an event “changed everything.” What would you need to believe that? If you believed it, what would you do next in this chapter?

Write both columns: *believe?* and *do?* Put one reason under each. If you do not have a reason yet, write the knowledge you would need and do not have. That count as honest work.

Explain it back

Tell someone at the table, in your own words, what critical thinking is. Then point to the rain example and say why it is this work, not just a mood. Then say whether “be skeptical of everything” is the same thing, and why or why not.

Challenge

Someone says, “I already think critically. That’s just how I am. I don’t need practice.”

What would that person be treating the words as — an activity, a personality, or a compliment? What would Ennis’s sentence ask them to do instead, this week, on a real page?

You are allowed to struggle. You may use the chapter in front of you. You write. If you get stuck, ask for a hint — not the answer. Then try again.

A decision with no knowledge under it is a guess. A slogan with no decision in it is a poster. You can catch both. Name the decision. Name the reason. Then write.

If it isn’t clicking

Three diagnostics. Each one has a next move. None of them is a verdict on talent, and none of them is a reason to wait for a birthday.

1. Every answer is “be skeptical” or “question everything.”

The mood has replaced the activity. Next move: smaller claims, from a subject they actually know. A fact they already retrieved this week. Ask only: what would you need to believe this, and what would you do? Require a reason that uses a fact from the page. If they cannot produce a reason, the hole is knowledge, and you fill the hole first. Stay here until “accept when the grounds are good” is as available as “doubt.” A human tutor who will sit with a real chapter, not a skepticism poster, is a reasonable next step if the mood is still the only move after several weeks of short daily work.

2. “I already have it” or “I’m just not a critical thinker.”

The phrase has become a personality. Next move: take the trait-word off the table for a week. Use Ennis's sentence as a *task*: decide, with reasons, on this page. Three claims a day, unaided exit ticket, no identity talk. Slow down the "new" workbook until the activity is hearable. A student who is fast at arguing and still writes "I already have it" is not ready to skip this. Go ahead once they can name *believe* and *do* on a new claim without prompting, and can catch the personality sentence as a miss. A tutor is optional here; the formats are cheap and you can hear them. Bring someone in if the identity claim is still the default after a couple of weeks of daily task-work *and* the hour has become a fight.

3. They can say the definition and cannot use it on this week's chapter.

The sentence is a recitation. Next move: close the definition card. Open the subject. One claim from today's page. No slogan. If they freeze, the model was too fast or the subject facts are missing. Retrieve two facts, then return to the claim. Copying a definition is what a stuck novice does. Slow the model tomorrow. Keep today's pencil theirs. Go ahead once the unaided exit ticket on a *new* sentence from the same chapter is reliable. A tutor is useful if recitation is still the only success after the subject-first work is consistent.

When to slow down. The mood is the only move. The personality claim is the default. Recitation without a decision on a real page. Those are brakes. "Not ready" because of age is the brake this book will not use. "Not ready" because the picture is missing is a real brake.

When to go ahead. They can say the sentence in their own words. They can mark *believe* and *do* on a new claim from this week's subject. They can hear "be skeptical of everything," "I already have it," and "like a bicycle" as misses. Then Chapter 2's knowledge constraint has somewhere to sit.

When to get a human tutor. You have run the small-claim work, or the task-not-trait work, or the subject-first work, for a stretch of daily sittings, and the same diagnosis is still the one in the room, and the hour has become a fight. A tutor is a release valve, not a failure of the sitting. Keep yourself as the person who can still hear "be skeptical of everything." Outsourcing the hearing is the thing to avoid, not asking for help.

Hear a thin strand without ranking the student. A child who can recite Ennis and cannot decide on a page is missing practice. A child who can argue at dinner and cannot name a reason from the chapter is missing knowledge. Hearing which is thin is enough.

Tools, including AI

Optional helpers for you, the adult. The full rules live in the practice-hour box in the front of this book. This chapter is a pointer, not a second policy paper.

The short version: the student attempts first. You hold the answer key. A tool may explain today's idea *to you*, make extra practice with answers on a separate page you keep, write a hint script you read aloud, or help you diagnose work the student has already done. The student writes on paper.

A language model will happily invent a “critical thinking worksheet” with no subject under it. That worksheet is the thing this chapter is trying to retire. Generate claims *from this week's actual chapter* if you like, with the answers hidden. Then your ears do the work. Bastani and colleagues found that an unguarded ChatGPT-like tutor raised assisted practice scores and then cut the unaided exam; a hint-only tutor did not hurt the exam.³² Soderstrom and Bjork had already named the confusion: performance during practice is often an unreliable index of learning.³³ “Finish your thinking with a chatbot so we can go to soccer” is that pattern with bigger words.

Paste the worksheet, “what's the answer,” photo-to-key, and an unsupervised chatbot during the attempt stay out of the hour. You using your own adult account to prepare, with the student not in the loop until the page is on paper, is the lane the practice-hour box describes.

A poster of maxims is a tool too. Put Ennis's sentence on the table this week. Take the rest of the poster down until Chapter 3. A wall of slogans trains recitation.

What “done enough” looks like

Placement is by skill, not birthday. A “grade 8 critical thinking” book is a publisher's scope, not a legal grade. Programs cut the grain differently. Placement

³²Hamsa Bastani et al., “Generative AI without Guardrails Can Harm Learning: Evidence from High School Mathematics,” *Proceedings of the National Academy of Sciences* 122, no. 26 (2025): e2422633122. Unguarded tutor raised assisted practice and cut the unaided exam; hint-only tutor did not hurt the exam. Not a homeschool study. The pattern is the sitting condition.

³³Nicholas C. Soderstrom and Robert A. Bjork, “Learning Versus Performance: An Integrative Review,” *Perspectives on Psychological Science* 10, no. 2 (2015): 176–99. Current performance is “often an unreliable index” of learning.

checks exist because the labels lie. Choose by fit. This book names programs as options later. It does not rank them.

Checklist before moving on

- The student can say, in their own words, that the work is reasonable, reflective thinking about what to believe or do.
- On a new claim from this week's actual subject, they can mark *believe?* and *do?* and give one reason that uses a fact from the page — or name the knowledge they would need and do not have.
- “Be skeptical of everything” is hearable as a miss. They can accept a claim when the grounds are good.
- “I already think critically” is hearable as a personality claim, not as evidence. The activity is a task this week, not a trait.
- “Like riding a bicycle” is hearable as the wrong picture. They can say, plainly, that the move needs knowledge in *this* subject.
- “We are the critical thinkers” is hearable as a compliment, not as the activity.
- A practice score, a debate, and a mission stamp are not treated as the meaning.
- The unaided exit ticket on a new sentence works more days than it fails. You can hear a wrong answer and ask a good question without taking the pencil.

If most of that list is true, go on to Chapter 2 — you need something to think about — even if the birthday says otherwise. If the birthday says “freshman” and the mood is still the only move, stay. The next chapter is the knowledge gate. This chapter's job was to make the sentence hearable.

A path through judgment is the promise. A percentile is not.

Chapter 2

You need something to think about



Figure 3: A lantern, a stack of bound books, a tied folio, and a spiral notebook on a dark desk.

Why this matters

Chapter 1 gave you a sentence: reasonable, reflective thinking about what to believe or do. This chapter gives you the constraint that makes the sentence expensive, and useful. You cannot think carefully about a topic you do not know. Knowledge is not the thing critical thinking is “higher than.” It is the medium of the thought.

Daniel Willingham’s 2007 essay is the load-bearing page. The processes of thinking are mixed with the content of thought. You can teach maxims about how a person ought to think. Without background knowledge and practice, they probably will not be able to implement the advice they memorize.³⁴ Even a three-year-old can think carefully in a domain they know. Even a trained scientist can fail outside theirs. Children are not as limited as a birthday suggests. Adults are not as portable as a compliment suggests. Stage, here, is domain knowledge, not a candle on a cake.

What this idea unlocks is every later Tuesday. If you wait to “do the facts” until after a thinking unit, the thinking unit has nothing to operate on. If you wait to think until every fact is in, you never start. The move is both, in the same subject, the same week: retrieve what they know, then decide, with reasons, on a claim that needs those facts. Chapter 3 will spend why maxims fail. Chapter 4 will spend how to practice this week. This chapter installs the gate. Skip it because a student is “good at arguing,” and later pages arrive as fluent guesses.

You do not need to be a cognitive scientist. You do need to understand today’s idea well enough to hear “we don’t want mere facts” as a skip of the medium, not as a high-minded aim, and “they’ll pick it up if we discuss” as hope, not as a method.

This week you can learn to hear a fluent guess with nothing under it. Today the student can retrieve two facts from the chapter already on the table, then decide, with reasons, on one new sentence that needs those facts. That pairing is the whole lesson. It is also the kindness: you are not asking them to be a general thinker. You are asking them to use what they know.

³⁴Daniel T. Willingham, “Critical Thinking: Why Is It So Hard to Teach?,” *American Educator* 31, no. 2 (Summer 2007): 8–19. “You can teach students maxims about how they ought to think, but without background knowledge and practice, they probably will not be able to implement the advice they memorize.” Even three-year-olds can think critically in a domain they know; even trained scientists fail outside theirs. Access date for URLs in these notes: 30 August 2026.

Age is not the gate. A nine-year-old who knows this week's science chapter can decide on a fair test. A sixteen-year-old who has not yet placed the parties in a treaty cannot decide what the treaty did, no matter how sharp the dinner-table argument sounds. Willingham's three-year-olds and trained scientists are the same fact in two directions. Knowledge, in the domain in front of you, is what makes the activity possible.

For the parent: understand it yourself

Many adults feel rusty. That is ordinary. Five minutes of this section, then the warm-up at the end, is enough for tomorrow.

Everyday picture. A chessboard. If you do not know what a knight does, you cannot judge a knight move. Curiosity will not supply the rule. A maxim — “consider both sides” — will not supply it either. The sides are made of legal moves, and legal moves are knowledge. The same picture at a kitchen table: a sports argument about a foul, a science claim about a vaccine schedule, a history sentence about a treaty. If the student does not know the rules, the schedule, or the parties, the argument is theater. They may sound sharp. They are not deciding with reasons that can be checked.

A second everyday picture, still ordinary. You ask a young child whether a toy “makes the light go.” In a domain they have played in, they can test, wait, and change their mind. Willingham's point, said kindly: they can think critically there. Take the same child to an unfamiliar machine and the same care has nothing to hold. The difference is not maturity as a personality. The difference is knowledge.

Precise picture. Working memory is small. It holds the pieces you are comparing right now. Background knowledge is what lets those pieces *be* pieces instead of noise. When a historian sources a document, the move is not a free-floating skill. It is knowing what a source of that kind usually is, what would count as corroboration, what the period's language did. When a scientist controls variables, the move is knowing which variables exist in this setup. Willingham's 2019 restatement is a four-step program, not a slogan: say what critical thinking *means in each subject*; name the content those tasks need; sequence knowledge and skills; plan years of revisiting, because even well-learned content fades.³⁵ “Think like a historian” is

³⁵Daniel T. Willingham, paper for the NSW Department of Education, 2019. Four-step plan: identify what critical thinking means in each domain; identify the content those tasks need; se-

not a goal until sourcing, corroboration, and contextualization are named. Those moves are not the ones a scientist needs.

Surface structure captures attention. Deep structure is what would transfer. Novices do not see the analog. Willingham reports a band-marching word problem that is the same mathematics as a garden problem solved minutes earlier: 19 percent saw the analog; a hint that “this is like the garden problem” raised solution only to 35 percent. Most still could not implement.³⁶ Mary Gick and Keith Holyoak’s older demonstration is the same noticing gap: Duncker’s radiation problem is solved spontaneously by about 10 percent; after an analogous fortress story, about 30 percent without a hint; with a hint to use the story, about 75 percent.³⁷ The gap is not application skill. It is noticing. Chapter 3 will spend that as the reason maxims fail. This chapter needs the prior fact: noticing deep structure depends on having the knowledge that makes the structure visible.

Experts and novices sort the world differently. Micheline Chi, Paul Feltovich, and Robert Glaser had physicists sort problems. Experts grouped by the principle needed — conservation of energy, Newton’s second law. Novices grouped by surface — inclines, pulleys, springs.³⁸ Representation, not raw brainpower, was the difference. Chess is the famous cousin. William Chase and Herbert Simon showed that masters reconstruct game positions after a five-second glance far better than novices — and lose the advantage on random, illegal boards. The skill

quence; plan multi-year revisiting. “Think like a historian” is not a goal until sourcing, corroboration, and contextualization are named. Punchline: “We are not even sure the general skills exist, but we are quite sure there is no proven way to teach them directly.” Forgetting of well-learned content: Harry P. Bahrick, “Semantic Memory Content in Permastore,” *Journal of Experimental Psychology: General* 113 (1984): 1–29; Bahrick and Lynda K. Hall, “Lifetime Maintenance of High School Mathematics Content,” *Journal of Experimental Psychology: General* 120 (1991): 20–33; Andrew Pawl et al. on introductory physics, about half forgotten in three years, as cited by Willingham 2019.

³⁶Willingham, 2007: band-marching word problem isomorphic to a garden/least-common-multiple problem solved minutes earlier; 19 percent saw the analog; a hint raised solution only to 35 percent.

³⁷Mary L. Gick and Keith J. Holyoak, “Analogical Problem Solving,” *Cognitive Psychology* 12 (1980): 306–55, and Gick and Holyoak, “Schema Induction and Analogical Transfer,” *Cognitive Psychology* 15 (1983): 1–38. Duncker radiation: about 10 percent spontaneous; analog without hint about 30 percent; analog plus hint about 75 percent. The gap is noticing, not application skill.

³⁸Micheline T. H. Chi, Paul J. Feltovich, and Robert Glaser, “Categorization and Representation of Physics Problems by Experts and Novices,” *Cognitive Science* 5 (1981): 121–52. Experts grouped by principle; novices by literal surface features.

is chunked, domain-specific pattern knowledge, not a general visual memory.³⁹ That is why chess as a “thinking gym” is the wrong metaphor. The chunks do not exist in algebra.

Expertise is encapsulated. Neurologists do not diagnose cardiac cases well. Technical writers of appliance pamphlets do not thereby write newspaper articles. Professional philosophers remain susceptible to order and wording effects in everyday moral judgment.⁴⁰ The knowledge that makes a person fast and accurate in a practiced environment does not travel, unchanged, to an unfamiliar one. Edward Glaser already wrote the educational version in 1941: skill in applying methods is limited by pertinent knowledge.⁴¹ John McPeck’s conceptual version: thinking is always about something; the something is never “everything in general.”⁴² Dewey’s lantern sentence, from 1910, said thinking “is specific, not a machine-like, ready-made apparatus to be turned indifferently and at will upon all subjects.”⁴³ Four denials of content-freeness, a century apart. This week’s teaching is the practical one: put knowledge on the table first, then the decision.

A National Research Council science-education committee, as Willingham quotes it, refused both skips: teaching content alone is not likely to lead to proficiency in science, nor is engaging in inquiry experiences devoid of meaningful science content.⁴⁴ Facts without a decision are a list. A decision without facts is a pose. You need both, in the same subject.

³⁹William G. Chase and Herbert A. Simon, “Perception in Chess,” *Cognitive Psychology* 4 (1973): 55–81. Masters reconstruct game positions after a brief glance and lose the advantage on random, illegal boards.

⁴⁰Encapsulated expertise as cited in Willingham, 2019: Remy M. J. P. Rikers, Henk G. Schmidt, and Henny P. A. Boshuizen, neurologists and cardiac cases (2002); Ronald T. Kellogg on technical writers and newspaper articles (2018); Eric Schwitzgebel and Fiery Cushman, professional philosophers and irrelevant problem features (2015).

⁴¹Edward M. Glaser, *An Experiment in the Development of Critical Thinking* (Teachers College, Columbia, 1941), 175, via Hitchcock: skill in applying methods is limited by pertinent knowledge; the attitude of wanting evidence transfers better.

⁴²John E. McPeck, *Critical Thinking and Education* (New York: St. Martin’s Press, 1981). Thinking is always about X; X is never “everything in general.”

⁴³John Dewey, *How We Think* (Boston: D. C. Heath, 1910). Thinking “is specific, not a machine-like, ready-made apparatus to be turned indifferently and at will upon all subjects, as a lantern may throw its light as it happens upon horses, streets, gardens, trees, or river.”

⁴⁴National Research Council science-education committee, as quoted by Willingham, 2007: “Teaching content alone is not likely to lead to proficiency in science, nor is engaging in inquiry experiences devoid of meaningful science content.”

Willingham is explicit that choosing content expresses values, and that not choosing is still a choice. Everyday life is not a school subject and not “everything in general.” For a household, the subjects are the ones you already teach, plus the domains a life actually contains later — money, health, news, work, friendship, citizenship. This chapter does not ask you to build all of those this week. It asks you to stop treating knowledge as optional warm-up. Pick the chapter on the table. Load the pieces. Then decide.

McPeck’s *about what?* is a constraint you can say in one breath. Thinking is always about something. “I teach thinking” with no object is empty. This book keeps that demand. It does not need his ban on any general thread. Ennis’s sentence can still be said once, as aim. The practice still happens inside X. X, this week, is the chapter.

Wrong answers you should be able to hear

1. “*We don’t want mere facts. We want them to think.*” Facts are not the enemy of the activity. They are what the activity runs on. Ennis’s sentence still holds: the work is deciding what to believe or do. You cannot decide, with reasons, about a treaty you cannot place. Hear this as a skip of the medium. Next move: retrieve two facts, then ask the decision.
2. “*If we just discuss, the thinking will come.*” Discussion without named knowledge is talk. Abrami and colleagues found that improvement cannot be a matter of implicit expectation; immersion — hoping the activity leaks from rigorous content without naming the moves — was the weakest cell.⁴⁵ P. Karen Murphy and colleagues found that more student talk is not automatically more reasoning.⁴⁶ Hear this as hope. Next move: name the fact the discussion needs, then name the decision.
3. “*Once they know the chapter, the thinking will take care of itself.*” That is the other skip: knowledge as a magic leak. Content-specific practice of a named move — control of variables, sourcing a document — can be taught. It does not appear on its own because the facts were covered. Hear this as

⁴⁵Philip C. Abrami et al., “Instructional Interventions Affecting Critical Thinking Skills and Dispositions: A Stage 1 Meta-Analysis,” *Review of Educational Research* 78, no. 4 (2008): 1102–34. Improvement “cannot be a matter of implicit expectation.” Immersion was the weakest cell.

⁴⁶P. Karen Murphy et al., “Examining the Effects of Classroom Discussion on Students’ Comprehension of Text: A Meta-Analysis,” *Journal of Educational Psychology* 101, no. 3 (2009): 740–64. Several approaches increased student talk and improved comprehension; few increased literal/inferential comprehension *and* critical thinking/reasoning.

immersion in a second costume. Next move: after retrieval, one explicit judgment on a new claim from the same chapter.

4. *“I’m smart, so I can think about anything.”* Cognitive ability helps. It does not supply the knight’s move. Myside bias is often independent of cognitive ability; “be smarter” is not a cure for defending your side.⁴⁷ Hear this as the bicycle in a new hat. Next move: pick a topic they do *not* know and let them feel the stall, kindly, then return to a topic they do.
5. *“We’ll do knowledge later. This week is thinking.”* There is no later that makes a maxim implementable without content. Willingham: without background knowledge and practice, they probably will not be able to implement the advice they memorize.⁴⁸ Hear this as a sequence error. Next move: same week, same subject, facts then decision.

A sixth you will also hear: “This is just memorization.” Retrieval of known facts is not the enemy of judgment. It is how the pieces get into working memory long enough to compare. Time only facts they can already get right untimed. Then use them.

Five-minute parent warm-up

Do this before the lesson, on a scrap of paper, no student in the room.

Minute 1. Write: “You cannot think about what you do not know.” Say it out loud. Then write Ennis’s sentence next to it, so the aim and the constraint sit together.

Minute 2. Open this week’s actual chapter. Write two facts the student should already have. Write one claim from the same chapter that needs those facts.

Minute 3. True or false, said aloud: “Facts get in the way of thinking.” “Discussion is enough.” “Knowing the chapter is enough, by itself.” You are practicing the hearing, not racing.

Minute 4. Look at a student sentence such as “I can think about anything.” Decide, without answering them yet, which wrong answer that is. Install the error so you

⁴⁷Keith E. Stanovich and Richard F. West, papers in 2007 and 2008: myside bias is often independent of cognitive ability. “Be smarter” is not a myside cure.

⁴⁸Daniel T. Willingham, “Critical Thinking: Why Is It So Hard to Teach?,” *American Educator* 31, no. 2 (Summer 2007): 8–19. “You can teach students maxims about how they ought to think, but without background knowledge and practice, they probably will not be able to implement the advice they memorize.” Even three-year-olds can think critically in a domain they know; even trained scientists fail outside theirs. Access date for URLs in these notes: 30 August 2026.

can hear it tomorrow.

Minute 5. Write one sentence you will actually say: “We need the facts from this chapter before we can decide.” Put the pencil down. That sentence is the lesson.

If you can do those five minutes, you are ready to sit down. The student generates. You hear.

How to teach it this week

Run the session as in *The Practice Hour*. The idea on the table is the knowledge constraint. The subject on the table is whatever you already teach this week. That pairing is the method.

Warm-up (3–5 minutes, unaided). Three facts from this week’s chapter they can already get right. Mixed, not a chant of the same item. Paper. No device.

Short model (5 minutes). One claim from the same chapter. You retrieve the facts out loud. Then you decide, with reasons, what to believe or do. Then you stop.

Student attempt (10–15 minutes). Three to five items. First: retrieve. Second: decide, using what they retrieved. You wait. Ask, wait, hint, then model. Leave the pencil in their hand.

One good question. Not “did you think?” A question that names the constraint: “What did you have to know to decide that?” “What would you need to know that you do not know yet?” “If this fact were different, would the decision change?”

Mixed practice (5–10 minutes). Yesterday’s fact next to today’s claim. A science retrieval next to a history decision, so they feel which knowledge is missing. Pull one older item onto the page.

Exit ticket (5 minutes, unaided). One retrieval, one decision on a *new* sentence from the same chapter. Phone in another room. You hold the key.

Exact wording you can say

On the constraint:

“You cannot think carefully about a topic you do not know.”

“Knowledge is not the enemy of this work. It is what the work runs on.”

“Let’s get the facts first. Then we decide.”

On a stall:

“Show me what you already know from this chapter.”

“Name one fact you would need and do not have yet.”

“This is not a personality problem. This is a missing piece.”

On the two skips:

“Covering the chapter is not the same as deciding.”

“Discussing without the facts is talk. We’ll retrieve, then decide.”

When you are about to take over:

“Your pencil. I’ll ask a question.”

Then wait.

First problem for the student

Write this on a card, using this week’s actual subject. Illustration, not a reported family. If this week is a science chapter on experiments:

Here are two facts you already know from this chapter: [fact A], [fact B]. Here is a new claim: “The student who tested two variables at once found the cause.” What do you believe? What would you do next in the lab write-up? Which fact did you need?

Say: “Get the facts first. Then decide, with reasons.”

Wait. If the student jumps to “that’s bad science” with no fact, they have a maxim and no medium. If they retrieve and then decide, they heard the constraint. If they stare, show a fully worked different item — a history claim with two named facts, said out loud — then return to the lab. Fade: full example, completion, conventional item.

Later the same week, the diagnostic item:

“We don’t need the chapter. We can just think about it.”

Hear “facts are the enemy.” Hear “thinking first, knowledge later.” After the attempt, show a worked incorrect example that decided with no facts from the page. Ask what that person thought the work ran on. The student names the error.

How to fade help

Day of a new move: you model retrieval-then-decision fully. The student names the facts on a second. A third is missing only the decision line. A fourth is conventional. When they are accurate, stop showing full examples. Bring the retrieval back whenever a new unit arrives. Fading is for accuracy, not for a calendar date.

When to stop talking

After the question. After the hint. After the model of a *different* item. If you hear yourself supplying the chapter facts while their pencil is idle, you have started doing the retrieval for them. Stop. Point to the page. Wait.

Instead of “this is the hard part” or “I was never good at this subject,” say today’s sentence: “Let’s get the facts first.” The first pair fills their working memory with your feeling. The second sentence teaches.

Praise a named hole, a used fact, a changed mind — not speed and not a sharp tone.

“You noticed you didn’t have the date yet. That’s the move I want.”

A short pass on the constraint, then a handful of retrieve-then-decide items, then a return to the sentence, then more items including one from a subject they know less. Withhold neither the facts nor the unaided decision for weeks. Understanding and practice grow together. A picture with no practice does not hold. A decision with no picture does not transfer.

Here is an illustration, not a reported family. A parent has spent five minutes writing “you cannot think about what you do not know” next to two facts from a science chapter on fair tests. The student warms up by retrieving those two facts. The parent models one new claim out loud, using the facts, then leaves the decision line blank. The student finishes it. The parent asks, “Which fact did you need?” Mixed practice: one retrieval, one decision, one claim from a sport they do not follow so they can name the hole. Exit ticket: a new sentence from the science chapter, book closed. Right or wrong, the constraint was visible. Tomorrow’s first move is already on the page.

Practice that actually builds learning

Blocked, for a new move. The day you introduce the constraint, give a short set that is only retrieve-then-decide, all from one chapter they know. Four to six claims. Facts first, decision second, every time.

Mixed, for when to use it. Later the same week, mix: a pure retrieval, a decision that needs a fact from last month, a claim from a domain they know poorly so they can *name* the missing knowledge. Mixing teaches when the constraint is the point.

Brief retrieval of facts already known. Two or three minutes, mixed, untimed until they are right, then optionally brief timing on those known facts. Cover the answers. The facts are the subject's. This is not a thinking-gym. It is loading working memory with pieces that can be compared.

One incorrect example to diagnose. After they attempt, show a worked solution that decided with no facts from the page, or that recited facts and never decided. Ask which skip this is. The student names it.

Kitchen, money, and the news. A price comparison needs the prices. A forecast decision needs the map. That talk can motivate. It does not replace the chapter. A student who can argue about a commercial and cannot retrieve this week's science fact has performed, not learned.

A weekly shape. Day of the constraint, a short blocked retrieve-then-decide set. Same week: mixed with older facts. Later this week and next: a claim from a less-known topic so they practice saying "I do not know this yet." Monthly: an unannounced mixed check without notes.

Willingham's revisiting point is the long shape: even well-learned content fades. Plan to meet the same facts again in months, not only this Tuesday.⁴⁹ That is not extra busywork. It is how the medium stays available.

⁴⁹Daniel T. Willingham, paper for the NSW Department of Education, 2019. Four-step plan: identify what critical thinking means in each domain; identify the content those tasks need; sequence; plan multi-year revisiting. "Think like a historian" is not a goal until sourcing, corroboration, and contextualization are named. Punchline: "We are not even sure the general skills exist, but we are quite sure there is no proven way to teach them directly." Forgetting of well-learned content: Harry P. Bahrick, "Semantic Memory Content in Permastore," *Journal of Experimental Psychology: General* 113 (1984): 1–29; Bahrick and Lynda K. Hall, "Lifetime Maintenance of High School Mathematics Content," *Journal of Experimental Psychology: General* 120 (1991): 20–33; Andrew Pawl et al. on introductory physics, about half forgotten in three years, as cited by Willingham 2019.

For the student

You are learning a limit that makes you more honest, not smaller.

You cannot think carefully about a topic you do not know. That is not an insult. It is how thinking works.

Knowledge is not the enemy of critical thinking. It is what critical thinking runs on. The work is still deciding what to believe or do, with reasons. The reasons need facts.

Imagine a chessboard. If you do not know how a knight moves, you cannot judge a knight move. A slogan will not help. “Consider both sides” does not tell you the rule. You need the rule first. Then you can decide.

The same thing is true of a history chapter, a science lab, a news story, and a price. Get the facts. Then decide.

This is not like riding a bicycle. A bicycle, once learned, works on any road. Thinking is mixed with what you know. You can be careful in a domain you know. You need knowledge before the same care will work in a domain you do not know.

Tiny worked example

Fact from the chapter: a fair test changes one thing at a time.

New claim: “We changed the ramp *and* the mass, and we found the cause.”

What I know: a fair test changes one thing at a time.

Believe? I do not believe they found *the* cause. Two things moved.

Do? I would rerun the test, changing only one thing.

The decision used a fact. Without the fact, “that sounds bad” is only a mood.

Two tries

1. Write two facts you already know from *your* chapter this week. Then take one new sentence from that chapter. What do you believe? What would you do? Which fact did you need?
2. Take a topic you do *not* know well — a sport you do not follow, a period of history you have not studied, a tool you have not used. Try to decide a

claim about it. What knowledge are you missing? Write the hole. That is the right answer.

Explain it back

Tell someone at the table, in your own words, why you need something to think about. Then point to the ramp example and say which fact did the work. Then say whether “we don’t want mere facts” is a good plan for this week, and why or why not.

Challenge

Someone says, “I’m smart, so I can think about anything. Knowledge is just memorization.”

What would that person be skipping? Can you show them, with a topic *you* do not know, that smart is not the same as having the pieces?

You are allowed to struggle. You may open the chapter. You write. If you get stuck, ask for a hint — not the answer. Then try again.

A guess with no facts under it is still a guess. A list of facts with no decision is still a list. You can catch both. Retrieve. Then decide.

If it isn’t clicking

Three diagnostics. Each one has a next move. None of them is a verdict on talent, and none of them is a reason to wait for a birthday.

1. They refuse facts as “mere memorization.”

The medium has been framed as the enemy. Next move: smaller, useful facts from a chapter they already like. Retrieve three. Then one decision that is *impossible* without those three. Let them feel the stall, then the click. Stay here until “facts first” is a tool, not an insult. A tutor who will sit with the actual chapter, not a speech about curiosity, is a reasonable next step if the refusal is still the only move after several weeks of short daily work.

2. They recite the chapter and never decide.

Knowledge as a leak. Next move: after every retrieval, a forced *believe or do* line. Close the book. One new sentence. If they freeze, the model was too fast. Show

one worked decision, then they try. Slow down the “new” content until retrieve-then-decide is hearable. Go ahead once the unaided exit ticket includes both a fact and a decision. A tutor is optional here; the format is cheap. Bring someone in if recitation is still the only success after a couple of weeks of daily decide-lines *and* the hour has become a fight.

3. They decide fluently on topics they do not know.

Theater. Next move: a topic they cannot fake. Ask what they would need to know. Require the hole to be named before any argument. If they fill the air anyway, stop the speech. Point to the blank. Wait. This is not cruelty. It is the constraint, made visible. Go ahead once they can say “I do not know this yet” without shame, then return to a known chapter and decide well. A tutor is useful if fluent guessing is still the default after the hole-naming work is consistent, especially if the same student also refuses retrieval.

When to slow down. Facts treated as the enemy. Recitation with no decision. Fluent guessing on empty knowledge. Those are brakes. “Not ready” because of age is the brake this book will not use. “Not ready” because the medium is missing is a real brake.

When to go ahead. They retrieve known facts without a fight. They use those facts in a decision. They can name a hole in a domain they do not know. They hear “mere facts,” “just discuss,” and “thinking first, knowledge later” as misses. Then Chapter 3’s maxims have somewhere to fail in public, and somewhere to succeed when the knowledge is there.

When to get a human tutor. You have run the useful-facts work, or the forced-decision work, or the hole-naming work, for a stretch of daily sittings, and the same diagnosis is still the one in the room, and the hour has become a fight. A tutor is a release valve, not a failure of the sitting. Keep yourself as the person who can still hear a fluent guess with nothing under it.

Hear which strand is thin. A student who will not retrieve is missing the medium. A student who only retrieves is missing the activity. A student who decides without retrieval is missing both, covered by fluency. Hearing which is enough.

Tools, including AI

Optional helpers for you, the adult. The full rules live in *The Practice Hour*. This chapter is a pointer.

The student attempts first. You hold the key. A tool may explain the knowledge constraint *to you*, make extra retrieve-then-decide items from this week's chapter with answers hidden, write a hint script, or help you diagnose work already done.

A language model will happily write a “critical thinking” page with no facts from your chapter. Throw that page away. Generate items *from the text you are actually teaching*. Facts are not a chatbot job. The student retrieves from the chapter. You listen.

Bastani and colleagues: an unguarded chatbot raised practice scores and then cut the unaided exam.⁵⁰ Tonight's helped fluency is not this morning's judgment. Paste the worksheet, “what's the answer,” photo-to-key, and an unsupervised chatbot during the attempt stay out of the hour.

Search can be a tool for *you* when you need a second picture of a fact. It is not a substitute for the student's retrieval. If they look the fact up during the attempt, they did not have it in working memory to compare.

What “done enough” looks like

Placement is by skill, not birthday. A publisher's thinking workbook is a scope, not a legal grade.

Checklist before moving on

- The student can say, in their own words, that you cannot think carefully about a topic you do not know.
- On a new sentence from this week's chapter, they retrieve the needed facts, then decide what to believe or do, with a reason that uses those facts.
- They can name a knowledge hole in a domain they do not know, without filling the air with a guess.

⁵⁰Hamsa Bastani et al., “Generative AI without Guardrails Can Harm Learning,” *Proceedings of the National Academy of Sciences* 122, no. 26 (2025): e2422633122. Unguarded practice gains, then a cut on the unaided exam. Not a homeschool study. The pattern is the sitting condition.

- “Mere facts” is hearable as a skip of the medium. “Just discuss” is hearable as hope. “Knowing the chapter is enough, by itself” is hearable as the other skip. “Thinking first, knowledge later” is hearable as a sequence error. “I’m smart so I can think about anything” is hearable as the bicycle in a new hat.
- Retrieval of known facts is a normal part of the hour, mixed, short, unaided.
- The unaided exit ticket includes both a fact and a decision, more days than it fails.
- You can hear a fluent guess with nothing under it and ask a good question without taking the pencil.

If most of that list is true, go on to Chapter 3 — maxims are not enough — even if the birthday says otherwise. If the birthday says “eighth grade” and fluent guessing is still the default, stay. The next chapter is why slogans fail. This chapter’s job was to put something on the table to think about.

A path through judgment is the promise. A percentile is not.

Chapter 3

Maxims are not enough



Figure 4: A chessboard mid-game, a wine glass, and two books on a dark wooden table.

Why this matters

Chapter 2 put something on the table to think about. This chapter spends the next expensive fact: a good sentence about how to think will not, by itself, make the thinking happen.

“Consider both sides.” “Look for deep structure.” “Evaluate the source.” Those are maxims. They are not foolish. Daniel Willingham’s point is stricter. You can teach students maxims about how they ought to think. Without background knowledge and practice, they probably will not be able to implement the advice they memorize.⁵¹ The maxim cues a search. It does not supply the knowledge the search requires, and it does not make the student *notice* the analog sitting in memory.

What this idea unlocks is a Tuesday that is quieter and more honest. If a poster of slogans were enough, you could hang it and go to soccer. It is not enough. The work this week is to catch the empty maxim when you hear it, then to practice one maxim *on a claim the student actually knows*, until the move is real. Chapter 4 will put that practice inside the week’s real subject as a habit. This chapter teaches you to hear the failure, and to know what to do instead.

You do not need a list of fifty named biases, and you do not need a new poster. You do need to understand today’s idea well enough to hear “we covered consider-both-sides” as a recitation, not as evidence, and “look for deep structure” as advice most people cannot implement without the relevant knowledge.

This week you can learn one replacement that actually has a laboratory record: consider the opposite, on a claim the student already knows, using a fact from the page. Today the student can catch one empty maxim and write one opposite fact. That is enough for Tuesday. The zoo of named fallacies can wait. Most of it can wait forever.

The replacement is small on purpose. “Consider both sides” without knowledge produces two empty columns. “Consider the opposite” with a fact produces a checkable line. “Look for deep structure” as a sentence usually fails the noticing test. Putting yesterday’s problem next to today’s, and asking what is the same underneath, is the design that Gick and Holyoak’s hint condition already pointed

⁵¹Daniel T. Willingham, “Critical Thinking: Why Is It So Hard to Teach?,” *American Educator* 31, no. 2 (Summer 2007): 8–19. Maxims without background knowledge and practice will not implement. Access date for URLs in these notes: 30 August 2026.

at. Cue the comparison. Then hide the first problem and test.

For the parent: understand it yourself

Many adults feel rusty. That is ordinary. Five minutes of this section, then the warm-up at the end, is enough for tomorrow.

Everyday picture. Two people at a table. One says, “Consider both sides.” The topic is a sports foul, a historical cause, or a medical number. Neither person knows the rule, the period, or the base rate. They now have two empty sides. The maxim fired. Nothing was implemented. A second picture, still ordinary: a student who can say “look for something similar I already know,” and then cannot see that this week’s garden problem is the same mathematics as yesterday’s band problem. The advice was given. The noticing did not happen.

Precise picture. Surface structure captures attention — tumors and rays, gardens and marching bands, the nouns in the story. Deep structure is what would transfer — disperse force and converge; least common multiple; control of variables. Novices search memory by surface. That is why maxims fail as a teaching engine.

Mary Gick and Keith Holyoak’s numbers are the noticing gap in one line. Duncker’s radiation problem is solved spontaneously by about 10 percent of people. After they read an analogous fortress-and-army story, about 30 percent produce the convergence solution without a hint. With a hint to use the story, about 75 percent.⁵² The gap is not application skill. It is noticing. You can tell people to look for deep structure. You can even tell them which prior problem is similar. Most still cannot implement the strategy without the relevant knowledge. Willingham’s band-and-garden pair: 19 percent saw the analog; a hint that “this is like the garden problem” raised solution only to 35 percent. Most still could not implement.⁵³

Edward Thorndike and Robert Woodworth, in 1901, already limited what a general faculty can do: spread of practice only where identical elements are concerned.⁵⁴

⁵²Mary L. Gick and Keith J. Holyoak, “Analogical Problem Solving,” *Cognitive Psychology* 12 (1980): 306–55; Gick and Holyoak, “Schema Induction and Analogical Transfer,” *Cognitive Psychology* 15 (1983): 1–38. About 10 percent / 30 percent / 75 percent. The gap is noticing.

⁵³Willingham, 2007: band/garden isomorphic pair; 19 percent saw the analog; hint raised solution only to 35 percent.

⁵⁴Edward L. Thorndike and Robert S. Woodworth, “The Influence of Improvement in One Men-

Douglas Detterman's later summary: the default prediction is failure, and likelihood tracks similarity.⁵⁵ David Perkins and Gavriel Salomon distinguish low-road transfer — automatic triggering by similar conditions, mostly near — from high-road transfer — mindful abstraction, the only plausible route to far, effortful and rare unless the learner is cued to abstract.⁵⁶ A book that never forces comparison of structurally same, surface-different cases is betting on a weak mechanism. The pedagogical moral is not “tell them to look for deep structure.” It is: force the comparison, vary the surface, cue the abstraction, then test later, unaided, in a new context. That is design. It is not a maxim.

One maxim has better classical support than a zoo of named biases, and it still needs something to think about. **Consider the opposite** — generate a reason the other side could be right, or a way your favored claim could be wrong — has the best laboratory record against confirmation-related errors.⁵⁷ Naming a new bias every Monday is a poor bet. Hal Arkes called psychoeducation alone, in one blunt line, absolutely worthless as a correction method.⁵⁸ Scott Lilienfeld and colleagues' honest review: psychology has catalogued biases far more thoroughly than it has learned to correct them.⁵⁹ Carey Morewedge and colleagues showed that a serious game or a short video can move *lab bias scores* for months, on items that resemble training, in a paid sample. That is not nothing. It is not a civic vaccine, and it is not unaided spotting of myside in an identity-laden argument

tal Function upon the Efficiency of Other Functions,” *Psychological Review* 8 (1901): 247–61. Spread of practice only where identical elements are concerned; formal discipline fails.

⁵⁵Douglas K. Detterman, “The Case for the Prosecution: Transfer as an Epiphenomenon,” in *Transfer on Trial*, ed. Detterman and Robert J. Sternberg (Norwood, NJ: Ablex, 1993). Transfer is rare; likelihood tracks similarity; default prediction is failure.

⁵⁶David N. Perkins and Gavriel Salomon, “Are Cognitive Skills Context-Bound?,” *Educational Researcher* 18, no. 1 (1989): 16–25; Salomon and Perkins, “Rocky Roads to Transfer,” *Educational Psychologist* 24, no. 2 (1989): 113–42. Low-road versus high-road transfer.

⁵⁷Charles G. Lord, Mark R. Lepper, and Elizabeth Preston, “Considering the Opposite: A Corrective Strategy for Social Judgment,” *Journal of Personality and Social Psychology* 47 (1984): 1231–43. Also Edward R. Hirt and Keith D. Markman, 1995; Craig A. Anderson, 1982, on counter-explanation against theory perseverance. Best classical support among confirmation-related techniques in Lilienfeld's review.

⁵⁸Hal R. Arkes, “Impediments to Accurate Clinical Judgment and Possible Ways to Minimize Their Impact,” *Journal of Consulting and Clinical Psychology* 49 (1981): 323–30. Psychoeducation alone described as “absolutely worthless.”

⁵⁹Scott O. Lilienfeld, Rachel Ammirati, and Kristin Landfield, “Giving Debiasing Away: Can Psychological Research on Correcting Cognitive Errors Promote Human Welfare?,” *Perspectives on Psychological Science* 4, no. 4 (2009): 390–98. PsycINFO counts at the time: 1,211 “cognitive bias(es)” versus 158 “debias/debiasing.”

a year later.⁶⁰ This week you steal the small set, not the bestiary: consider the opposite, name the missing knowledge, compare two stories that share a structure. Inside a subject the student knows.

Deanna Kuhn's interviews are the everyday-argument cousin. On ordinary social questions — school failure, unemployment, crime — most people are weak at genuine evidence, alternatives, and counterarguments. Explanation is not argument. An explanation is a story about why something happens. An argument is a reason to prefer this claim to that one, under uncertainty.⁶¹ "Explain why you believe X" often elicits a story. "What would you need to see to abandon X, and what does the other side say?" elicits argument. Ask for both, named separately. "Just discuss the controversy" is not instruction.⁶²

Reflection is useful when it is tied to a standard, to feedback, and to knowledge the person can actually use. It is theater when it is uninformed — a journal prompt to "consider both sides" of a topic they do not know; fluency-fooled — a polished paragraph that feels like knowing; identity-serving; post-hoc — a clever reason for the first impulse; or unscored — no later unaided check.⁶³ Kruger and Dunning's loop: the same incompetence that causes the errors can block recognition of the errors; improving the skill improved the monitoring.⁶⁴ "Rate your confidence" without teaching the skill leaves the miscalibration in place.

⁶⁰Carey K. Morewedge et al., "Debiasing Decisions: Improved Decision Making With a Single Training Intervention," *Policy Insights from the Behavioral and Brain Sciences* 2, no. 1 (2015): 129–40. Games and videos moved researcher-built bias scales for months, paid convenience sample, items resembling training — not job performance or civic behavior.

⁶¹Deanna Kuhn, *The Skills of Argument* (Cambridge: Cambridge University Press, 1991), as reported by Jonathan Baron in *Informal Logic* (1992) and by Petra Barchfeld and Beate Sodian (2009): about 16 percent consistently generated genuine evidence; about 30 percent generated pseudoevidence; about one third consistently generated alternatives or counterarguments. Jonathan Osborne and Amanda Patterson, "Scientific Argument and Explanation: A Necessary Distinction?," *Science Education* 95, no. 4 (2011): 627–38. Explanation accounts for a phenomenon; argument justifies a claim under uncertainty.

⁶²Jonathan Osborne, Sibel Erduran, and Shirley Simon, "Enhancing the Quality of Argumentation in School Science," *Journal of Research in Science Teaching* 41, no. 10 (2004): 994–1020. Valid argument does not come naturally; it has to be taught with task structure and modelling.

⁶³John H. Flavell's metacognitive apparatus; Asher Koriati on experience-based judgments of learning and fluency as a treacherous cue. Reflection as theater when uninformed, fluency-fooled, identity-serving, post-hoc, or unscored — the constraint as used in this book's source record.

⁶⁴Justin Kruger and David Dunning, "Unskilled and Unaware of It," *Journal of Personality and Social Psychology* 77, no. 6 (1999): 1121–34. Bottom-quartile performers (actual about the 12th percentile) estimated themselves at about the 62nd; improving the skill improved the monitoring.

What to do instead, said as a parent picture rather than a review article. After they claim, ask for a fact. After the fact, ask for an opposite fact from the same page. After the opposite, ask what they now believe, and what they would do. If they cannot produce the opposite fact, the hole is knowledge, and you drop back to Chapter 2 for a few minutes of retrieval. If they produce a story — a script that elaborates their first idea without testing it — change the prompt: “What would you need to see to abandon this, and what does the other side say on this page?” Kuhn’s interviews found that kind of generation is hard for many people on everyday social issues. Osborne, Erduran, and Simon’s pedagogical moral still holds: valid argument does not come naturally; it has to be taught with task structure and modelling.⁶⁵ You are doing that structure at the table. You are not running a bias-of-the-week course.

Myside deserves one plain sentence so “think harder” does not sneak back in as the cure. People often evaluate evidence in a way that protects their own prior view, and that tendency is often independent of cognitive ability.⁶⁶ Slowing down can help when the knowledge is there. Slowing down can also produce a better-sounding press release for the first impulse. The opposite fact is the interruption. The unaided new claim is the check.

Wrong answers you should be able to hear

1. *“Consider both sides” — with no knowledge of either side.* Two empty columns. Hear a maxim with no medium. Next move: retrieve facts for *one* side from this week’s chapter, then generate one opposite that uses a fact, not a mood.
2. *“We did bias of the week. They can name the fallacy.”* Naming is not implementing. Hear psychoeducation as the weak cell. Next move: drop the name. Give a new surface from the same structure. Ask them to decide, then to consider the opposite.
3. *“Look for deep structure” — as if the sentence completed the search.* Willingham’s 35 percent, Gick and Holyoak’s noticing gap. Hear a cue that most people cannot execute without knowledge. Next move: put yester-

⁶⁵Jonathan Osborne, Sibel Erduran, and Shirley Simon, “Enhancing the Quality of Argumentation in School Science,” *Journal of Research in Science Teaching* 41, no. 10 (2004): 994–1020. Valid argument does not come naturally; it has to be taught with task structure and modelling.

⁶⁶Keith E. Stanovich and Richard F. West, 2007 and 2008: myside bias often independent of cognitive ability.

day's problem and today's problem on the same desk. Force the comparison. Then hide yesterday's and test.

4. "*Slow down and think harder.*" Sometimes useful. Incomplete. Smart people still defend their side; myside is often independent of cognitive ability.⁶⁷ Slowing down can produce a better-sounding justification of the same error. Hear a slogan that skips the knowledge and the opposite. Next move: "What would the other side say, using a fact from this page?"
5. "*They said the maxim, so they can do it.*" Recitation is performance. Soderstrom and Bjork: current performance is often an unreliable index of learning.⁶⁸ Hear a practice score treated as a skill. Next move: unaided exit ticket, new surface, book closed.

A sixth you will also hear: "Just discuss it." Talk is not automatically reasoning. Murphy and colleagues: more student talk is not more critical thinking.⁶⁹ Structure the talk: a claim, a fact, an opposite, a decision.

Five-minute parent warm-up

Do this before the lesson, on a scrap of paper, no student in the room.

Minute 1. Write: "Maxims cue. Knowledge and practice implement." Say it out loud.

Minute 2. Take one maxim — "consider both sides" or "consider the opposite." Apply it to a claim from this week's actual chapter. Write one fact for the favored side and one fact that could support the other side. If you cannot, you have found the knowledge hole.

Minute 3. True or false, said aloud: "Naming a bias is enough." "Consider both sides is enough without facts." "If they can say the maxim, they can do it." You are practicing the hearing.

Minute 4. Look at a student sentence such as "I considered both sides." Decide which wrong answer that is if no facts from the page appear. Install the error.

⁶⁷Keith E. Stanovich and Richard F. West, 2007 and 2008: myside bias often independent of cognitive ability.

⁶⁸Nicholas C. Soderstrom and Robert A. Bjork, "Learning Versus Performance," *Perspectives on Psychological Science* 10, no. 2 (2015): 176–99.

⁶⁹P. Karen Murphy et al., "Examining the Effects of Classroom Discussion on Students' Comprehension of Text," *Journal of Educational Psychology* 101, no. 3 (2009): 740–64.

Minute 5. Write one sentence you will actually say: “Tell me the opposite, using a fact from this chapter.” Put the pencil down. That sentence is the lesson.

If you can do those five minutes, you are ready to sit down. The student generates. You hear.

How to teach it this week

Run the session as in *The Practice Hour*. The idea on the table is that maxims are not enough. The subject on the table supplies the something to think about.

Warm-up (3–5 minutes, unaided). Two facts from this week’s chapter. One claim they already decided last week, restated in one line. Paper. No device.

Short model (5 minutes). One maxim, used *and* shown failing without knowledge, then used successfully on a known claim. You talk for a few minutes. Then you stop.

Student attempt (10–15 minutes). Three to five items. First: an empty maxim they catch. Second: consider-the-opposite on a known claim. You wait. Ask, wait, hint, then model.

One good question. Not “did you consider both sides?” A question that names the implementation: “What fact did the opposite use?” “What would you need to know to consider the other side, and do you have it?” “Show me yesterday’s problem next to this one. What is the same underneath?”

Mixed practice (5–10 minutes). A retrieve, an opposite, a new surface of an old structure. Pull one older item onto the page.

Exit ticket (5 minutes, unaided). One new claim from the same chapter. Book closed. Consider the opposite, with a fact. You hold the key.

Exact wording you can say

On the failure:

“A maxim is advice. It is not the thinking. The thinking needs facts from this chapter.”

“‘Consider both sides’ with no knowledge is two empty columns.”

On the replacement:

“Consider the opposite. Use a fact from this page, not a mood.”

“What would you need to see to change your mind?”

“Put yesterday’s problem next to today’s. What is the same underneath?”

On a recitation:

“You said the words. Show me the fact the opposite uses.”

When you are about to take over:

“Your pencil. I’ll ask a question.”

Then wait.

First problem for the student

Illustration, not a reported family. If this week is a history chapter:

Claim: “The new law caused the change.”

1. Write one fact from this chapter that supports the claim.
2. Consider the opposite: write one fact from this chapter that could support a different cause, or that could make the law unnecessary for the change.
3. Decide what you believe, and what you would do next — close the book, or look for one more source.

Say: “The maxim is consider the opposite. The work is using a fact. Empty sides do not count.”

Wait. If they write “the other side has a point” with no fact, they recited. If they use a fact for the opposite, they implemented. If they stare, show a fully worked different item — a science claim with one supporting fact and one confounding fact — then return to the law. Fade: full example, completion, conventional item.

Later the same week, the diagnostic item:

“I looked for deep structure.”

Hear recitation. After the attempt, show two problems that share a structure and differ in cover story. Ask whether saying the maxim was enough, or whether they had to see the pair. The student names the gap.

How to fade help

Day of a new move: you model empty-maxim-then-implemented-opposite fully. The student fills the opposite fact on a second. A third is missing only the decision. A fourth is conventional. When they are accurate, stop showing full examples. Bring the forced comparison back when a new unit arrives with a new surface.

When to stop talking

After the question. After the hint. After the model of a *different* item. If you hear yourself generating the opposite for them, you have started doing the search. Stop. Point to the page. Wait.

Instead of “consider both sides” as a mantra, say today’s sentence: “Tell me the opposite, using a fact from this chapter.” The mantra fills the air. The sentence teaches.

Praise an opposite that uses a page fact, a catch of their own empty maxim, a pair of problems they can see as the same underneath — not a list of fallacy names.

Here is an illustration, not a reported family. A parent has spent five minutes applying consider-the-opposite to a history claim, with one supporting fact and one competing cause from the same chapter. The student warms up on those two facts. The parent models an empty “both sides” and crosses it out, then writes the two facts and a decision. The student tries a new claim. The parent asks, “What fact did the opposite use?” Mixed practice: one retrieval, one opposite, two cover stories that share a structure. Exit ticket: a new sentence, book closed. The maxim was named. The fact did the work.

A short pass on why maxims fail, then a handful of implemented opposites, then a forced comparison of two surfaces, then an unaided new claim. Withhold neither the maxim’s name nor the fact it needs. Understanding and practice grow together.

Practice that actually builds learning

Blocked, for a new move. The day you introduce consider-the-opposite, give a short set that is only that, all from one known chapter: claim, supporting fact, opposite fact, decision. Four to six items.

Mixed, for when to use it. Later the same week: a pure retrieval, an empty maxim to catch, two stories that share a structure, a new claim. Mixing teaches *when* the maxim is the right cue, and when it is a recitation.

Brief retrieval of facts already known. Two or three minutes. The opposite cannot be generated from a blank. Cover the answers. You listen.

One incorrect example to diagnose. After they attempt, show a worked “I considered both sides” with no facts. Ask what that person thought a side was. Show a named-fallacy recitation that still picked the first impulse. The student names the error.

Kitchen, money, and the news. A price, a forecast, a share button can host consider-the-opposite *if* they have the numbers or the source. That talk can motivate. It does not replace the chapter. A student who can say “both sides” at dinner and cannot produce an opposite fact from this week’s page has performed, not learned.

A weekly shape. Day of the maxim, a short blocked implemented-opposite set. Same week: mixed with a forced comparison of two surfaces. Later this week and next: the same structure in a new cover story, unaided. Monthly: an unannounced mixed check without notes.

Near is not far. A same-week success on a similar item is real and small. It is not proof that the maxim will fire a year later, at work, with identity on the line.⁷⁰ Design for identical elements and cued comparison. Then test later, unaided. That is enough honesty for a kitchen table.

For the student

You are learning why a good slogan is not yet good thinking.

A **maxim** is a short piece of advice: “consider both sides,” “look for something similar you already know,” “slow down.” Advice can help. Advice is not the work.

The work is still deciding what to believe or do, with reasons, on something you actually know.

If you do not know either side, “consider both sides” gives you two empty columns.

⁷⁰Susan M. Barnett and Stephen J. Ceci, “When and Where Do We Apply What We Learn? A Taxonomy for Far Transfer,” *Psychological Bulletin* 128, no. 4 (2002): 612–37. Near and far are positions on several dimensions (knowledge domain, time, social context, modality), not a binary. Same-week similar items are near.

If you cannot see that today's story is the same kind as yesterday's, "look for deep structure" is a sentence you said, not a seeing you did.

What to do instead, this week:

1. Get a fact from the page for the claim you like.
2. **Consider the opposite:** get a fact from the page that could support a different claim, or that could make your claim weaker.
3. Then decide. Name what you believe, and what you would do next.

That is the maxim *implemented*. The fact is what makes it real.

Tiny worked example

Claim: "The plant grew better because we talked to it."

Fact for the claim: the talked-to plant is taller.

Opposite fact from the same lab: that plant also got more sun, because it sat by the window.

Believe? I do not yet believe talking was the cause. Two things moved.

Do? I would run a fair test: same sun, only talking changes.

"Consider both sides" without those facts would have been a shrug.

Two tries

1. Take one claim from *your* chapter this week. Write a fact that supports it. Write a fact that considers the opposite. Decide what to believe or do.
2. Here is a maxim: "Look for something similar you already know." Find two problems from this week and last week that look different on the surface and might be the same underneath. If you cannot find the pair, write what knowledge you would need. That is honest work.

Explain it back

Tell someone at the table why a maxim is not enough. Then point to the plant example and say which fact did the opposite work. Then say whether "I considered both sides" is enough, by itself, and why or why not.

Challenge

Someone says, “We learned the names of ten biases. I can spot them anywhere now.”

What would that person be treating as the skill — naming, or implementing on a new page? Can you show them, with a new claim from *your* chapter, that the name is not the move?

You are allowed to struggle. You may open the chapter. You write. If you get stuck, ask for a hint — not the answer. Then try again.

Empty sides are a miss. A named bias with no opposite fact is a miss. You can catch both. Use a fact. Then decide.

If the opposite will not come, that is usually a missing fact, not a missing personality. Open the chapter. Find one line that could support a different claim. Write it. Then decide again. That second decision is the work. The first impulse can stay on the page as a draft. The opposite fact is how you check it.

If it isn't clicking

Three diagnostics. Each one has a next move. None of them is a verdict on talent, and none of them is a reason to wait for a birthday.

1. Every answer is the maxim, with no fact.

Recitation. Next move: ban the slogan for three days. Require two facts from the page before any decision. If they cannot produce the facts, the hole is Chapter 2, and you drop back. Stay here until an opposite without a fact feels unfinished to them, not only to you. A tutor who will sit with the chapter, not a fallacy poster, is a reasonable next step if recitation is still the only move after several weeks of short daily work.

2. They can consider the opposite in last week's chapter and freeze on this week's.

The move did not travel because the knowledge did not. That is the noticing gap, not stubbornness. Next move: put last week's worked opposite next to this week's claim. Force the comparison. Then hide last week's and try again. Slow down the new unit until the pair is visible. Go ahead once the unaided exit ticket on a *new* sentence from the current chapter includes an opposite fact. A tutor is optional;

the pair on the desk is cheap. Bring someone in if the freeze continues after daily paired work *and* the hour has become a fight.

3. They generate a story instead of an opposite.

Explanation posing as argument. Next move: change the prompt. Not “why do you believe that?” — “what would you need to see to abandon it, and what fact on this page could support the other side?” Model one. Then they try. Stay here until a story without a competing fact is hearable as a miss. A tutor is useful if story-as-argument is still the default after the prompt change is consistent.

When to slow down. Maxims with no facts. Transfer freeze because the new chapter’s knowledge is missing. Stories standing in for opposites. Those are brakes. “Not ready” because of age is the brake this book will not use. “Not ready” because the maxim has nothing to implement on is a real brake.

When to go ahead. They catch an empty maxim. They implement consider-the-opposite with a fact from the page. They can put two different-looking problems side by side and say what is the same underneath, or name the knowledge they would need to see it. Then Chapter 4’s weekly practice has a move worth embedding.

When to get a human tutor. You have run the no-slogan work, or the paired-chapter work, or the evidence-not-story work, for a stretch of daily sittings, and the same diagnosis is still the one in the room, and the hour has become a fight. A tutor is a release valve, not a failure of the sitting. Keep yourself as the person who can still hear “I considered both sides” with nothing under it.

Hear which strand is thin. Recitation is a missing implementation. Freeze on a new chapter is a missing medium. A story is a missing argument. Hearing which is enough.

Tools, including AI

Optional helpers for you, the adult. The full rules live in *The Practice Hour*. This chapter is a pointer.

The student attempts first. You hold the key. A tool may explain why maxims fail *to you*, make extra consider-the-opposite items from this week’s chapter with answers hidden, write a hint script, or help you diagnose work already done.

A language model will happily produce a “both sides” worksheet with no facts from your book. That worksheet is the empty maxim in digital form. Generate items *from this week’s actual chapter*. The student supplies the opposite fact from the page.

Bastani and colleagues: unguarded practice gains, then a cut on the unaided exam.⁷¹ A fluent “here are both sides” paragraph from a chatbot is performance, not later unaided judgment. Paste the worksheet, “what’s the answer,” photo-to-key, and an unsupervised chatbot during the attempt stay out of the hour.

A bias poster is a tool too. If you keep one, keep it small: consider the opposite; name the missing knowledge; compare two surfaces. Take the fifty-name zoo off the wall this week.

What “done enough” looks like

Placement is by skill, not birthday. A publisher’s “logical fallacies” packet is a scope, not a legal grade.

Checklist before moving on

- The student can say, in their own words, that a maxim is advice, not the thinking, and that the thinking needs facts.
- On a new claim from this week’s chapter, they can write a supporting fact, an opposite fact, and a decision about what to believe or do.
- “I considered both sides” with no facts is hearable as a miss.
- “We named the bias” is hearable as recitation, not as implementation.
- “Look for deep structure” is hearable as a cue that still needs a comparison they can actually see.
- “Slow down and think harder” is hearable as incomplete.
- They can put two different-looking problems side by side and name the shared structure, or name the knowledge they would need.
- The unaided exit ticket on a new sentence includes an implemented opposite, more days than it fails.
- You can hear an empty maxim and ask a good question without taking the pencil.

⁷¹Hamsa Bastani et al., “Generative AI without Guardrails Can Harm Learning,” *PNAS* 122, no. 26 (2025): e2422633122.

If most of that list is true, go on to Chapter 4 — how to practice this week inside a real subject — even if the birthday says otherwise. If the birthday says “freshman logic” and the maxim is still empty, stay. The next chapter is the weekly habit. This chapter’s job was to retire the poster as the engine.

A path through judgment is the promise. A percentile is not.

Chapter 4

How to practice this week



Figure 5: A chalkboard, a torn list on parchment, wooden puzzle cubes, and an old desk clock.

Why this matters

Chapters 1 through 3 gave you a sentence, a constraint, and a warning. The work is deciding what to believe or do. You need something to think about. Maxims fail without knowledge and practice. This chapter is the Tuesday. This week's practice sits inside a real subject.

A standalone “critical thinking hour” with nothing to think about is the wrong gym. Stamping the words on every lesson and hoping is the other wrong gym. Philip Abrami and colleagues put the hope in one line that still designs a school: improvement “cannot be a matter of implicit expectation.”⁷² Objectives have to be explicit. The pedagogical combination that actually showed up, in a later analysis, is not a course title. It is dialogue, a problem that is real to the student, and mentoring.⁷³ You are the mentor. The problem is a claim from this week's history, science, mathematics, or news. The dialogue is one good question, then waiting.

What this idea unlocks is a week you can actually run. You do not need a new curriculum. You need the subject you already teach, the session shape from *The Practice Hour*, and a named move from Chapters 1 through 3. Domain-embedded practice is the load-bearing beam. Control of variables can be taught in a science setup. Sourcing can be taught with documents. Diagnosis, in Geoffrey Norman's line of work, is knowledge plus cases, not a missing general trait.⁷⁴ Those are different recognitions attached to different knowledge. This week you pick one subject and practice one named decision inside it.

You do not need to be a researcher. You do need to understand today's idea well

⁷²Philip C. Abrami et al., “Instructional Interventions Affecting Critical Thinking Skills and Dispositions: A Stage 1 Meta-Analysis,” *Review of Educational Research* 78, no. 4 (2008): 1102–34. Improvement “cannot be a matter of implicit expectation.” 117 studies, 20,698 participants, 161 effects; $g^+ = 0.341$; immersion weakest. Access date for URLs in these notes: 30 August 2026.

⁷³Philip C. Abrami et al., “Strategies for Teaching Students to Think Critically: A Meta-Analysis,” *Review of Educational Research* 85, no. 2 (2015): 275–314. 341 effect sizes from quasi- or true experiments using standardized critical-thinking measures; weighted random-effects mean $g^+ = 0.30$. Dialogue, authentic or situated problems, and mentoring combined, 19 studies, $g = 0.57$, as summarized in Hitchcock's methods supplement and the TUM Clearing House Unterricht short review of Abrami 2015. Content-specific critical-thinking outcomes $g = 0.57$ (97 effects); disposition $g = 0.23$ (25 effects). 0.30 is not 0.57.

⁷⁴Geoffrey Norman et al., “The Etiology of Diagnostic Error: A Psychological Autopsy,” *Academic Medicine* 92, no. 1 (2017): 23–30. Diagnostic error is knowledge access, not a missing general critical-thinking trait; teaching is many cases with feedback.

enough to hear “we think critically in everything” as a hope, not as a plan, and “we’ll think after we finish the chapter” as a sequence error.

This week you can learn the session shape well enough to run it inside one subject you already teach. Today the student can try one new sentence from that subject, unaided, after a short model. That is the whole promise of the chapter. Later domains — a live page, a clinic number, a sentence a machine will finish — will use this Tuesday. They will not replace it.

Pick one gym for the week — one real subject already on the table. History, science, mathematics, or the news you were going to read anyway. Pick one named move: Ennis’s sentence, retrieve-then-decide, or consider-the-opposite. Run the hour. The move stays. The page supplies the something to think about. If you remember only one sentence from this chapter, remember that one: this week’s practice is inside a real subject, with the move named, and with an unaided check at the end of the hour.

For the parent: understand it yourself

Many adults feel rusty. That is ordinary. Five minutes of this section, then the warm-up at the end, is enough for tomorrow.

Everyday picture. The history book is already on the table. So is a science lab, a math page, or a news story you were going to read anyway. That page is the gym. You retrieve two facts. You model one judgment. They try. You ask what they would need to see to change their mind. They practice a short blocked set, then a mixed set. They finish with one unaided claim. Soccer still happens. The thinking happened *inside* the subject, not instead of it.

A second everyday picture: a purchased workbook of generic puzzles, brain-teasers, and “what’s wrong with this argument” items about topics nobody in the house studies. That workbook can be a warm-up. It is not the week’s practice. Identical elements do not live there. The elements live in the chapter you are actually teaching.

Precise picture. Robert Ennis named four course architectures.⁷⁵ **General:** a

⁷⁵Robert H. Ennis, “Critical Thinking and Subject Specificity: Clarification and Needed Research,” *Educational Researcher* 18, no. 3 (1989): 4–10. Four approaches: general, infusion, immersion, mixed. Ennis later argued (2013, 2018) for a mixed undergraduate design — a pro-

standalone thinking course; content incidental. **Infusion:** deep subject instruction, encouragement to think, *and* the principles named. **Immersion:** the same depth without naming the principles. **Mixed:** a general thread plus infusion or immersion in the subjects. Abrami's 2015 analysis found that those four types did **not** differ significantly from one another on standardized tests. Mixed was numerically largest; immersion smallest. Design mixed anyway, for transfer reasons Chapters 2 and 3 already gave: identical elements have to be practiced in the places a life actually contains, and high-road abstraction has to be cued. The 2015 paper is not proof that mixed won a horse race. Numerically largest is not a race won.⁷⁶

What *is* settled: making the objective explicit beats hoping. Dialogue plus authentic or anchored problems plus mentoring is the combination with the largest reported effects in that analysis — $g = 0.57$ from 19 studies, against an overall $g+$ of 0.30 on standardized tests.⁷⁷ Those numbers are not interchangeable, and they are not far transfer, job performance, or civic wisdom. They are a reason to name the move, use a real claim, and stay in the chair as a mentor — which, at a kitchen table, means you ask, wait, hint, and then model, without taking the pencil.

Domain-embedded practice is the beam. A meta-analysis of control-of-variables instruction found a mean $g = 0.61$; assessment format was a major source of variance, which is a reminder that what you count as “having the skill” depends on how you ask.⁷⁸ Avishag Reisman's *Reading Like a Historian* study was a quasi-experiment, not a randomized trial: 236 eleventh-graders, five San Francisco high schools, six months of document-based curriculum versus textbook instruction,

positional, not a completed multi-institution randomized trial.

⁷⁶David Hitchcock, *Stanford Encyclopedia of Philosophy*, methods supplement, reporting Abrami 2015 by Ennis type: mixed +0.38, infusion +0.29, general +0.26, immersion +0.23. The difference was not statistically significant. Design mixed for transfer reasons; 2015 did not pick a winner.

⁷⁷Philip C. Abrami et al., “Strategies for Teaching Students to Think Critically: A Meta-Analysis,” *Review of Educational Research* 85, no. 2 (2015): 275–314. 341 effect sizes from quasi- or true experiments using standardized critical-thinking measures; weighted random-effects mean $g+ = 0.30$. Dialogue, authentic or situated problems, and mentoring combined, 19 studies, $g = 0.57$, as summarized in Hitchcock's methods supplement and the TUM Clearing House Unterricht short review of Abrami 2015. Content-specific critical-thinking outcomes $g = 0.57$ (97 effects); disposition $g = 0.23$ (25 effects). 0.30 is not 0.57.

⁷⁸Martin Schwichow, Christoph Croker, Corinne Zimmerman, Tim Höffler, and Hendrik Hartig, “Teaching the Control-of-Variables Strategy: A Meta-Analysis,” *Developmental Review* 39 (2016): 37–63. 72 intervention studies, 226 pairwise comparisons, mean $g = 0.61$ (95% CI 0.53–0.69); assessment format a major source of between-study variance.

significant effects on historical thinking, transfer of those strategies to contemporary issues, factual knowledge, and general reading.⁷⁹ It is not a generic-critical-thinking result. It is an existence proof that a named disciplinary move can be taught in ordinary schools. Norman and colleagues: diagnostic error is knowledge access, not a missing general trait; teaching is many cases with feedback.⁸⁰ Steal the shape. Teach a named move, in a subject, with practice and an unaided check.

Christopher Huber and Nathan Kuncel's college picture is the long version of the same warning. Ordinary college already moves generic tests about half a standard deviation in four years. Curriculum-wide campaigns to improve critical thinking did not necessarily add a reliable increment. Nursing's mandated critical-thinking curriculum did not beat other majors on domain-general tests.⁸¹ Stamping the phrase on every syllabus is immersion-by-committee. Abrami 2008 predicts that will not move the needle. Your house does not need a stamp. It needs this week's subject and a named decision.

Childhood still needs a page, without turning a pedagogy into an attainment engine. Matthew Lipman's community of inquiry — judgment, criteria, self-correction, sensitivity to context, talk with children about reasons — is a way of practicing the activity on stories they can actually handle.⁸² Two large independent trials of a related UK programme told different stories on school attainment: a 2015 efficacy trial moved reading and maths a little; a 2021 effectiveness trial, larger and more

⁷⁹Avishag Reisman, "Reading Like a Historian: A Document-Based History Curriculum Intervention in Urban High Schools," *Cognition and Instruction* 30, no. 1 (2012): 86–112. Quasi-experiment, 236 eleventh-graders, five San Francisco high schools, six-month document-based curriculum versus textbook. MANCOVA significant on historical thinking, transfer to contemporary issues, factual knowledge, and general reading. Not an RCT; not a generic-critical-thinking result.

⁸⁰Norman et al., 2017. Case specificity: lots of cases, because a general inventory will not capture case-specific knowledge.

⁸¹Christopher R. Huber and Nathan R. Kuncel, "Does College Teach Critical Thinking? A Meta-Analysis," *Review of Educational Research* 86, no. 2 (2016): 431–68. Mixed-design four-year estimate for critical-thinking skills 0.59 SD; purely longitudinal 0.46 SD; disposition (CCTDI) 0.55 SD over four years. Curriculum-wide efforts did not necessarily produce incremental long-term gains; nursing's mandated curriculum did not outperform other majors on domain-general tests.

⁸²Matthew Lipman, "Critical Thinking — What Can It Be?," *Educational Leadership* 46, no. 1 (1988): 38–43. "Skillful, responsible thinking that facilitates good judgment because it (1) relies upon criteria, (2) is self-correcting, and (3) is sensitive to context." Community of inquiry as childhood pedagogy.

secure, moved neither, and showed no harm.⁸³ Dialogic practice may still be worth the hour if the goal is the hour of dialogue. It is not a proven engine of a generic faculty, and it was never measured that way at scale. Keep the talk. Keep it on something they know. Keep the unaided check.

Wrong answers you should be able to hear

1. *"This is our thinking hour. We don't need the history book."* A content-free hour. Hear the bicycle gym. Next move: open the history book. The thinking hour is the history hour, with the move named.
2. *"We think critically in everything we do."* A stamp. Hear implicit expectation. Next move: name today's move in one sentence, pick one claim, check unaided. "Everything" is not a plan.
3. *"We had a great discussion."* Talk without a named move, a real problem, and mentoring. Hear Murphy: more talk is not automatically more reasoning.⁸⁴ Next move: one authentic claim, one good question, an unaided line at the end.
4. *"The workbook of puzzles is the program."* Generic items as the engine. Hear Thorndike: identical elements. Next move: keep a puzzle as a two-minute warm-up if you like. Put this week's chapter in the center.
5. *"We'll think after we finish the chapter."* Sequence error. Understanding and practice grow together. Next move: retrieve, then decide, *this* week, on a new sentence from the same chapter.

A sixth you will also hear: "They liked the lesson, so it worked." Enjoyment is not the unaided check. Teacher glow survived in a trial that moved zero months on attainment.⁸⁵ Like the hour. Still give the exit ticket.

⁸³Stephen Gorard, Nadia Siddiqui, and Beng Huat See, Education Endowment Foundation efficacy trial of Philosophy for Children, 2015: 48 English primary schools; KS2 n = 1,529; reading +0.12, maths +0.10, writing +0.03, CAT4 +0.07; larger for FSM-eligible pupils. EEF effectiveness trial, 2021, evaluator NFER: 198 schools, 3,601 pupils, 5/5 padlock; **0 months** additional progress on FSM reading and on reading and maths for the whole cohort; no harm to attainment. Dialogic practice may be worth the hour as dialogue. It is not a proven engine of generic critical thinking at scale.

⁸⁴P. Karen Murphy et al., "Examining the Effects of Classroom Discussion on Students' Comprehension of Text," *Journal of Educational Psychology* 101, no. 3 (2009): 740–64.

⁸⁵Stephen Gorard, Nadia Siddiqui, and Beng Huat See, Education Endowment Foundation efficacy trial of Philosophy for Children, 2015: 48 English primary schools; KS2 n = 1,529; reading

Five-minute parent warm-up

Do this before the lesson, on a scrap of paper, no student in the room.

Minute 1. Write: “This week’s practice is inside [name the subject].” Say it out loud. Write today’s named move: Ennis’s sentence, retrieve-then-decide, or consider-the-opposite.

Minute 2. Open this week’s actual chapter. Work two claims yourself, on paper, without the key until you finish.

Minute 3. True or false, said aloud: “A separate thinking hour is the engine.” “Stamping the words is enough.” “A good discussion is the check.” You are practicing the hearing.

Minute 4. Look at a student sentence such as “we already think in everything.” Decide which wrong answer that is. Install the error.

Minute 5. Write one sentence you will actually say: “Open the chapter. We’ll decide on a new sentence, with reasons.” Put the pencil down. That sentence is the lesson.

If you can do those five minutes, you are ready to sit down. The student generates. You hear.

How to teach it this week

Run the session as in *The Practice Hour*. This chapter does not rebuild the hour. It tells you what to put on the table: a real subject, a named move, an authentic claim, you as mentor.

Warm-up (3–5 minutes, unaided). Facts from this week’s subject they can already get right. Mixed. Paper. No device. A generic puzzle, if you use one, lasts two minutes and then you close it.

Short model (5 minutes). One claim from the chapter. Name the move. Retrieve. Decide, or consider the opposite. Then you stop.

+0.12, maths +0.10, writing +0.03, CAT4 +0.07; larger for FSM-eligible pupils. EEF effectiveness trial, 2021, evaluator NFER: 198 schools, 3,601 pupils, 5/5 padlock; **0 months** additional progress on FSM reading and on reading and maths for the whole cohort; no harm to attainment. Dialogic practice may be worth the hour as dialogue. It is not a proven engine of generic critical thinking at scale.

Student attempt (10–15 minutes). Three to five items from the same chapter. You wait. Ask, wait, hint, then model. Pencil in their hand.

One good question. Not “did we think critically?” A question that names the week’s move: “What are you deciding?” “What fact did you need?” “What is the opposite, using this page?”

Mixed practice (5–10 minutes). Today’s move next to yesterday’s type. A science claim next to a history retrieval if both are live this week. Pull one older item onto the page.

Exit ticket (5 minutes, unaided). One or two new sentences from the same chapter. Book closed. No hints. You hold the key.

That is infusion with the principle named, plus a thin general thread — Ennis’s sentence, said once — which is as close to mixed as a kitchen table needs. Explicit. Dialogic. Authentic. Mentored. Domain-embedded.

Exact wording you can say

On the gym:

“This week’s practice is inside this chapter. The chapter is the gym.”

“We are not doing a separate thinking hour. We are deciding, with reasons, on a new sentence from the book we already have.”

On the named move (pick one per week):

“Today’s move: decide what to believe or do, with a fact from this page.”

“Today’s move: retrieve first, then decide.”

“Today’s move: consider the opposite, using a fact.”

On mentoring:

“Show me what you tried.”

“I’ll be quiet.”

“Your pencil. I’ll ask a question.”

On the stamp:

“‘We think in everything’ is not today’s plan. This claim is the plan.”

First problem for the student

Write this on a card, using this week's actual subject. Illustration, not a reported family.

From today's chapter, copy one sentence you have not discussed yet.

1. What are you deciding — what to believe, what to do, or both?
2. Which fact from this chapter do you need?
3. What is one opposite fact, or one knowledge hole?
4. Decide. Write the reason.

Say: "This chapter is the gym. The move has a name. I'll be quiet."

Wait. If they ask for a generic puzzle instead, they are looking for the wrong gym. If they use the sentence from the chapter, they heard this week. If they stare, show a fully worked different sentence from the same chapter, then leave the last line blank. Fade: full example, completion, conventional item.

Later the same week, the diagnostic item:

"We already think critically in everything, so we can skip the exit ticket."

Hear the stamp. Hear enjoyment as a check. After the attempt, give the exit ticket anyway. The student names what the stamp skipped: the unaided new sentence.

How to fade help

Day of a new subject-unit: you model one fully. The student names the move on a second. A third is missing only the decision. A fourth is conventional. When they are accurate, stop showing full examples. When a new unit arrives, the facts are new; the session shape is not. Bring the full model back for the new knowledge, not for the shape.

When to stop talking

After the question. After the hint. After the model of a *different* item. If you hear yourself turning the hour into a lecture about "how we think," you have left the subject. Stop. Point to the chapter. Wait.

A short named move, then a handful of claims from the chapter, then mixed practice, then an unaided new sentence. Withhold neither the subject nor the check. Understanding and practice grow together. Practice the new move this week, not

next month. A slogan with no page and no “why” is a chant. A page with a named move, a real attempt, and an unaided line is the hour.

Here is an illustration, not a reported family. Monday: science chapter, retrieve-then-decide, blocked set of four claims about fair tests. Tuesday: mixed — two fair-test claims, one retrieval from last week’s measurement unit, one “what would you do.” Wednesday: consider-the-opposite on a new lab sentence, using a fact. Thursday: two cover stories that share the same structure, one from the lab and one from a kitchen comparison of recipes. Friday: two unaided sentences from the chapter, book closed. The subject did not change. The named move had a week to live in it. That is infusion with the principle named, plus Ennis’s sentence said once on Monday, which is as close to mixed as a kitchen needs.

Practice that actually builds learning

Blocked, for a new move. The day a named move is new in this subject, give a short set that is only that, all from this chapter. Four to six claims. Same structure, so they learn to execute.

Mixed, for when to use it. Later the same week: yesterday’s type, last month’s type, a retrieval, today’s move. Mixing teaches *when*. A page of twenty identical “what do you think?” prompts feels like mastery at 9 p.m. It often is not still there in two weeks.⁸⁶

Brief retrieval of facts already known. Two or three minutes, mixed, from this subject. Cover the answers. You listen. The medium has to be available or the move has nothing to hold.

One incorrect example to diagnose. After they attempt, show a worked hour that never opened the chapter, or a discussion with no unaided line, or a puzzle workbook treated as the program. Ask what that person thought the gym was. The student names the error.

Kitchen, money, and the news. Ordinary decisions can host the same named move — a price, a forecast, a share — *if* the facts are on the table. They can

⁸⁶Doug Rohrer, Robert F. Dedrick, and Kaleena Burgess, *Psychonomic Bulletin & Review* 21 (2014): 1323–30; Rohrer, Dedrick, Marissa K. Hartwig, and Chi-Ngai Cheung, *Journal of Educational Psychology* 112, no. 1 (2020): 40–52. Grade-7 interleaved mathematics. This chapter borrows the practice of mixing types after a short blocked introduction, not those coefficients as a thinking-hour finding.

motivate or give a second picture. They do not replace the chapter, the written reasons, or the unaided problem. A student who can argue at dinner and cannot decide on a new sentence from this week's science has performed, not learned.

A weekly shape. Pick one subject as the gym this week. Pick one named move. Day 1: blocked. Days 2–3: mixed with retrieval. Day 4: a new surface of the same structure. Day 5: unannounced unaided check, two items, book closed. Next week the subject may change. The shape does not.

If you teach two ages, you may still run one shape. The younger student gets claims from their chapter; the older student gets claims from theirs. The named move can be the same sentence. The facts cannot. That is ordinary, not a scheduling failure. Two children, two pages, one hour-shape. *The Practice Hour* already said the shape does not change when the idea does.

Word problems, lab write-ups, and news stories are types, not magic words. A “study says” sentence is a claim about evidence, not a reason to believe. Vary the cover story so they match the type — a cause, a comparison, a source, a number — not the noun. Kitchen, money, and the news can host the same type *after* the chapter has supplied the structure. They do not replace the chapter as first picture.

Dialogue here is not a seminar. It is one good question and a wait. Authentic means the claim is from a page they are actually studying, or from a decision they will actually make. Mentoring means you stay in the chair, you hold the key, and you do not take the pencil. That combination is the 0.57 cell, scaled to a kitchen, labelled as a scaling. It is a useful study of teaching ingredients, not a promise that every home will see the same result.⁸⁷

For the student

You are learning where the practice lives.

This week's practice is inside a real subject — the history chapter, the science lab,

⁸⁷Philip C. Abrami et al., “Strategies for Teaching Students to Think Critically: A Meta-Analysis,” *Review of Educational Research* 85, no. 2 (2015): 275–314. 341 effect sizes from quasi- or true experiments using standardized critical-thinking measures; weighted random-effects mean $g^+ = 0.30$. Dialogue, authentic or situated problems, and mentoring combined, 19 studies, $g = 0.57$, as summarized in Hitchcock's methods supplement and the TUM Clearing House Unterricht short review of Abrami 2015. Content-specific critical-thinking outcomes $g = 0.57$ (97 effects); disposition $g = 0.23$ (25 effects). 0.30 is not 0.57.

the math page, the news you are already reading. That page is the gym. There is not a separate hour called “thinking” with nothing to think about.

The work is still deciding what to believe or do, with reasons. You still need facts. A maxim still needs something to implement on. This week you *do* it, on a new sentence from the book in front of you.

A puzzle book can be a warm-up. It is not the program. “We think in everything” sounds nice. It is not a plan. A good discussion is not the check. The check is one or two problems you do alone, book closed.

Tiny worked example

Subject: this week’s science chapter.

New sentence from the chapter: “The bigger the wheels, the faster the car.”

Named move: retrieve, then decide, then consider the opposite.

Fact I retrieve: last lab, the heavy car with big wheels was *slower* on the carpet.

Opposite: wheel size and weight both changed in some trials.

Believe? I do not yet believe “bigger wheels, faster car” as a rule.

Do? I would test wheel size with weight held still.

The gym was the lab. The move had a name. The check was a new sentence.

Two tries

1. Open *your* chapter. Copy one sentence you have not discussed. Decide what to believe or do. Use a fact. Consider the opposite with a fact, or name a hole.
2. Someone wants to skip the chapter and do a puzzle book labeled “critical thinking” instead. What would they be treating as the gym? What would you put on the table instead, this week?

Explain it back

Tell someone at the table where this week’s practice lives. Then point to the wheels example and say why the lab was the gym, not a separate thinking hour. Then say whether “we think in everything” is a plan, and why or why not.

Challenge

Design tomorrow's exit ticket for yourself. Two new sentences from the same chapter. Book closed. No hints. After you try, check: did you name the decision, use a fact, and consider an opposite? If you needed the book, that is useful. Tomorrow you fade.

You are allowed to struggle. You write. If you get stuck, ask for a hint — not the answer. Then try again.

The chapter in front of you is enough, if you know what to do with it. Name the move. Use a fact. Decide. Then do one alone.

If it isn't clicking

Three diagnostics. Each one has a next move. None of them is a verdict on talent, and none of them is a reason to wait for a birthday.

1. They want a separate thinking hour, or a puzzle book, instead of the chapter.

Wrong gym. Next move: two minutes of a puzzle if it keeps the peace, then the chapter, every day this week. The exit ticket is always from the chapter. Stay here until the chapter feels like the place the move lives. A tutor who will sit with the actual subject, not a brain-teaser packet, is a reasonable next step if the wrong gym is still the request after several weeks of short daily work.

2. The hour is all talk, and the unaided line never happens.

Discussion as the whole program. Next move: timer on the talk — five minutes, then pencil, then exit ticket. You may like the talk. The check is still unaided. Slow down new content until the unaided line is a habit. Go ahead once the exit ticket from the chapter works more days than it fails. A tutor is optional; the timer is cheap. Bring someone in if talk-without-check is still the default after a couple of weeks *and* the hour has become a fight.

3. They can do the move in last week's subject and not in this week's.

That is Chapter 2 traveling with you, not a failure of this week's shape. Next move: retrieve this week's facts first, longer warm-up, then the same named move. If the facts are not there, stay on retrieval until they are. The session shape does not change. The medium does. Go ahead once retrieve-then-decide works in the

current chapter. A tutor is useful if the stall repeats at every unit change after the longer warm-up is consistent.

When to slow down. Wrong gym. Talk without a check. Unit change with no facts. Those are brakes. “Not ready” because of age is the brake this book will not use. “Not ready” because this week’s subject is not yet available as a medium is a real brake.

When to go ahead. The chapter is the gym. The move is named. Dialogue is one good question, not the whole hour. The unaided new sentence works more days than it fails, in the subject you are actually teaching. Then later chapters — the test is not the thing, leaving the page, stakes, machines — have a weekly habit to sit on.

When to get a human tutor. You have run the chapter-as-gym work, or the timer-and-ticket work, or the longer-retrieval work, for a stretch of daily sittings, and the same diagnosis is still the one in the room, and the hour has become a fight. A tutor is a release valve, not a failure of the sitting. Keep yourself as the person who can still hear a stamp, a puzzle-as-program, and a discussion with no check.

Hear which strand is thin. Wrong gym is a missing place. All talk is a missing check. Freeze in a new unit is a missing medium. Hearing which is enough.

For a student already struggling in the subject itself, explicit instruction in that subject — clear models, practice, think-alouds, feedback — has the first claim on the hour.⁸⁸ The thinking move sits on that instruction. It does not replace it.

A useful study is not a promise that every home will see the same result. Abrami’s combination cell is a teaching ingredient list, scaled to a kitchen and labelled as a scaling. If Friday’s unaided lines are still empty, stay in this chapter’s shape another week. Name the move again. Open the same subject. Mix a little more. The later chapters can wait. A path is the promise. A calendar date is not.

If two subjects are live in the same week, pick one as the gym and let the other be retrieval only. Splitting the named move across two thin pages is how the hour becomes a stamp again. One gym. One move. The shape from *The Practice Hour*. That is enough.

⁸⁸National Mathematics Advisory Panel, *Foundations for Success* (2008), Finding 27; Lynn S. Fuchs et al., *Assisting Students Struggling with Mathematics*, WWC 2021006 (IES, 2021), Recommendation 1. Explicit instruction has consistently positive effects for students already struggling with the content. The thinking move sits on that instruction.

Tools, including AI

Optional helpers for you, the adult. The full rules live in *The Practice Hour*. This chapter is a pointer.

The student attempts first. You hold the key. A tool may explain infusion and the session shape *to you*, make extra claims from this week's chapter with answers hidden, write a hint script, or help you diagnose work already done.

A language model will happily invent a week of “critical thinking activities” with no connection to your book. Throw that week away. Generate items *from the chapter you are teaching*. The student works those items on paper.

Bastani and colleagues: an unguarded chatbot raised practice scores and then cut the unaided exam.⁸⁹ Tonight's helped discussion, with a tool in reach, is not the exit ticket. Paste the worksheet, “what's the answer,” photo-to-key, and an unsupervised chatbot during the attempt stay out of the hour.

You can run this entire week with no AI and no extra workbook. Many families will. The subject you already have is enough. The hour still has the same shape. You still hold the key.

What “done enough” looks like

Placement is by skill, not birthday. A publisher's “critical thinking, grade 8” book is a scope, not a legal grade. You may use such a book as extra mixed practice. You do not owe it the center of the hour.

Checklist before moving on

- This week's practice happened inside a real subject, not in a content-free hour.
- A named move was on the table — Ennis's sentence, retrieve-then-decide, or consider-the-opposite — said in ordinary words.
- The session shape ran: warm-up, short model, student attempt, one good question, mixed practice, unaided exit ticket.
- “We think in everything” is hearable as a stamp, not as a plan.

⁸⁹Hamsa Bastani et al., “Generative AI without Guardrails Can Harm Learning,” *PNAS* 122, no. 26 (2025): e2422633122. Unguarded practice gains, then a cut on the unaided exam. Not a homeschool study. The pattern is the sitting condition.

- A puzzle workbook is hearable as a warm-up, not as the program.
- A good discussion is hearable as incomplete without an unaided line.
- “We’ll think after the chapter” is hearable as a sequence error.
- The unaided new sentence from this week’s chapter works more days than it fails.
- You can hear a wrong gym, a stamp, and talk-without-check, and ask a good question without taking the pencil.

If most of that list is true, you have the weekly habit this book will keep using. Later chapters will change the domain — a test score, a live page, a clinic number, a sentence a machine will finish. They will not rebuild the hour. If the birthday says “ready for media literacy” and the chapter is still not the gym, stay. The later chapters need this Tuesday.

A path through judgment is the promise. A percentile is not.

Chapter 5

The test is not the thing



Figure 6: A teacher's desk with books, a tied manuscript, pencils, and a chalkboard diagram in window light.

Why this matters

A score on a short “critical thinking” test is a sitting. It is not the skill.

Robert Ennis’s public sentence, already in this book, names the work: reasonable reflective thinking focused on deciding what to believe or do.⁹⁰ A forty-minute inventory of unstated assumptions is a thin sample of items, taken on a particular morning, under whatever motivation the sitting had. A percentile on that inventory can be real. It is still a percentile on those items. The life you are preparing the student for is a source, a number, a hire, a share, a sentence that already looks finished. Those stakes do not grade the sitting.

This chapter is worth the struggle because families get offered the sitting as if it were the work. A campus reports that scores went up four points. A prep book promises a Watson–Glaser gain. A co-op wants a “critical thinking” badge on the transcript. None of those objects is worthless. Each of them is easy to confuse with judgment. Confusion is expensive. You can spend a semester raising a coached score and still have a student who cannot source a document, recode a percent, or explain a conclusion with the notes closed.

What you can learn this week is how to tell a test from a judgment, and how to collect evidence at home that the student actually decided. What the student can try today is one unseen case in a subject they already know, written out, then explained without the page.

You do not need to become a psychometrician. You do need today’s idea well enough to hear “I scored well, so I can think” as a category mistake, not as a compliment.

Daniel Willingham’s constraint still holds: thinking is not a bicycle-skill you learn once and ride everywhere.⁹¹ Domain knowledge plus practiced evaluation is the work. A short test cannot be the meaning of that work. This chapter teaches you how to look at evidence of judgment without treating a commercial inventory as

⁹⁰Robert H. Ennis, working definition as restated in Ennis, “The Nature of Critical Thinking: An Outline of Critical Thinking Dispositions and Abilities,” criticalthinking.net (2011); already the 1985 *Educational Leadership* working definition: “Critical thinking is reasonable reflective thinking focused on deciding what to believe or do.”

⁹¹Daniel T. Willingham, “Critical Thinking: Why Is It So Hard to Teach?,” *American Educator* 31, no. 2 (Summer 2007). Thinking is not a bicycle-skill; processes are intertwined with domain knowledge; maxims without knowledge and practice will not implement. Domain-specific critical thinking can be taught.

the thing.

For the parent: understand it yourself

Many adults feel rusty around tests with names like Watson–Glaser or CCTST. That is ordinary. Five minutes of this section, then the warm-up at the end, is enough for tomorrow.

Everyday picture. A written driving exam. The student names the meaning of a yellow light and the distance for a stop. That sitting is useful. It is not the same object as driving in rain, at night, with a child in the back. A high score on the written exam is still a high score on the written exam. You would not call it “driving.” You would not skip the road.

Precise picture. Peter Facione’s 1990 Delphi report was expert consensus *for purposes of educational assessment and instruction*: forty-six panelists, six rounds, a shared construct so programs could test and teach at scale.⁹² From that construct came the California Critical Thinking Skills Test and a companion disposition inventory. The report did not discover a mental organ. It operationalized a consensus statement. A book that treats a CCTST score as “what critical thinking is” has confused an instrument with Ennis’s sentence.

Nathan Kuncel’s review, as Christopher Huber and Kuncel restated it, is the constraint the inventories live under. Common critical-thinking tests show little discriminant validity from general cognitive ability. There is little evidence that they predict grades or job performance *better than IQ*.⁹³ Two senses of the phrase sit on the same page, and they are not the same claim. One is field-specific expertise that does not readily transfer — the historian’s sourcing, the physician’s differential. The other is a thin set of reasoning moves useful when problems share those elements: sample size, correlation versus causation, unstated assumptions. If a test samples that thin set, a gain means “better at those items.” That can be a real

⁹²Peter A. Facione, *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction* (The Delphi Report), ERIC ED315423 (1990). Forty-six panelists; six rounds, 11 February 1988–25 September 1989. Consensus for assessment and instruction, not a discovery of nature. CCTST is one operationalization of that construct.

⁹³Nathan R. Kuncel, review of critical-thinking tests and general cognitive ability (2011), as restated in Christopher R. Huber and Nathan R. Kuncel, “Does College Teach Critical Thinking? A Meta-Analysis,” *Review of Educational Research* 86, no. 2 (2016): 431–68. Little discriminant validity from *g*; little evidence of prediction better than IQ.

gain. It is still not a life of judgment.

The map, in ordinary language, so you can hear the names when they arrive.

Watson–Glaser: selected-response items in five advertised buckets — inference, recognition of assumptions, deduction, interpretation, evaluation of arguments. Widely used as an employment screen, especially in law-firm hiring. Independent studies have reported subscale consistencies as low and wide as 0.17 to 0.74. A meta-analysis of sixty published studies found a single-component structure, which undercuts the five-score story.⁹⁴ A hiring screen can be useful without being a learning outcome. It can also be a general-ability test with a critical-thinking label.

Cornell Critical Thinking Test, with Ennis among the authors: multiple-choice, closer to Ennis’s outline than Facione’s list. An open-ended cousin, the Ennis–Weir essay, scores a letter to the editor on argument moves. Closer to the conception is a reason to prefer it as a *teaching* task. It is not a reason to treat a scored letter as the meaning of deciding what to believe or do.

CCTST: multiple-choice, about thirty-four items in about forty-five minutes, widely used in nursing because accreditation asked for a critical-thinking outcome. Independent subscale alphas have been reported as low as 0.21 to 0.51, against author-reported figures around 0.68 to 0.70.⁹⁵ An alpha of 0.21 is not a diagnostic. It is noise with a name. Huber and Kuncel found that nursing students, despite the requirement and heavy use of this inventory, did not show larger long-term generic-test gains than other majors.⁹⁶ That null is not “thinking cannot be taught.” It is a null on these tests, in the profession that tried hardest to move them.

CLA and CLA+: a document-based memo plus selected-response items. The con-

⁹⁴Watson–Glaser Critical Thinking Appraisal, current forms including WGCTA and Watson–Glaser II. Robert Loo and K. Thorpe, independent subscale consistencies 0.17–0.74 (1999), cited by Ou Lydia Liu, Lois Frankel, and Katrina Crotts Roohr, ETS Research Report RR-14-10 (2014). Robert M. Bernard et al., meta-analysis of 60 published studies (2008): single-component structure.

⁹⁵Independent CCTST subscale alphas 0.21–0.51, Leppa (1997), against author-reported about 0.68–0.70; pattern noted by Kelly Y. L. Ku (2009). Liu, Frankel, and Roohr, ETS RR-14-10, is the map.

⁹⁶Huber and Kuncel, “Does College Teach Critical Thinking?” Nursing students did not show larger long-term generic-test gains than other majors despite accreditation demand and heavy CCTST use.

struct is a blend of reasoning and writing. Richard Arum and Josipa Roksa's *Academically Adrift* reported that 45 percent of students showed no statistically significant CLA gain in the first two years.⁹⁷ That is a serious book about effort and curriculum. It is a poor warrant for "critical thinking cannot be taught." The 45 percent is a CLA finding, a two-year window, a writing-plus-reasoning blend, and, as the next paragraph will say, partly a motivation number.

The Halpern Critical Thinking Assessment used everyday scenarios with both open-ended and forced-choice items. It had one of the rare correlations with negative life events. It did not control for general cognitive ability. By 2024 it was retired.⁹⁸ There is no standard instrument named HALO. What exists, and gets confused with that acronym, is a holistic scoring rubric, the halo effect in rating (one strong trait coloring the rest), and the retired Halpern. AAC&U VALUE rubrics for critical thinking have been studied for rater severity; one report found 36 percent perfect agreement across raters.⁹⁹ Useful as a teaching tool. Weak as a high-stakes measure.

Ou Lydia Liu, Lois Frankel, and Katrina Crotts Roohr's ETS map is the single best tour of these instruments.¹⁰⁰ You do not need the tour memorized. You need the rule: each operationalizes a conception under time and format constraints. None of them is Ennis's sentence.

Two further threats, so a campus press release cannot do your thinking for you.

Daniel Koretz distinguished meaningful gain — more teaching, harder work, better methods — from reallocation, alignment, coaching, and cheating.¹⁰¹ Critical-thinking inventories are short and coachable. A semester of practice items is coach-

⁹⁷Richard Arum and Josipa Roksa, *Academically Adrift* (Chicago: University of Chicago Press, 2011). Twenty-four institutions; more than 2,300 students; 45 percent no statistically significant CLA gain in the first two years; 0.18 SD over three semesters; 0.47 SD over four years. CLA finding, partly motivation; not "thinking cannot be taught."

⁹⁸Halpern Critical Thinking Assessment: everyday scenarios; Butler (2012) $r = -.38$ with negative life events, not controlled for g . Described as retired by 2024 in a *Journal of Intelligence* review. No standard instrument named HALO was found.

⁹⁹AAC&U VALUE critical-thinking rubric: Finley (2012) reported 36 percent perfect agreement across raters. Halo is a rater effect, not a test. The Holistic Critical Thinking Scoring Rubric is a four-level teaching tool; weak as a high-stakes measure.

¹⁰⁰Liu, Frankel, and Roohr, "Assessing Critical Thinking in Higher Education," ETS RR-14-10 (2014).

¹⁰¹Daniel Koretz, D. F. McCaffrey, and L. S. Hamilton (2001); Koretz (2005/2008): meaningful gain versus reallocation, alignment, coaching, and cheating.

ing until a far task the coach did not teach says otherwise.

Motivation on a low-stakes sitting can flip the conclusion. Liu, Brent Bridgeman, and Rachel Adler showed that conclusions about college learning can reverse depending on whether students are induced to try.¹⁰² End-of-semester inventories that do not affect a grade are typically low-stakes for the student. A student who does not try has not demonstrated that two years taught nothing. The student has demonstrated that the occasion did not require trying.

CLA+'s sixty-minute constructed-response reliability has been reported at 0.43; the higher test-level figure is largely driven by the multiple-choice section.¹⁰³ The part that looks like judgment is the unreliable part. Authenticity and reliability trade. A memo that looks like work is not, by looking like work, a precise measure of a person.

If the construct is judgment — deciding what to believe or do in an unseen, under-specified case — then a short selected-response inventory is the wrong genre. Redesign language, not a fitted Monday protocol, points at three kinds of evidence: unseen cases, sampled many times because one case is not the next case; process that makes the move visible, with a live check because process can be forged; and a short oral defense as verification that the artefact had a student attached.¹⁰⁴ Those are validity ideas borrowed from medicine's many cases, history's new documents, and viva theory. They are not a trial that "oral defense raises CCTST." They are how a grade, or a kitchen folder, could be evidence of judgment rather than evidence of a sitting.

Wrong answers you should be able to hear

1. *"I scored in the ninetieth percentile, so I can think critically."* The sitting went well. The items were those items. Ask what the student would do with a new document in a domain they know, notes closed.

¹⁰²Ou Lydia Liu, Brent Bridgeman, and Rachel M. Adler, "Measuring Learning Outcomes in Higher Education: Motivation Matters," *Educational Researcher* 41, no. 9 (2012): 352–62. Conclusions about college learning can flip depending on whether students are induced to try.

¹⁰³Zahner (2013, CAE): 60-minute CLA+ constructed-response reliability 0.43; test-level reliability 0.87 largely driven by the 30-minute multiple-choice section, as reported by Liu, Frankel, and Roohr.

¹⁰⁴Geoffrey Norman et al. (2017): many cases, because of case specificity. Gordon Joughin (2010) on viva as verification. Role-play as anchored instruction, Abrami et al. 2015, $g = 0.61$ from 5 studies — suggestive, small k , not a proven oral-defense accelerator of a generic trait. Redesign is language, not a fitted protocol.

2. *“The ‘evaluation of arguments’ subscore is low, so we should drill evaluation.”* Independent work finds those subscales too unreliable to steer instruction. Watson–Glaser often measures one factor. A named cell on a report is not five skills.
3. *“The take-home essay proves they can think.”* A take-home essay is a writing sample. A language model can finish one. That problem is older than the models; Norman Frederiksen already argued that convenient artefacts shape the skills we get.¹⁰⁵ The essay can be *part* of the evidence. It cannot be the whole.
4. *“Those tests are fake, so thinking cannot be taught.”* The inventories are thin. Domain-specific judgment can be taught. Chapter 4 already spent that sentence on control of variables, historical thinking, and cases. A weak instrument is not an impossibility proof.
5. *“I can tell they’re thinking. We do not need a record.”* Fluency fools. A finished-looking paragraph is easy to process. Ease is not evaluation. A live check on an unseen case is the record this chapter wants.

Five-minute parent warm-up

Do this before the lesson, on a scrap of paper, no student in the room.

Minute 1. Write Ennis’s sentence in your own words: deciding what to believe or do, reasonably, after reflection. Circle *or do*. The test items usually sample *believe*. A life includes sharing, consenting, refusing.

Minute 2. Look at one multiple-choice “assumption” item from any free sample, or invent a thin one: “All dogs bark. Fido is a dog. Therefore Fido barks. What was assumed?” Notice what the item is. It is a logic move. It is not a source, a number, or a hire.

Minute 3. Write one unseen case in a subject you already teach: a short primary paragraph in history, a lab claim in science, a news headline in government. The student has not seen this exact page. That is the point.

¹⁰⁵Norman Frederiksen, “The Real Test Bias: Influences of Testing on Teaching and Learning,” *American Psychologist* 39, no. 3 (1984): 193–202. Convenient artefacts shape the skills we get. Generative models made the take-home-prose hole unskippable; they did not invent it. Mike Perkins and colleagues’ AI Assessment Scale (2024) is language for aligning allowed tools with evidence; dissemination is not an effect size.

Minute 4. Decide what “process” would look like for that case: a sourcing note, a count, a search log, a map of claim and objection. Decide the live check: two minutes of “explain it back” with the page turned over.

Minute 5. Write one sentence you will actually say: “A score is a sitting. Today we are collecting evidence that you decided.” Put the pencil down. That sentence is the lesson.

If you can do those five minutes, you are ready to sit down. The student generates. You hear.

How to teach it this week

A good critical-thinking hour this week has the same shape as the rest of this book: warm-up, short model, student attempt, one good question, mixed practice, exit ticket. The content is evidence of judgment, not a prep book.

Warm-up (3–5 minutes, unaided). Two facts the student already knows in this week’s subject. One small “what would count as evidence?” question. Paper. No device.

Short model (5 minutes). One new idea, one picture, one written sentence. The picture is the driving-exam versus the road, or a sample inventory item next to an unseen case. You talk for a few minutes. Then you stop.

Student attempt (10–15 minutes). One unseen case in a domain they know. They write what they believe, what they would do, and what would change their mind. You wait. Struggle before rescue: ask, wait, hint, then model. Leave the pencil in their hand.

One good question. Not “what’s the answer?” A question that names the idea: “What sitting would this be evidence of? What would it *not* be evidence of?” “If this were a test item, what would the test be sampling? What is the actual stake?”

Mixed practice (5–10 minutes). Yesterday’s subject next to today’s. A logic item next to a document. A percent next to a headline. The mix is the point: the student has to notice which kind of evidence is being asked for.

Exit ticket (5 minutes, unaided). Two items. One thin reasoning item they can already do. One live check: explain the unseen case with the page closed. You hold the key. Phone in another room.

That shape is a practice you impose on whatever book is on the table. It is not a 180-day plan, and it is not a commercial inventory.

Exact wording you can say

On the sitting:

“A score tells us how the sitting went. It does not tell us whether you can decide this new case.”

“What was the test sampling? Assumptions? Writing? Motivation to try?”

On the case:

“This page is new. You know the subject. What do you believe, and what would you do?”

“What would change your mind? Write that down before you defend the first answer.”

On process:

“Show me the move, not only the conclusion. Where did you look? What did you count? What did you reject?”

On a fluent paragraph:

“This looks finished. Looking finished is not the same as being checked.”

When you are about to take over:

“Your pencil. I’ll ask a question.”

Then wait. Ten seconds of silence is teaching.

First problem for the student

Write this on a small card or the top of the page, labelled as an illustration, not as a reported family:

A one-paragraph claim in a subject you already teach. History: a short excerpt that names a battle and a motive. Science: “this fertilizer doubled yield.” Government: a headline about a poll. The student has background. They have not seen this paragraph.

Say: “This is not a test with a published key. Decide what you believe, what you would do next, and what would change your mind. Then turn the page over and tell me.”

Wait. If they copy the paragraph’s conclusion in more polite words, they treated fluency as evidence. If they ask who produced it, what the numbers were, or what another source says, they have started judging. If they stare, show a fully worked *different* case — a weather claim you recode together — then return to theirs. That is fading: full example, then the item.

Later the same week, the diagnostic item: a sample “critical thinking” multiple-choice question next to the same kind of unseen case. Ask which object each one is. Hear “they’re the same.” Hear “the test is fake so neither matters.” Hear “the case is the work; the item is a sitting.” After they attempt, you may show a worked *incorrect* example that treated a high practice-test score as proof of judgment and ask what that person confused. Generation first, then the named error.

How to fade help

Day of a new move: you model one fully. The student explains a second. A third is missing only the last written step. A fourth is conventional. When they are accurate at naming “sitting versus case,” stop showing full examples. Bring the model back when a new domain arrives — health numbers in Chapter 7, a live page in Chapter 6 — because the distinction has to be rebuilt in the new content.

When to stop talking

After the question. After the hint. After the model of a *different* case. If you hear yourself explaining the student’s current case in real time while their pencil is idle, you have started doing the work. Stop. Point to the two objects. Wait.

Practice that actually builds learning

Blocked practice for a new move. Mixed practice for when to use it. Brief retrieval of facts already known. One incorrect example to diagnose. Kitchen talk can motivate. It does not replace the unaided case.

Blocked, this week. Three unseen cases in the *same* subject, same kind of move. If the subject is history, three short documents, each new, each asking who produced this and what would corroborate it. If the subject is science, three claims that need

a control. The student is learning to see the move, not to hunt a keyword.

Mixed, later in the week. A logic item, a document, a percent, a headline. The cover stories differ. The job is to notice which kind of evidence is being asked for. Keyword lists fail here the way they fail in word problems. “Evaluate” appears on tests, in headlines, and in lab write-ups. The word does not tell you the object.

Retrieval. Two minutes of already-known domain facts before the case. You cannot judge a Civil War letter without Civil War knowledge. You cannot judge a fertilizer claim without a picture of yield. Willingham’s second point still holds: metacognitive prompts help when the student already has the knowledge the prompt requires.¹⁰⁶ A retrieval burst is not a quiz for its own sake. It is how you make the case thinkable.

One incorrect example. After the student has attempted, show a worked error: a student (composite, labelled as an illustration) who raised a practice-test score by twenty points and then trusted a fluent take-home paragraph they could not explain with the page closed. Ask: what was the sitting evidence of? What was missing? The student names the error. You do not narrate it as a verdict on talent.

Keep the practice short enough to finish. A folder of unfinished “critical thinking journals” is not process evidence. Three completed cases with a live check beat a dozen prompts.

Abrami and colleagues found that making thinking an explicit objective beats hoping it emerges as a by-product, and that dialogue plus authentic problems plus mentoring is the combination with the largest reported effects on standardized inventories.¹⁰⁷ Those numbers measure the inventories. This week’s practice is the authentic problem in a domain you already teach, with you as the mentor, and the move named out loud. That is mixed design in kitchen form: a short explicit strand plus subject-matter work. It is not a promise of a percentile.

¹⁰⁶Daniel T. Willingham, “Critical Thinking: Why Is It So Hard to Teach?,” *American Educator* 31, no. 2 (Summer 2007). Thinking is not a bicycle-skill; processes are intertwined with domain knowledge; maxims without knowledge and practice will not implement. Domain-specific critical thinking can be taught.

¹⁰⁷Philip C. Abrami et al., 2008: improvement “cannot be a matter of implicit expectation.” Abrami et al., 2015: $g+ = 0.30$ on standardized critical-thinking tests; dialogue plus authentic problems plus mentoring $g = 0.57$ from 19 studies. Ennis-type mixed / infusion / general / immersion differences were not statistically significant.

For the student

You are not a percentile. You are a person who will have to decide what to believe or do when nobody is scoring the morning.

A test can be useful. It samples a few kinds of items: assumptions, arguments, maybe a memo. If you got those items right, you got those items right. That is honest. What it does not prove is that you can handle the next new page in a subject you know — a letter, a lab claim, a headline — and then act.

Here is the move. When someone shows you a score, ask: a sitting of what? When someone shows you a finished paragraph, ask: what did I decide, and can I still say it with the page closed?

Tiny worked example. A paragraph says a new after-school program “cut failing grades in half.” Looks finished. Looks like success. A sitting might ask you to spot an unstated assumption in a similar sentence. The case asks you to decide. What was the base number? Half of what? Who was in the program? What else changed that year? What would you do next — celebrate, ask for the counts, wait? Write the decision. Turn the page over. Say it.

Two tries.

1. A sample logic item: “If it rains, the picnic is cancelled. The picnic is cancelled. Therefore it rained.” What, if anything, went wrong? Write one sentence.
2. A new paragraph in a subject you already study this year. Your teacher or parent will pick it. Decide what you believe, what you would do, and what would change your mind. Then close the page and say those three things out loud.

Explain it back. In your own words: what is the difference between a test sitting and a judgment on a real case? If you can say that without this chapter open, you have the idea.

Challenge. Find one score you have been given this year — a quiz, a rubric, a practice test — and write two columns: *what this is evidence of* and *what this is not evidence of*. Be kind to the sitting. Be honest about the gap. Then pick one real case this week where you will collect the missing evidence: a source, a count, or a two-minute explanation with the notes closed.

You are allowed to struggle. Struggle is the work, not a delay before the work. A fluent answer that you cannot rebuild is not yet yours.

If it isn't clicking

Three diagnostics, and the next move for each. No shame. Placement is by skill, not by birthday. A “grade 8 critical thinking workbook” is a publisher’s scope, not a legal grade.

1. They treat every score as the skill. Next move: two columns on paper, every time a score appears this week. Left: sitting of what. Right: what we still do not know. Keep it to one line each. When the columns fill without prompting, the confusion is easing.

2. They freeze on an unseen case even in a subject they know. Next move: shrink the case. One paragraph, one question (“who produced this?” or “what would count as evidence?”), one live sentence. Knowledge first, then the evaluation prompt. If they still freeze, the hole is often the subject, not “thinking.” Go back to the content. A prompt cannot supply the Civil War.

3. They produce fluent pages they cannot explain closed. Next move: the live check becomes the exit ticket, every sitting, for two weeks. Two minutes. Page down. You ask what they believe, what they would do, what would change their mind. If they cannot say it, the page was performance. That is useful information. It is not a character verdict. Shorten the writing. Lengthen the talk.

When to slow down: the student cannot yet retrieve the domain facts the case needs. Stay in the subject. Evaluation waits on knowledge.

When to go ahead: they can name sitting versus case, complete three unseen problems in one subject with a live check, and catch a fluent paragraph that they cannot rebuild.

When to get a human tutor: you cannot yet hear the difference yourself in this year’s subject, or the student’s writing difficulty is so large that the live check always collapses into a reading problem. A tutor in the *subject* is the right valve. A generic “critical thinking coach” with a prep book is the sitting again.

Huber and Kuncel found that ordinary college already moves generic tests about half a standard deviation over four years, and that campus-wide campaigns stamp-

ing the phrase on every syllabus did not add a reliable increment.¹⁰⁸ Hoping is the weakest design. Named practice in a domain is the stronger one. If this week is not clicking, name a smaller move in a known subject. A longer inventory is the wrong next purchase.

Tools, including AI

Optional helpers for the adult. The student attempts first. You hold the key. A tool may explain the *idea* to you, make extra practice with answers hidden, write a hint script, or help you diagnose work already done.

Keep this short. This is not an AI policy chapter. Chapter 8 is the life condition that a machine will finish the sentence. Here the job is assessment hygiene.

A box of rules

- Child attempts first.
- Parent holds the answer key, including the live-check questions.
- Use a language model to explain Delphi versus Ennis to *you*, to draft three extra unseen cases in this week's subject with answers hidden until you check them, or to help you write a two-minute oral-check script.
- A tool does not grade the live check. You do.
- During the attempt: no worksheet pasted into a chat, no "what's the answer," no photo-to-key, no unsupervised chatbot.

A language model can draft a transcript *template* later, in the records chapter, with fields a stranger can read. Asking it to invent a score, a sitting that did not happen, or a "critical thinking" percentile the family never earned is a false statement on a document. Drafting a template is not signing.

Detectors that claim to catch machine-written prose fail often, and they fail more often on honest writers whose English is not a first language.¹⁰⁹ A detector score is not evidence of judgment and not evidence of cheating. The live check is the check.

¹⁰⁸Huber and Kuncel, 2016: college about 0.5–0.6 SD over four years on generic tests; curriculum-wide campaigns did not add a reliable increment.

¹⁰⁹Weixin Liang, Mert Yuksekgonul, Yining Mao, Eric Wu, and James Zou, "GPT Detectors Are Biased against Non-Native English Writers," *Patterns* 4, no. 7 (2023): 100779. Average false-positive rate 61.22 percent on human TOEFL essays. A detector score is not a judgment score.

If you want extra practice items, write them in the subject you already teach. Hide the answers from the student. You may ask a model to generate distractors *for you*. You still choose which case is fair.

What “done enough” looks like

Short checklist before you move on. Placement by skill, not birthday.

- You can say, in one sentence, why a test score is a sitting and not the skill.
- You can hear the five wrong answers above and know what each usually means.
- The student can complete three unseen cases in *one* known subject, with process visible, and explain each with the page closed.
- The student can look at a sample inventory item and a real case and say which object each one is.
- You have one folder, however thin, of those cases dated. That folder is evidence. A commercial percentile, by itself, is not.
- You are not steering next week’s lessons from a subscore with a pretty name.

A “done” student in this chapter is not a student who has passed a Watson–Glaser. A done student can tell a sitting from a judgment and can produce a small trail of actual decisions in a domain they know.

Move on to Chapter 6 when that trail exists. Chapter 6 is the first life domain written as a week of teaching: leave the page. The civic habit has trial evidence. It is still not a badge. The test remains not the thing.

If a co-op, an umbrella school, or a relative asks for a critical-thinking score, you now have a calm answer. We can sit an inventory if a destination requires the sitting. We will not confuse the sitting with the work. The work is in history, science, government, and the cases in this folder.

You can do this. Tomorrow’s sentence is already written: “A score is a sitting. Today we are collecting evidence that you decided.”

Chapter 6

Leave the page

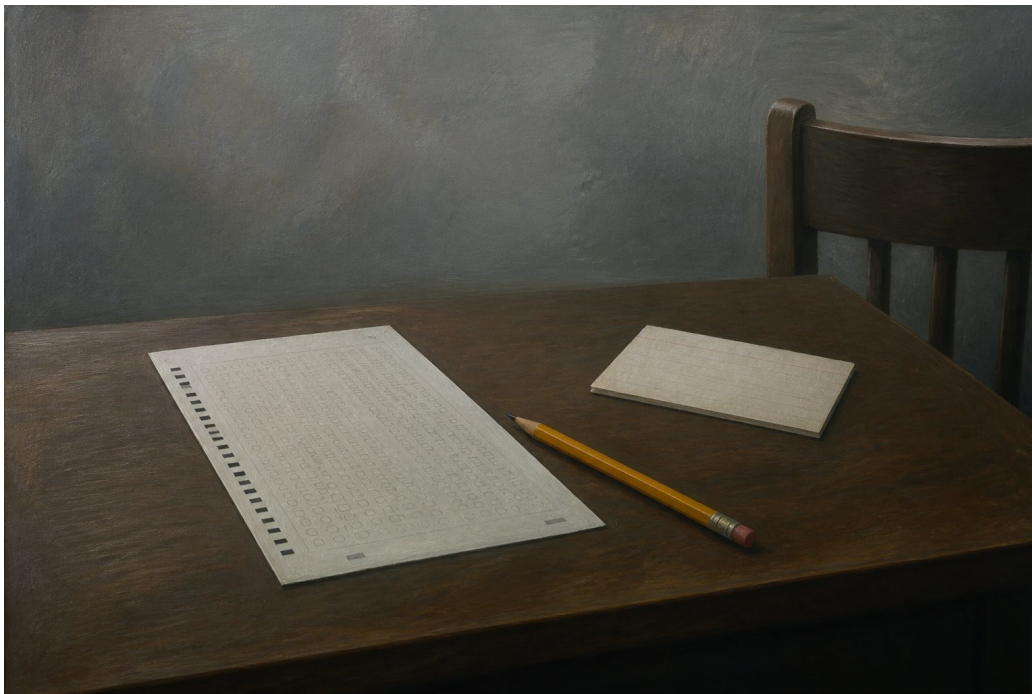


Figure 7: A scantron sheet, a yellow pencil, and a stack of index cards on a dark desk.

Why this matters

When 3,446 high-school students were asked to investigate a climate site that claimed to publish factual reports, 96 percent never found the group’s fossil-fuel ties.¹¹⁰

That is why this week’s work is leaving the page.

The civic habit with classroom trial evidence is not a thirty-item checklist. It is a short departure: open a new tab, search the organization and the people named, then come back knowing who is speaking. Sam Wineburg and Sarah McGrew watched professional fact checkers do that on live websites. Historians and Stanford undergraduates stayed on the page — About pages, .org domains, prose quality, internal consistency — and were slower and less accurate.¹¹¹ Reading carefully was not the skill. Knowing when *not* to read yet was.

Ennis’s sentence includes *do*. Sharing is a doing. Declining to amplify a video is a doing. Voting is a doing. A book that treated thinking as the correct scoring of statements could stop at accuracy judgments. This one cannot. The feed will not wait for a course to finish.

What you can learn this week is the move itself, in language you can say at the table, on real pages, in a subject you already teach — usually government, history, or current events sitting inside those courses. What the student can try today is one unfamiliar page, two new tabs, and three questions spoken out loud.

This is a useful study, not a promise that every home will see the same result. Then teach.

¹¹⁰Joel Breakstone, Mark Smith, Sam Wineburg, Amie Rapaport, Jill Carle, Marshall Garland, and Anna Saavedra, “Students’ Civic Online Reasoning: A National Portrait,” *Educational Researcher* 50, no. 8 (2021): 505–15. 3,446 high-school students; live internet; six constructed-response tasks. Climate site: 96 percent never uncovered fossil-fuel ties. Two thirds missed news versus ads. More than half treated an anonymous Facebook video shot in Russia as strong evidence of U.S. voter fraud. Scores varied by SES, race, maternal education, and free-or-reduced-price lunch.

¹¹¹Sam Wineburg and Sarah McGrew, “Lateral Reading and the Nature of Expertise,” *Teachers College Record* 121 (2019). Ten professional fact checkers, ten Ph.D. historians, twenty-five Stanford undergraduates. Fact checkers left; historians and undergraduates stayed; checkers were faster and more accurate.

For the parent: understand it yourself

You do not need to become a professional fact checker. You do need today's idea well enough to hear "it looks professional, so it is trustworthy" as a stay, not as evaluation.

Everyday picture. A stranger knocks. The porch light is on. The clipboard looks official. The speech is fluent. You do not decide from the clipboard. You look at the badge, you call the number on the city's site — a *different* number than the one on the clipboard — and then you open the door or you do not. Lateral reading is that call. The clipboard is the page.

Precise picture. Civic Online Reasoning, the Stanford History Education Group curriculum that grew out of this work, distills three questions, in the project's own words:

1. Who's behind the information?
2. What's the evidence?
3. What do other sources say?¹¹²

Not "is there an About page," but *leave and find out*. Not "does the page have citations," but *what kind of evidence is this, and does it show what the headline claims*. Not "is the site internally consistent," but *what do independent accounts, including ones that disagree, report*.

Two companion moves sit next to those questions. *Click restraint*: Google's ranking is not a trust ranking; the first result is a commercial and algorithmic fact, not an epistemic one. *Wise Wikipedia use*: the encyclopedia is a map, not a destination. Fact checkers harvest references and read Talk pages, then leave again.¹¹³ Wikipedia as a first hop is a lead generator. It is not "Wikipedia is the source."

The contrast case is a feature checklist. Currency, Relevance, Authority, Accuracy, Purpose — CRAAP — was recommended on college library pages. Students used it. A Stanford History Education Group working paper argued that those "best

¹¹²Stanford History Education Group, Civic Online Reasoning, <https://cor.stanford.edu/>, fetched 30 August 2026. Three questions in the project's own words. Free lessons; the site is not itself an outcome trial.

¹¹³Sarah McGrew, Teresa Ortega, Joel Breakstone, and Sam Wineburg, "The Challenge That's Bigger Than Fake News," *American Educator*, Fall 2017. Click restraint; wise Wikipedia use; checklist critique. Quoted sentence: providing an author, a reference list, and a page free of typos hardly establishes credibility.

practices” can make students *more* susceptible, because they train attention onto features an adversary can fake.¹¹⁴ Anyone can register a .org. Anyone can write an About page. Anyone can list an author, throw up a reference list, and hire a designer. Providing an author, a reference list, and a page free of typos hardly establishes a source as credible.¹¹⁵ A cloaked real site with a real funder will look finished. The student who was taught to inspect look, top-level domain, About pages, and quantity of information will find those signals, stay, and score the costume.

The 2021 national portrait is what that stay produces. Besides the climate-site result that opened this chapter, two thirds of those 3,446 students could not distinguish news stories from advertisements on a popular homepage, and more than half treated an anonymously posted Facebook video, shot in Russia, as “strong evidence” of U.S. voter fraud.¹¹⁶ Three failures, one mechanism: trust in fakeable signals on the page. “Digital native” does not name a skill. Growing up with a phone is not growing up with leaving.

Then there is a trial. Wineburg, Joel Breakstone, McGrew, Mark Smith, and Teresa Ortega tested six lessons of about fifty minutes each, Civic Online Reasoning, in a required high-school government course. Treatment 271 students, matched control 228. Experimental classrooms grew significantly in judging digital credibility.¹¹⁷

¹¹⁴Sam Wineburg, Joel Breakstone, Nadav Ziv, and Mark Smith, *Educating for Misunderstanding*, SHEG working paper A-21322 (2020). Vertical, feature-based checklists (including CRAAP as recommended on library pages) can train attention onto fakeable signals. Working paper, not a journal article.

¹¹⁵Sarah McGrew, Teresa Ortega, Joel Breakstone, and Sam Wineburg, “The Challenge That’s Bigger Than Fake News,” *American Educator*, Fall 2017. Click restraint; wise Wikipedia use; checklist critique. Quoted sentence: providing an author, a reference list, and a page free of typos hardly establishes credibility.

¹¹⁶Joel Breakstone, Mark Smith, Sam Wineburg, Amie Rapaport, Jill Carle, Marshall Garland, and Anna Saavedra, “Students’ Civic Online Reasoning: A National Portrait,” *Educational Researcher* 50, no. 8 (2021): 505–15. 3,446 high-school students; live internet; six constructed-response tasks. Climate site: 96 percent never uncovered fossil-fuel ties. Two thirds missed news versus ads. More than half treated an anonymous Facebook video shot in Russia as strong evidence of U.S. voter fraud. Scores varied by SES, race, maternal education, and free-or-reduced-price lunch.

¹¹⁷Sam Wineburg, Joel Breakstone, Sarah McGrew, Mark D. Smith, and Teresa Ortega, “Lateral Reading on the Open Internet: A District-Wide Field Study in High School Government Classes,” *Journal of Educational Psychology* 114, no. 5 (2022): 893–909. Six ~50-minute COR lessons over three months in a required government course; treatment n = 271, matched control n = 228; experimental classrooms grew significantly on digital-credibility tasks.

University communications summarizing that paper reported that students roughly doubled their pre-test scores and still earned only about half the possible points.¹¹⁸ Six hours, in a course that already had a civic job, moved a test of the same habit. The remaining miss rate is the story as much as the doubling. Steal the object of practice. A credibility-task gain is not informed citizens, higher turnout, or less polarization.

Willingham still holds. You cannot laterally read a climate-advocacy site if you have no idea what a fossil-fuel trade association is. You cannot evaluate a voter-fraud claim if you do not know what would count as evidence of it.¹¹⁹ The three questions are cues. Doing what they call for depends on civic knowledge and on practice with real pages. That is why the trial lived in a government course, and why this week lives inside a subject, not in a free-floating “media hour” with empty content.

A few neighboring tools, named so they cannot collapse into this week’s move.

Technique inoculation — short videos or games that expose emotionally manipulative language, false dichotomies, scapegoating, ad hominem — raises *technique recognition* in the lab and more weakly on YouTube, and it decays without boosters.¹²⁰ That is resistance to a named move, not page evaluation. An accuracy prompt before sharing can raise the quality of subsequent shares by putting accuracy in working memory; it is not a belief vaccine, and whether it works the same across partisan groups is a live dispute.¹²¹ Corrections are still worth making. Large later experiments found worldview backfire elusive; the surviving problem

¹¹⁸Stanford Graduate School of Education news, 19 April 2022, summarizing Wineburg et al. 2022: roughly doubled pre-test scores; still about half the possible points. University communications, not the paper. Funded by Google.org.

¹¹⁹Daniel T. Willingham, “Critical Thinking: Why Is It So Hard to Teach?,” *American Educator* 31, no. 2 (Summer 2007); Willingham, “How to Teach Critical Thinking,” NSW Department of Education, 2019 (identify the meaning in each domain; identify needed content; sequence; plan three to five years of revisiting).

¹²⁰Jon Roozenbeek, Sander van der Linden, Beth Goldberg, Steve Rathje, and Stephan Lewandowsky, *Science Advances* (2022): lab technique-recognition d 0.28–0.68; YouTube field h = 0.09. Rakoén Maertens, Roozenbeek, Melisa Basol, and van der Linden, *Journal of Experimental Psychology: Applied* (2021): Bad News game effects decay to nonsignificance over about two months without regular testing.

¹²¹Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio Arechar, Dean Eckles, and David G. Rand, *Nature* (2021): accuracy prompt raises quality of subsequent shares; sharing discernment, not belief. Partisan moderation remains disputed (Rathje et al. letter; 2024 adversarial collaboration).

is continued influence — a myth that still fills a causal slot after a person has nodded at the correction.¹²² Offer a true cause, not a bare “that’s false.” Twenty-five state media-literacy laws are laws. A professional association’s principles are a framework. Classrooms under those laws have not been evaluated as a set.¹²³ A statute that says the words is not this week’s three questions on a live page.

Wrong answers you should be able to hear

1. “*It’s a .org, so it’s nonprofit, so it’s fair.*” Anyone can register the domain. Leave. Search the organization.
2. “*The About page names an author and lists references, so it’s solid.*” Those are cheap signals. The question is who is behind the organization, checked off the page.
3. “*The first Google result is the best source.*” Ranking is not a trust ranking. Click restraint: scan, then choose.
4. “*It’s a video, so we can see it happened.*” A video is a clip. Anonymous posting is not a named witness. What would count as evidence of the *claim*?
5. “*Wikipedia is banned, so we have to trust the site’s self-description.*” Wikipedia as a first hop is a map. Harvest the references. Then leave again.

Five-minute parent warm-up

Do this before the lesson, on a scrap of paper, no student in the room. Use a real unfamiliar page — a cause, a health claim, a local organization you do not already know.

Minute 1. Open the page. Do *not* read it yet. Write the organization’s name.

Minute 2. New tab. Search the name plus a word such as “funding,” “about,” or “Wikipedia.” Write one sentence: who is behind this?

Minute 3. Back to the original claim. What is the evidence on the page, in one line? What kind of evidence would actually show it?

¹²²Thomas Wood and Ethan Porter, *Political Behavior* (2019): five experiments, more than 10,000 participants, 52 issues; no item triggered a worldview backfire. *The Debunking Handbook 2020*: do not skip correction for fear of backfire; continued influence is the surviving problem; offer a causal alternative.

¹²³Media Literacy Now policy map (updated January 2026): 25 states with media-literacy laws; MLN has not evaluated classroom implementation. NAMLE Core Principles (2023) are a professional association’s framework, not a trial.

Minute 4. Third tab. What do two independent sources say? They may disagree. That is allowed. Write the disagreement.

Minute 5. The sentence you will say tomorrow: “Leave the page before you trust the page.” Put the pencil down.

If you can do those five minutes, you are ready. The student leaves. You hear whether they left.

How to teach it this week

Same session shape. The object is a live page, not a worksheet about a cartoon hoax. The Pacific Northwest tree octopus is a useful joke and a weak civic education. Cloaked real sites with real funders are the live web’s problem.¹²⁴

Warm-up (3–5 minutes, unaided). Two civic facts the student already knows: what a trade association is, what an advertisement is doing on a news homepage, what a primary source is. One tiny prompt: “If you needed to know who funds a group, would you start on their site or off it?” Paper first. Then the browser.

Short model (5 minutes). You project or sit beside one unfamiliar page. You do not read the article. You say the three questions. You open a new tab. You search. You come back with one sentence about who is behind it. Then you stop talking.

Student attempt (10–15 minutes). One live page you have already previewed so the sitting is teachable, not a trap. The student leaves, writes who / evidence / other sources, and does not start with the About page. You wait. Struggle before rescue.

One good question. “Where did you look *second*?” If the answer is “I kept reading,” they stayed. If the answer is a new tab, they left.

Mixed practice (5–10 minutes). Yesterday’s history document next to today’s live page. A homeschool science claim next to a homepage that mixes news and ads. The mix trains *when* to leave, not only how.

¹²⁴Sarah McGrew, Teresa Ortega, Joel Breakstone, and Sam Wineburg, “The Challenge That’s Bigger Than Fake News,” *American Educator*, Fall 2017. Click restraint; wise Wikipedia use; checklist critique. Quoted sentence: providing an author, a reference list, and a page free of typos hardly establishes credibility.

Exit ticket (5 minutes, unaided). A new unfamiliar URL, or a screenshot plus permission to search. Three written lines: who, evidence, other sources. Then two spoken sentences with the original page closed.

Exact wording you can say

“Leave the page before you trust the page.”

“Who’s behind this — not ‘is there an About page,’ but leave and find out.”

“What’s the evidence — not ‘are there citations,’ but does this evidence show the claim?”

“What do other sources say, including ones that disagree?”

“Google’s first result is a ranking, not a medal.”

“Wikipedia is a map. Harvest, then leave.”

“Looking finished is cheap. Checking is the work.”

When you are about to take over:

“Your tabs. I’ll ask who you found.”

Then wait.

First problem for the student

A live page, previewed by you, on a topic already in this year’s government or history work. Label the sitting as practice, not as a reported family.

Say: “Leave before you read. Write the organization’s name. Open a new tab. Who’s behind this? Come back and tell me before you read further.”

Wait. If they scroll the original, they stayed. If they search the name, they left. If they freeze, model a *different* page fully — you search, you read a second source, you return — then hand theirs back.

Later the same week, the diagnostic: a homepage with news and ads mixed, or a video posted without a named account. Hear “it looks like news.” Hear “it’s a video, so it’s strong.” Hear “I don’t know who is behind this yet, so I have not evaluated it.” After they attempt, you may show a worked incorrect stay: a composite student who scored the About page, the .org, and the reference list, and never found the funder. They name the error.

How to fade help

Day one: you model the whole departure on a different page. Day two: they search; you sit silent unless they return to vertical reading. Day three: they run all three questions. When they leave without being told, stop narrating the tabs. Bring the model back when the domain changes — health sites in Chapter 7, a fluent machine paragraph in Chapter 8 — because the page will look different and the departure is the same.

When to stop talking

After “who’s behind this?” After the hint “new tab.” After modeling a different URL. If you are reading the original article aloud while they watch, you have started staying for them. Stop. Point at the address bar. Wait.

Government class was the trial’s home for a reason you already know from Chapter 2: the habit needs civic knowledge. A week of empty “media literacy” with no history, no government, no science will produce the maxim without the knowledge. Teach this week inside a subject. Name the move out loud. That is mixed design in kitchen form.

Practice that actually builds learning

Blocked for the new move. Mixed for when to use it. Retrieval of civic facts. One incorrect stay to diagnose.

Blocked. Three live pages, same week, same three questions. All in one domain you already teach: three organizations in a civics unit, or three claims about one historical event’s afterlife online. The student is learning to leave, not to collect trivia.

Mixed. A primary document from the history book (source on the page, because the book’s apparatus is the apparatus) next to a live advocacy site (leave). A peer-reviewed abstract next to a supplement ad. A poll headline next to the poll’s methods page, found in a new tab. The student has to notice which object needs departure.

Retrieval. Two minutes, before the browser: what is a trade association, a think tank, a news article versus an advertorial, a primary versus a secondary source, a sample versus a population. Without those names, “who’s behind this” has

nowhere to land.

One incorrect example. After they have attempted, a worked stay: professional layout, .org, About page, dense references, all on a cloaked site. Ask what the composite student trusted, and what a second tab would have shown. They name the fakeable signals. You do not lecture the 2021 portrait again. One sentence was enough.

Wineburg 2022 is six lessons over three months, teachers with professional development, live internet, constructed-response tasks.¹²⁵ You are not copying a district. You are copying the object: short, repeated departures on real pages, inside a civic subject. Spread three or four sittings across the week rather than one long assembly. McGrew’s smaller eight-lesson study found gains on investigating backers, critiquing social-media evidence, and locating reliable information on a contentious question.¹²⁶ Small relative to the district trial; still a reason to practice those three tasks, not a reason to declare the civic job done.

Inoculation videos can sit in a Friday slot as technique recognition, with a later booster, if you want a second object. They are not a substitute for leaving. An accuracy question before anyone in the house shares — “do you think this is true?” — is a ten-second family prompt about sharing, not about belief.¹²⁷ Keep the objects separate. This week’s practice is the tabs.

A hoax-site lesson trains a laugh. Retire it for now. The live web rewards the adversary who invests in looking finished.

News versus ads, as its own sitting. Two thirds of the national sample missed the costume of native advertising.¹²⁸ One blocked sitting this week can be a single

¹²⁵Sam Wineburg, Joel Breakstone, Sarah McGrew, Mark D. Smith, and Teresa Ortega, “Lateral Reading on the Open Internet: A District-Wide Field Study in High School Government Classes,” *Journal of Educational Psychology* 114, no. 5 (2022): 893–909. Six ~50-minute COR lessons over three months in a required government course; treatment n = 271, matched control n = 228; experimental classrooms grew significantly on digital-credibility tasks.

¹²⁶Sarah McGrew, “Learning to Evaluate,” *Computers & Education* 145 (2020): eight-lesson COR intervention; gains on three of four online tasks.

¹²⁷Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio Arechar, Dean Eckles, and David G. Rand, *Nature* (2021): accuracy prompt raises quality of subsequent shares; sharing discernment, not belief. Partisan moderation remains disputed (Rathje et al. letter; 2024 adversarial collaboration).

¹²⁸Joel Breakstone, Mark Smith, Sam Wineburg, Amie Rapaport, Jill Carle, Marshall Garland, and Anna Saavedra, “Students’ Civic Online Reasoning: A National Portrait,” *Educational Researcher* 50, no. 8 (2021): 505–15. 3,446 high-school students; live internet; six constructed-

homepage: circle three items, label each news, opinion, or ad, then leave the page to check one label you were unsure about. Exact wording: “What kind of thing is this, before we ask whether it is true?” Genre first. Truth second. A student who cannot tell an advertorial from a report cannot yet run “what’s the evidence?” on it, because they have not named the speaker’s job.

Click restraint, as a thirty-second habit. Before the first click on a search page, read the results list out loud: who seems to be behind each hit, as far as the title and the address reveal. Then choose. The first hit is allowed to lose. This is not a new theory. It is how you keep Google’s ranking from becoming the evaluation.

If a sitting produces only “I still don’t know,” that is a legal exit ticket. Not knowing who is behind a page, and saying so, is already better than scoring the costume. Write “unknown — need another tab” as an answer you will accept. The feed rewards a confident stay. You are training an honest departure.

For the student

You already know how to read. This week is about when *not* to start reading yet.

A page can look finished — clean design, a serious domain name, an About page, lots of words and citations — and still be a costume. The move is to leave before the costume has time to feel true.

Three questions, in this order:

1. Who’s behind this?
2. What’s the evidence?
3. What do other sources say?

Tiny worked example. A site says it “disseminates factual reports” on climate. Looks like a public-interest group. Instead of reading the report, you open a new tab and search the group’s name. A second source says a fuel-industry association helps fund it. Now you know who is speaking. Then you ask what evidence the report actually offers, and what independent scientists say. You may still read the original. You read it as a document with a speaker, not as the weather.

response tasks. Climate site: 96 percent never uncovered fossil-fuel ties. Two thirds missed news versus ads. More than half treated an anonymous Facebook video shot in Russia as strong evidence of U.S. voter fraud. Scores varied by SES, race, maternal education, and free-or-reduced-price lunch.

Two tries.

1. Your parent or teacher will give you an unfamiliar URL in a subject you already study. Leave before you read. New tab. Who's behind it? Write one sentence.
2. Same sitting or the next day: what is the evidence for the main claim, and what do two other sources say? They are allowed to disagree. Write the disagreement in one line.

Explain it back. Why is leaving the page a different move from reading the About page carefully? Say it without this chapter open.

Challenge. Tonight, before anyone in the house shares an article or a video, run the three questions on that one item. If you cannot answer “who's behind this” without the original page, you are not ready to share. Sharing is a doing. Ennis's sentence includes do.

You are allowed to be slow. Fact checkers in the study were faster *because* they left. Staying to read carefully was the slow, less accurate path. Speed will come from practice, not from skimming.

If it isn't clicking

1. They stay. They read. They score the costume. Next move: take the original article away for the first three minutes. Only the organization name, on paper. New tab from the name alone. When leaving happens without the costume on screen, bring the page back.

2. They leave, but they do not know what they are looking at. Next move: knowledge, not more maxims. A twenty-minute content burst — what a trade association is, what native advertising is, what a primary document is — then one page. A prompt cannot supply civic vocabulary.

3. They treat Wikipedia as either forbidden or final. Next move: a scripted first hop. Open Wikipedia (or a search that lands there). Read the first paragraph and the references. Open one reference. Close Wikipedia. That is the map-not-destination sitting. Practice it twice.

When to slow down: the student cannot yet tell a news story from an ad on a homepage, even after you name the label. Stay on genre for a sitting. Wineburg's

portrait found that failure in two thirds of the national sample.¹²⁹ It is common. It is teachable as “what is this kind of thing?”

When to go ahead: they leave without being told, they can answer the three questions on two new pages in a known subject, and they can explain why .org and an About page are cheap signals.

When to get a human tutor: the bottleneck is reading, not evaluation — decoding the search results is the wall. A reading teacher, or a librarian who will sit one hour and run lateral reading as a search skill, is the right valve. A “fake news” assembly is not.

Equity belongs here as a reason to teach, not as a verdict on the house. Breakstone and colleagues found that civic-online-reasoning scores already varied by socioeconomic status, race, maternal education, and free-or-reduced-price lunch — before chatbot news.¹³⁰ Growing up online is not the practice. The practice is this week’s tabs, inside a knowledge-rich subject. A poster on the wall will not close that gap. Your Tuesday might.

Older relatives share too. One 2016 Facebook study found sharing from fake-news domains was rare overall, and users over sixty-five shared nearly seven times as many such articles as the youngest group; that is a sharing study, not a finding that aging ruins thought.¹³¹ The habit does not retire at commencement. If an adult in the house will run one page with the student, the student has a second mentor. If

¹²⁹Joel Breakstone, Mark Smith, Sam Wineburg, Amie Rapaport, Jill Carle, Marshall Garland, and Anna Saavedra, “Students’ Civic Online Reasoning: A National Portrait,” *Educational Researcher* 50, no. 8 (2021): 505–15. 3,446 high-school students; live internet; six constructed-response tasks. Climate site: 96 percent never uncovered fossil-fuel ties. Two thirds missed news versus ads. More than half treated an anonymous Facebook video shot in Russia as strong evidence of U.S. voter fraud. Scores varied by SES, race, maternal education, and free-or-reduced-price lunch.

¹³⁰Joel Breakstone, Mark Smith, Sam Wineburg, Amie Rapaport, Jill Carle, Marshall Garland, and Anna Saavedra, “Students’ Civic Online Reasoning: A National Portrait,” *Educational Researcher* 50, no. 8 (2021): 505–15. 3,446 high-school students; live internet; six constructed-response tasks. Climate site: 96 percent never uncovered fossil-fuel ties. Two thirds missed news versus ads. More than half treated an anonymous Facebook video shot in Russia as strong evidence of U.S. voter fraud. Scores varied by SES, race, maternal education, and free-or-reduced-price lunch.

¹³¹Andrew Guess, Jonathan Nagler, and Joshua Tucker, *Science Advances* (2019): over-65s shared nearly seven times as many fake-news-domain articles as the youngest group, 2016 Facebook data, holding after partisanship. Nadia Brashier and Daniel Schacter (2020): do not reduce this to cognitive decline.

not, the student still has the three questions.

Tools, including AI

Optional helpers for the adult. Student attempts first. You hold the key — including which URL you chose and what a decent “who’s behind this” looks like for that page.

A box of rules

- Child attempts first, on a live page you previewed.
- Parent holds the key.
- Use a language model to explain lateral reading to *you*, to help you preview three URLs and hide a one-sentence “who” until after the attempt, or to write a hint script (“search the organization name plus ‘funding’”).
- During the attempt: no “is this site trustworthy?” pasted into a chat, no unsupervised chatbot evaluating the page in the student’s place, no photo of the article sent for a verdict.
- Wikipedia, a search engine, and a second news outlet are the tools of the move. They are not cheating. They *are* the departure.

A model can generate a fluent “this .org is a respected nonprofit” paragraph. That paragraph is another page. Leave it too. Chapter 8 will spend the machine that finishes the sentence. This week, the rule is the same: the finished-looking evaluation is not the evaluation.

A checklist app that scores currency and typos is the contrast case in software form. Skip it for now.

If you want published lessons rather than URLs you pick, Civic Online Reasoning offers free classroom tasks built around the three questions.¹³² Fit: government and history, live internet, constructed response. Parent load is real on previewing and on sitting beside the first two attempts; lighter once the student leaves without a nudge. Those lessons are a curriculum, not a trial you have replicated. Wineburg 2022 is the trial. The site is the practice object.

¹³²Stanford History Education Group, Civic Online Reasoning, <https://cor.stanford.edu/>, fetched 30 August 2026. Three questions in the project’s own words. Free lessons; the site is not itself an outcome trial.

What “done enough” looks like

- You can say the three questions without looking them up.
- You can hear the five wrong answers and know that each is a stay.
- The student, on two new pages in a subject they know, leaves before reading, writes who / evidence / other sources, and explains the decision with the original page closed.
- The student can tell a news story from an ad on at least one homepage sitting, or can say “I cannot tell yet” instead of guessing from layout.
- .org, About page, and “lots of references” are no longer treated as proof.
- You have dated notes on those two pages. That is evidence of a civic habit. It is not a media-literacy badge, and it is not a promise about votes.

Spread the next months. Willingham’s advice for domain-specific thinking is to identify what the skill means in the domain, identify the content it needs, sequence, and plan years of revisiting.¹³³ Six lessons doubled a still-failing score. Revisiting is the honest plan. Put a live page inside government or history once a week for a while. Boosters matter in the inoculation studies; they matter here as ordinary practice.

Move on to Chapter 7 when the departure happens without a lecture. Chapter 7 is stakes: health and work as domains, with numbers and stories, not as proof of a transferable badge. The page is still there to leave when a clinic pamphlet or a job post looks finished.

You can do this. Tomorrow’s sentence is already written: “Leave the page before you trust the page.”

¹³³Daniel T. Willingham, “Critical Thinking: Why Is It So Hard to Teach?,” *American Educator* 31, no. 2 (Summer 2007); Willingham, “How to Teach Critical Thinking,” NSW Department of Education, 2019 (identify the meaning in each domain; identify needed content; sequence; plan three to five years of revisiting).

Chapter 7

Stakes



Figure 8: Open tabbed books, a stack of newspapers, and eyeglasses under a banker's lamp.

Why this matters

Judgment in a life does not look like an untimed worksheet. Time is short. Information is incomplete. Other people are in the room. Values are the family's, or the patient's, not an answer key. A body, a hire, or a paycheck can be on the line.

This chapter treats health and work as *domains*, each with named knowledge, a representation that makes a move executable, and real error-costs. It does not treat them as proof that a transferable badge has been earned. Willingham's constraint travels: you cannot implement "consider the base rate" without the knowledge and the format.¹³⁴ Ennis's sentence travels: the decision is what to believe *or do*. In a clinic the doing is consent, refusal, a second question asked before a fluent label sticks. At work the doing is a situated story — a time you anticipated a problem — not a percentile on an inventory.

Two pictures open the week, so the error-costs are visible before the arithmetic.

In 1978, Ward Casscells, Arno Schoenberger, and Thomas Graboys asked Harvard Medical School faculty, staff, and students: a disease has prevalence of one in a thousand; a test has a false-positive rate of five percent; you have no other information; what is the chance that a person with a positive result has the disease? Twenty-seven of sixty said 95 percent. The correct answer is about 2 percent. Eleven of sixty gave it.¹³⁵ Prestige did not install the base rate.

The National Academies, in 2015, put the same structure on a life: most people will experience at least one diagnostic error; about 5 percent of U.S. adults seeking outpatient care each year; communication of the explanation sits inside the definition of error.¹³⁶ A 2023 modeling estimate of 795,000 Americans permanently disabled or dead annually from misdiagnosis is a model with a published range

¹³⁴Daniel T. Willingham, "Critical Thinking: Why Is It So Hard to Teach?," *American Educator* 31, no. 2 (Summer 2007).

¹³⁵Ward Casscells, Arno Schoenberger, and Thomas B. Graboys, "Interpretation by Physicians of Clinical Laboratory Results," *New England Journal of Medicine* 299 (1978): 999–1001. Harvard Medical School faculty, staff, and students; prevalence 1/1,000; false-positive 5 percent; 27 of 60 said 95 percent; 11 of 60 gave about 2 percent. *AMA Journal of Ethics* (2009) restates the item next to shared decision-making.

¹³⁶National Academies of Sciences, Engineering, and Medicine, *Improving Diagnosis in Health Care* (2015). Diagnostic error as failure to establish an accurate and timely explanation *or* to communicate it; about 5 percent of U.S. adults seeking outpatient care each year; postmortem work about 10 percent of deaths; 6–17 percent of hospital adverse events; most people at least one diagnostic error. Synthesis, not an RCT.

(598,000 to 1,023,000), not a census.¹³⁷ Print the scale. Keep the word *model*. None of those numbers is a trial of a critical-thinking course.

What you can learn this week is how to recode a frightening percent as a count, and how to hear a hiring story as domain practice rather than as a test of a faculty. What the student can try today is one natural-frequency tree on paper, and one “describe a time” paragraph in a subject they actually know.

You do not need to be a physician or a recruiter. You do need today’s idea well enough to hear “25 percent reduction” as a frame that still needs a count, and “we hire for critical thinking” as a survey sentence until someone asks for a story.

For the parent: understand it yourself

Many adults feel rusty around percentages in a clinic pamphlet and around interview talk. That is ordinary. Five minutes of this section, then the warm-up, is enough for tomorrow.

Everyday picture. A poster in a waiting room: “screening cuts deaths by 25 percent.” A quarter gone sounds like a lot. Ask for the count. In the mammography example Gerd Gigerenzer and colleagues made famous, that “25 percent reduction” is 1 in 1,000 in absolute risk.¹³⁸ One extra person in a thousand. Both sentences can be true. Only one of them is a number a person can prefer *with*. Shared decision-making without that number is a fluent brochure.

Precise picture. Imagine 1,000 people. One has the disease (the base rate in the Casscells item). The test’s false-positive rate is 5 percent, so about 50 of the 999 well people test positive. One true positive, about 50 false positives. Of the people who test positive, about 1 in 51 has the disease — about 2 percent. You can read the posterior off the counts. You do not have to remember Bayes’ theorem as a formula. You have to keep the 1,000, apply the base rate, apply the false-positive rate, and not invert. Ulrich Hoffrage and Gigerenzer showed that physicians’ diagnostic

¹³⁷David E. Newman-Toker et al., “Burden of Serious Harms from Diagnostic Error in the USA,” *BMJ Quality & Safety* (2023). Modeling estimate 795,000 serious harms per year; plausible range 598,000–1,023,000. Not a census.

¹³⁸Gerd Gigerenzer, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa M. Schwartz, and Steven Woloshin, “Helping Doctors and Patients Make Sense of Health Statistics,” *Psychological Science in the Public Interest* 8, no. 2 (2008): 53–96. Statistical illiteracy in patients, journalists, and physicians; mammography “25 percent reduction” as 1 in 1,000 absolute.

inferences improve when the same information is recoded as counts rather than as percentages and conditionals.¹³⁹ Named knowledge (what prevalence and false-positive *mean*) plus a format (counts in a population) plus an error-cost (a person about to be told they have a rare disease). That is this book's thesis in one object. It is not a transferable faculty.

The fluent wrong answer on Casscells is: the test is wrong 5 percent of the time, so it is right 95 percent of the time. That computation answers a different question. It answers how often a well person tests positive, inverted and subtracted, as if it were the chance of disease given a positive. William Thompson and Edward Schumann named the same inversion in court as the prosecutor's fallacy: treating the chance of the evidence given innocence as the chance of innocence given the evidence.¹⁴⁰ A DNA profile that would match 1 in 10,000 innocent people is not, by itself, a 1-in-10,000 chance of innocence. You still need the base rate. Clinic and court share the inversion. They do not share the value. In criminal law the burden is a designed error-cost: the state must prove; the cost of convicting the innocent has been set higher than the cost of acquitting the guilty. In a clinic the patient's goals decide which error is worse. Teaching both as one "base-rate skill" would erase the design. Teaching both as one representation serving different error-costs is the structure.

Geoffrey Norman and colleagues, from inside diagnosis, already said the binding constraint is access to relevant knowledge, not a missing general trait. Lots of cases, because of case specificity. The fluent premature label is the failure mode.¹⁴¹ A maxim "avoid bias" will not supply the cases. Natural-frequency training helps inferences on lab-like tasks. Whether it reduces real diagnostic error or improves shared-decision-making at scale was not shown in the sources this book stands on. Keep the week honest: you are teaching a recoding a family can use on a pamphlet, not a population-health result.

Glyn Elwyn and colleagues offered a three-step clinic sequence: choice talk (a choice exists, and preferences matter), option talk (name the options, including

¹³⁹Ulrich Hoffrage and Gerd Gigerenzer, "Using Natural Frequencies to Improve Diagnostic Inferences," *Academic Medicine* 73, no. 5 (1998): 538–40. Gigerenzer and Hoffrage, *Psychological Review* 102 (1995): frequency formats improve Bayesian reasoning without instruction.

¹⁴⁰William C. Thompson and Edward L. Schumann, "Interpretation of Statistical Evidence in Criminal Trials," *Law and Human Behavior* 11 (1987): 167–87. Prosecutor's fallacy: $P(\text{evidence} \mid \text{innocence})$ treated as $P(\text{innocence} \mid \text{evidence})$. Teaching object, not a courtroom RCT.

¹⁴¹Geoffrey Norman et al., *Academic Medicine* 92 (2017): 23–30. Diagnostic error as knowledge access; many cases because of case specificity.

doing nothing, with benefits and harms a person can use), decision talk (arrive at a decision that fits informed preferences).¹⁴² It is a practice model, not a civic-outcome trial. Option talk is where the count has to live. Relative risk standing alone is not a usable option. A patient who cannot recode a screening number cannot share the decision. Skepticism without the count is how a person refuses a useful screen or accepts a useless one, depending on which fluent sentence they heard last.

Now work, as a second domain, not as a sequel of the same badge.

Employers say they want critical thinking. What they test, when they test, is a story in an interview. The National Association of Colleges and Employers, *Job Outlook 2025* — 237 respondents, 162 members equal to 19.2 percent of eligible members — put communication at 4.57 and critical thinking at 4.49 on a five-point importance scale. Teamwork 4.43. Share rating critical thinking very or extremely important: 96.1 percent. Share rating graduates very or extremely proficient: 53.5 percent.¹⁴³ NACE's own definition is not a Facione skill list: identify and respond to needs based on situational context and logical analysis of relevant information. Sample interview item: describe a time when you anticipated a problem.¹⁴⁴ That is a situated story. It is closer to knowledge plus practice in a domain than to a generic course. A couple of hundred recruiters, not a probability sample of all employers. A Likert rating, not a skill assessment.

The World Economic Forum's *Future of Jobs 2025* puts *analytical thinking* at seven of ten companies as an essential core skill.¹⁴⁵ A survey of WEF respondents. The phrase is not "critical thinking." Unsourced "number-one skill employers want" is folklore. Sourced to NACE or WEF, it is still a recruiter survey, and not evidence that a stand-alone thinking course produces job performance.

If you are hiring, or helping a student be hired, ask for a story or a work sample in

¹⁴²Glyn Elwyn, Dominick Frosch, Richard Thomson, et al., "Shared Decision Making: A Model for Clinical Practice," *Journal of General Internal Medicine* 27 (2012): 1361–67. Choice talk / option talk / decision talk. Practice model, not a population-health RCT.

¹⁴³National Association of Colleges and Employers, *Job Outlook 2025* (revised January 2025). n = 237; communication 4.57, critical thinking 4.49, teamwork 4.43 on a 1–5 importance scale; CT proficiency 3.67; 96.1 percent very/extremely important versus 53.5 percent very/extremely proficient graduates. Survey of recruiters, not a skill assessment.

¹⁴⁴NACE, "Career Readiness Competencies: Critical Thinking" (December 2025). Sample interview item: describe a time when you anticipated a problem.

¹⁴⁵World Economic Forum, *The Future of Jobs Report 2025*. Analytical thinking as top core skill, seven of ten companies; employer survey; phrase is *analytical thinking*.

the domain of the job. If you are designing a year to satisfy a relative who wants “critical thinking on the transcript,” teach the domain and the evaluation habits the domain requires. Chapter 9 will refuse “Critical Thinking 101” as a magic course title. This chapter supplies the reason: work is a domain.

A neighboring hygiene, stealable when a domain has a score: start from a base rate, update, keep score. Superforecasting research described those practices in geopolitical tournaments.¹⁴⁶ Steal the hygiene. A forecasting tournament is not a hiring trial, and not a bicycle.

Life versus lab, in one paragraph so the week does not pretend clinic pamphlets are worksheets. Gary Klein’s recognition-primed picture of experts: they often do not generate three options and score them; they recognize a type, simulate the first plausible action, and modify or reject it.¹⁴⁷ Daniel Kahneman and Klein, in a joint statement, agreed on the boundary: intuition is trustworthy when the environment has stable cues and usable feedback, and the person has learned those cues.¹⁴⁸ Chess can be such an environment. Some medicine can. Geopolitics often cannot; Philip Tetlock showed that prestige is not calibration.¹⁴⁹ A curriculum that only trains untimed reflection trains a mode people will not have in a triage bay or a hiring meeting. Natural frequencies are not a general slowdown. They are a format that makes a Bayesian move executable when the fluent answer is 95 percent and the right count is 2 in 100.

Wrong answers you should be able to hear

1. “*The test is 95 percent accurate, so I am 95 percent sure I have it.*” Inversion. Ask for 1,000 people, the base rate, and the false positives.
2. “*Twenty-five percent reduction — that’s a lot.*” Relative risk without the absolute count. Ask “1 in how many?”
3. “*The doctor said so, so the number is already evaluated.*” Casscells was Harvard Medical School. Prestige does not install the recoding. A second question is allowed.

¹⁴⁶Barbara Mellers, Lyle Ungar, Jonathan Baron, et al., *Psychological Science* 25 (2014): 1106–15. Superforecasting hygiene: base rates, updating, outside view, scorekeeping. Not a workplace critical-thinking RCT.

¹⁴⁷Gary Klein, recognition-primed decision model; *Sources of Power* (1998; 2nd ed. 2008).

¹⁴⁸Daniel Kahneman and Gary Klein, “Conditions for Intuitive Expertise: A Failure to Disagree,” *American Psychologist* 64, no. 6 (2009): 515–26.

¹⁴⁹Philip E. Tetlock, *Expert Political Judgment* (Princeton: Princeton University Press, 2005).

4. *“Critical thinking is the number-one skill employers want, so we should take a thinking course.”* NACE 2025 ranks communication slightly above it; they ask a story. Teach the domain. Prepare the story.
5. *“I have a high thinking-test score, so I am ready for the clinic pamphlet and the interview.”* Chapter 5: a sitting is not the skill. Health and work are domains. The tree and the story are the week’s objects.

Five-minute parent warm-up

Do this before the lesson, on paper, no student in the room.

Minute 1. Write 1,000 people. Mark 1 sick. Mark about 50 false positives. Say out loud: of the people who test positive, about 1 in 51 is sick. About 2 percent.

Minute 2. Translate “25 percent reduction” into a count you make up for practice, labelled as an illustration: 4 in 1,000 become 3 in 1,000. One person. Both frames. Which one would you want on a pamphlet.

Minute 3. Write NACE’s sample prompt: “Describe a time when you anticipated a problem.” Answer it yourself, in four sentences, about a domain you actually know this week — a household repair, a lab, a budget. Notice that the answer is knowledge plus a story, not a slogan.

Minute 4. One sentence on error-costs: in this clinic number, which error is worse *for the person who will live with it*? You do not have to decide. You have to notice that the value is not given by the percent.

Minute 5. The sentence you will say tomorrow: “Give me the count in a thousand, then tell me the story.” Put the pencil down.

If you can do those five minutes, you are ready. The student recodes. You hear the inversion.

How to teach it this week

Same session shape. Two objects, not one: a tree of counts, and a situated story. Health one day. Work another. Mixed by Friday.

Warm-up (3–5 minutes, unaided). One already-known science fact (what a false positive is, if you taught it; if not, teach it as content first). One already-known workplace or project fact from the student’s year. No device.

Short model (5 minutes). Draw 1,000 as a box, or as ten rows of 100 marks if the student is young enough that marks help. Place the base rate. Place the false positives. Read the posterior off the paper. Then stop. On a later day, model a four-sentence “time I anticipated a problem” from a domain you both know, labelled as an illustration.

Student attempt (10–15 minutes). One recoding. Numbers you have already checked. Then, another sitting, one story in a domain they know: science fair, a budget, a lab, a family project. You wait. Struggle before rescue.

One good question. “Which question did that 95 percent answer?” “What was the base rate doing in the tree?” “In your story, what did you know, and what did you notice in time?”

Mixed practice (5–10 minutes). A relative-risk headline next to a Casscells-style item next to a “describe a time” prompt. The student names the object: count, inversion, or story.

Exit ticket (5 minutes, unaided). Recode one percent into a thousand. Tell one four-sentence story with notes closed.

Exact wording you can say

“Imagine a thousand people. How many have it before anyone is tested?”

“How many well people will test positive anyway?”

“Of the people who test positive, how many actually have it? Read it off the counts.”

“Twenty-five percent of what? 1 in how many?”

“Describe a time you anticipated a problem. What did you know? What did you do?”

“A survey of recruiters is a survey. The story is the work sample.”

When you are about to take over:

“Your tree. I’ll ask where the fifty came from.”

Then wait.

First problem for the student

Write on a card, labelled as an illustration:

A disease in 1 of 1,000 people. A test that wrongly flags 5 in 100 well people. A person tests positive. About what chance they have the disease?

Say: “Draw a thousand. Keep the base rate. Read the answer off the paper.”

Wait. If they write 95 percent, they inverted. If they write about 2 percent, they kept the base rate. If they stare, model a *different* set of numbers fully — 1 in 100, false-positive 10 percent — then return to this card.

Later the same week: “Describe a time when you anticipated a problem in [this year’s science or history project].” Hear a slogan (“I used critical thinking”). Hear a story with a context, information, and an action. After they attempt, a worked incorrect inversion: a composite student who heard “false-positive 5 percent” and wrote 95. They name the swapped question.

How to fade help

Day of the tree: you draw one fully. They fill a second. A third is missing only the last box. When they can recode an unseen percent into a thousand, stop drawing first. Bring the tree back when a new pamphlet arrives. For the story: they write from a list of four prompts (context, what I knew, what I noticed, what I did). Fade the list when the four appear without it.

When to stop talking

After “draw a thousand.” After “which question did 95 percent answer?” After modeling a different story. If you are filling the tree while they watch, you have started the inversion’s opposite error: doing the recoding for them. Stop. Point at the empty thousand. Wait.

Law can sit as an optional third sitting, not as a courtroom trial you have run. Burden of proof as a designed asymmetry. The prosecutor’s fallacy as the same inversion with liberty as the error-cost. The cab problem as a psychology-class object whose “correct” 41 percent is a live academic dispute — use Casscells when you need an uncontested teaching posterior.¹⁵⁰ Teaching objects. Not proof of a transferable legal-reasoning skill.

¹⁵⁰Amos Tversky and Daniel Kahneman, “Judgment under Uncertainty,” *Science* 185 (1974): 1124–31. Cab problem as teaching object; whether ~41 percent is the uncontested Bayesian answer is a live dispute (Birnbaum 1983 and later). Casscells ~2 percent, as the 1978 item was put, is the uncontested teaching posterior in this chapter.

Practice that actually builds learning

Blocked. Three recodings, same structure, different numbers, one week. Then, separately, three situated stories in one domain they know — not three slogans.

Mixed. A pamphlet percent, an interview prompt, a headline that left the absolute risk out, a history claim that needs a base rate (“rare, therefore surprising”). The student names which representation to reach for.

Retrieval. Two minutes of already-taught science: what prevalence means, what a control is, what a sample is. Two minutes of the project they will tell the story from. You cannot recode a mammogram number you have no picture of. You cannot tell a hiring story from a year you did not live.

One incorrect example. After they have attempted, a waiting-room poster that says “25 percent fewer deaths” with no count, next to a recruiter sentence that says “we test critical thinking” and then asks for a story. Ask what each object actually is. They name the frame and the survey.

Kitchen and money can motivate a tree — a sale that is “50 percent off” of a marked-up price — but they do not replace the clinic numbers or the unaided story. The error-cost in the week’s main sitting is a body or a hire, even when the numbers are practice numbers.

Elwyn’s three talks can be a role-play once the count exists: you are the clinician naming a choice; they are the patient who must hear a count; then you switch. Role-play appeared as an anchored-instruction cell in Abrami’s 2015 combination with a large effect from a small number of studies.¹⁵¹ Treat it as a way to make option talk carry a number, not as proof that drama raises a Watson–Glaser. Keep the live check: they recode with the pamphlet closed.

For the student

Some pages this year will be about a body. Some will be about a job. Neither page is a badge that you “have critical thinking.” Each is a domain with knowledge you have to bring, a format that helps, and a cost if the fluent answer is wrong.

¹⁵¹Philip C. Abrami et al. (2015): role-play $g = 0.61$ from 5 studies — suggestive anchored instruction, small k , not a proven accelerator of a generic trait.

The health move. When you hear a percent that wants you to decide, draw a thousand. How many have it before the test? How many well people will still test positive? Of the positives, how many are really sick? Read that last number off the paper. That is the number you can consent with, or refuse with, or take to a second question.

The work move. When someone asks whether you can think, they often mean: tell me a time you saw a problem coming, in a real context, with information, and did something. That story has to come from a domain you actually know — a lab, a budget, a build, a paper. A slogan is not a story.

Tiny worked example. Prevalence 1 in 1,000. False-positive 5 percent. Draw 1,000. One sick. About 50 well people flagged. About 1 in 51 of the flagged people is sick. About 2 percent. The fluent 95 percent answered a different question.

Two tries.

1. Recode this (illustration): 1 in 100 have a condition; a test wrongly flags 10 in 100 well people. A person tests positive. About what chance they have it? Draw a hundred or a thousand. Show the counts.
2. Four sentences, notes closed: a time this year you anticipated a problem in a subject you study. Context. What you knew. What you noticed. What you did.

Explain it back. Why is “95 percent accurate” not the same question as “given a positive, what’s the chance I have it?” Why is a recruiter’s importance rating not a work sample?

Challenge. Find one real pamphlet, article, or ad this week that uses a relative-risk percent. Recode it, or write what count is missing. Then tell one hiring-style story from a domain you know, out loud, to a person who will ask one follow-up.

You are allowed to struggle on the tree. Harvard samples struggled. Struggle means the inversion is tempting, not that you cannot think. The paper is there so you do not have to invert in your head.

If it isn't clicking

1. They keep writing 95 percent. Next move: take the word *percent* off the page. Only counts. 1 sick in 1,000. 50 false flags. “How many flagged people are there?

How many of those are sick?” When the fraction is easy, put the percent back as the last step, not the first.

2. They recode in science class and invert again on a pamphlet. Next move: mixed practice the same day, not a week later. One textbook item, one advertisement, one spoken sentence from a relative. Near is not far, as Chapter 3 already said. Similarity has to be practiced on purpose.

3. Their “interview story” is a slogan. Next move: four labelled sentences, always, for two weeks: context / knew / noticed / did. If a sentence cannot be labelled, it comes out. The domain has to be real. “I think critically” is not a context.

When to slow down: they do not yet know what a false positive is, or what prevalence means. Teach the science. The tree waits on the words.

When to go ahead: they recode two unseen percents into a thousand, catch an inversion in someone else’s work, and tell one four-sentence story from a known domain with notes closed.

When to get a human tutor: the hole is mathematics they have not finished — fractions of a thousand, or “percent of” versus “percent off.” A math tutor is the right valve. A “medical critical thinking” workbook is the sitting again. If the hole is interview anxiety, practice the story with a trusted adult; the content is still the year they lived.

Goals are not given. A patient may refuse a screen whose absolute benefit is 1 in 1,000 because 1 in 1,000 is not worth the harm *to them*. That refusal can be informed. A patient may accept a screen because a relative-risk poster frightened them. That acceptance can be fluent and uninformed. Shared decision-making is not a device for producing the clinician’s preferred option by nicer means. It is a device for making the person’s goals, and the actual numbers, jointly usable.

Tools, including AI

Optional helpers for the adult. Student attempts first. You hold the key — the counts, and what a decent story in this year’s domain looks like.

A box of rules

- Child attempts first, on paper, for the tree.

- Parent holds the answer key.
- Use a language model to explain natural frequencies to *you*, to generate extra count problems with answers hidden, or to write a hint script (“draw a thousand before you write a percent”).
- Use a tool after the attempt to check arithmetic, not to invert for them.
- During the attempt: no “what’s the chance I have this disease?” pasted into a chat, no unsupervised chatbot, no photo of the pamphlet sent for a verdict.

A model can emit “95 percent” fluently, with bedside manner. Fluency is the failure mode this chapter is interrupting. Chapter 8 will spend the machine. This week the tree is on paper so the fluent sentence has somewhere else to go.

Symptom-checker apps and hiring-coach bots are the same object in different costumes: a finished-looking answer. The student still recodes. The student still tells the story from a year they lived. A bot that invents a work story is a false statement if it is later treated as the student’s.

Gigerenzer’s papers and the AMA teaching restatement of Casscells are the adult reading if you want more than this chapter.¹⁵² ¹⁵³ They are not a curriculum to buy. Elwyn 2012 is a practice sequence you can steal in a role-play.¹⁵⁴ NACE’s sample interview item is public.¹⁵⁵ None of these is a franchise.

What “done enough” looks like

- You can draw the Casscells thousand and read about 2 percent off the paper.
- You can hear “25 percent reduction” and ask “1 in how many?”

¹⁵²Ward Casscells, Arno Schoenberger, and Thomas B. Graboys, “Interpretation by Physicians of Clinical Laboratory Results,” *New England Journal of Medicine* 299 (1978): 999–1001. Harvard Medical School faculty, staff, and students; prevalence 1/1,000; false-positive 5 percent; 27 of 60 said 95 percent; 11 of 60 gave about 2 percent. *AMA Journal of Ethics* (2009) restates the item next to shared decision-making.

¹⁵³Gerd Gigerenzer, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa M. Schwartz, and Steven Woloshin, “Helping Doctors and Patients Make Sense of Health Statistics,” *Psychological Science in the Public Interest* 8, no. 2 (2008): 53–96. Statistical illiteracy in patients, journalists, and physicians; mammography “25 percent reduction” as 1 in 1,000 absolute.

¹⁵⁴Glyn Elwyn, Dominick Frosch, Richard Thomson, et al., “Shared Decision Making: A Model for Clinical Practice,” *Journal of General Internal Medicine* 27 (2012): 1361–67. Choice talk / option talk / decision talk. Practice model, not a population-health RCT.

¹⁵⁵NACE, “Career Readiness Competencies: Critical Thinking” (December 2025). Sample interview item: describe a time when you anticipated a problem.

- You can hear “number-one skill” and remember a survey, a story prompt, and communication at 4.57 next to critical thinking at 4.49.
- The student recodes two unseen percents, catches one inversion, and tells one four-sentence situated story from a domain they know, notes closed.
- You have dated trees and one dated story in the folder. That is evidence in two domains. It is not a transferable badge, not a NACE percentile, and not 795,000 printed as a headcount.

Move on to Chapter 8 when the tree can be drawn without a lecture and the story has a context. Chapter 8 is the life condition that a machine will finish the sentence. Evaluation remains the human work. A fluent clinic number and a fluent cover letter are both sentences. The remaining work is the same work this week named: recode, check, refuse the inversion, say it with the window closed.

You can do this. Tomorrow’s sentence is already written: “Give me the count in a thousand, then tell me the story.”

Chapter 8

When a machine finishes



Figure 9: A stethoscope, a gavel, a stack of folders, and bound books on a dark desk.

Why this matters

A machine will finish the sentence. That is the life condition this chapter is about.

The remaining human work is evaluation.

In one high-school mathematics trial, unguarded chatbot practice raised session scores and then left students worse on the closed-book exam that followed — about 17 percent below classmates who never had the machine.¹⁵⁶

A cloaked .org fooled students in Chapter 6 because it looked finished. A relative-risk poster fooled a waiting room in Chapter 7 because it looked like a magnitude. A generated paragraph fools for the same reason: it looks like thought. The surface form of completed judgment is now cheap. Evaluation is not.

This is not a classroom-AI handbook. It is not five chapters of tutor design. It is one week of teaching the evaluative loop on a fluent page that nobody in the house had to write: interpret the claim; check it laterally, numerically, or against a source you actually have; refuse a fluent wrong answer; say the conclusion with the window closed.

What you can learn this week is that loop, in language you can say at the table. What the student can try today is one generated paragraph in a subject they already know, checked off the page, then defended with the model shut.

You do not need to become a computer scientist. You do need today's idea well enough to hear "it sounds right, so it is evaluated" as a stay, not as thought.

Ennis's sentence still names the work: decide what to believe or do, including when the sentence arrives finished.¹⁵⁷ Willingham still vetoes a content-free "AI literacy" module as the engine of that work.¹⁵⁸ Thinking about a generated medical

¹⁵⁶Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman, "Generative AI without Guardrails Can Harm Learning," *Proceedings of the National Academy of Sciences* 122, no. 26 (2025): e2422633122. Unguarded GPT-style practice about +48 percent on session work; about -17 percent versus never-AI on a closed-book isomorphic exam; students did not perceive the loss. Hint-tutor arm: practice up, unaided exam approximately at control. Cited here as the life condition, not retold as classroom-AI design.

¹⁵⁷Robert H. Ennis, "The Nature of Critical Thinking," criticalthinking.net (2011); 1985 *Educational Leadership* working definition: reasonable reflective thinking focused on deciding what to believe or do.

¹⁵⁸Daniel T. Willingham, "Critical Thinking: Why Is It So Hard to Teach?," *American Educator* 31, no. 2 (Summer 2007).

number is thinking about medicine. Thinking about a generated history paragraph is thinking about history. The machine does not lift the knowledge constraint. It makes the constraint more expensive to ignore, because the fluent output *looks* as if the constraint has already been met.

For the parent: understand it yourself

Many adults already use a language model for mail, summaries, and “what’s a good way to explain this.” That is ordinary. Five minutes of this section is about keeping evaluation as the human job when the draft is free.

Everyday picture. A friend texts a polished paragraph about a local school-board claim, a supplement, or a statistic. You did not watch them write it. It has topic sentences. It has a confident last line. The old cue — “this took time, so someone thought” — is gone. Time is no longer a filter. Your job is the same job Chapter 6 taught on a website: leave the paragraph, find who and what would count, then decide. The friend may be a model. The friend may be a person who pasted a model. Either way the object is a finished-looking sentence.

Precise picture. The evaluative loop, named as a loop rather than as a new faculty:

1. **Interpret.** What claim is being made, and what would have to be true for it to be usable?
2. **Check.** Laterally (who’s behind this, what’s the evidence, what do other sources say). Numerically (recode the percent as a count). Against a source you actually have (a document, a worked example, a number you can locate).
3. **Refuse.** A fluent wrong answer is still wrong. Refusal can be costly. Adoption without refusal is not a loop.
4. **Defend with the window closed.** If you cannot say the conclusion, the check, and the remaining uncertainty without the model, you have a document, not a judgment.

That loop is Ennis plus Willingham plus Wineburg plus Gigerenzer. It has no transfer trial showing that “evaluate the chatbot” becomes better judgment on non-machine pages, or the reverse.¹⁵⁹ Willingham predicts it will not, without shared domain knowledge. Teach the loop *in a subject they know*, on a paragraph that hap-

¹⁵⁹No randomized trial showing that “evaluate the chatbot” transfers to non-AI source evaluation, or the reverse, was found in the research this book stands on (access date 30 August 2026).

pens to be generated. Surface “AI tells” change. Fluency is the durable property. The durable move is leaving, recoding, and noticing when you do not know.

Yizhou Fan and colleagues found a related split on an essay task: the chatbot group improved essay scores more; knowledge gain and transfer did not differ; process mining showed fewer evaluation and orientation moves than a human-expert group. The authors called that “metacognitive laziness.”¹⁶⁰ An authors’ term for skipped steps, not a verdict on character. A better artefact is not evidence of better judgment.

Matthias Stadler, Maria Bannert, and Michael Sailer compared a model with a search engine on a short scientific-recommendation task. The model made the work easier and the justifications thinner; the drop in learning-relevant effort accounted for the quality difference.¹⁶¹ Ease is not depth. Offloading is old — calendars, lists, search. Evan Risko and Sam Gilbert described it as a metacognitive choice that can be wrong.¹⁶² Betsy Sparrow, Jenny Liu, and Daniel Wegner found that when people expect information to remain available, they remember where, not what.¹⁶³ A language model is a more aggressive store. It does not only hold the fact. It holds a finished-looking argument about the fact. The new risk is offloading *evaluation*, with fluency as the false cue that evaluation occurred.

Lisa Fazio and colleagues showed that repetition increases judged truth even when you know better: knowledge does not protect against illusory truth.¹⁶⁴ A machine that can emit fluent sentences at the scale of a feed is a fluency engine. Combined with Chapter 6’s look-finished and Chapter 7’s 95 percent, the mechanism is not mysterious. People use the feeling of a finished product as a cue that evaluation occurred. When the product is generated, that cue is a lie about the human’s work.

¹⁶⁰Yizhou Fan, Luzhen Tang, Huixiao Le, et al., including Dragan Gašević, “Beware of Metacognitive Laziness,” *British Journal of Educational Technology* 56, no. 2 (2025): 489–530. N = 117; better essays; no significant knowledge gain or transfer; fewer evaluation and orientation moves than a human-expert group.

¹⁶¹Matthias Stadler, Maria Bannert, and Michael Sailer, “Cognitive Ease at a Cost,” *Computers in Human Behavior* 160 (2024): 108386. N = 91; ChatGPT-3.5 versus Google; thinner justifications; germane-load drop mediated the quality difference.

¹⁶²Evan F. Risko and Sam J. Gilbert, “Cognitive Offloading,” *Trends in Cognitive Sciences* 20, no. 9 (2016): 676–88.

¹⁶³Betsy Sparrow, Jenny Liu, and Daniel M. Wegner, “Google Effects on Memory,” *Science* 333, no. 6043 (2011): 776–78.

¹⁶⁴Lisa K. Fazio, Nadia M. Brashier, B. Keith Payne, and Elizabeth J. Marsh, “Knowledge Does Not Protect against Illusory Truth,” *Journal of Experimental Psychology: General* 144, no. 5 (2015): 993–1002.

Two lanes are a design hypothesis, not a fitted life-span schedule, and not a district policy you must adopt.¹⁶⁵ *Formation*: the skill being built is practiced without the model, or with hints that withhold the finished answer. *Performance*: the model is allowed, logged, and judged as a tool, and the human is the evaluator. Collapsing formation into an unguarded chat window is the exam loss in the trial that opened this chapter. Collapsing performance into a detector is the other failure.

Detectors fail, and they fail unfairly. Weixin Liang and colleagues found GPT detectors misclassified 61.22 percent of human TOEFL essays as AI-generated, while remaining near-perfect on U.S. eighth-grade essays. Eighty-nine of ninety-one TOEFL essays were flagged by at least one of seven detectors.¹⁶⁶ Debora Weber-Wulff and colleagues found commercial tools neither accurate nor reliable, and worse after paraphrase.¹⁶⁷ A detector score is not critical thinking, not evidence of cheating, and not a substitute for the live check Chapter 5 already wanted. It punishes a writing profile — often honest writers whose English is not a first language.

A 2025 EEG preprint on essay writing with a chatbot will be cited in headlines as proof that machines rot brains. It is a preprint, with a small sample, measuring connectivity during a lab task. Connectivity is not Ennis's sentence.¹⁶⁸ Peer-reviewed reasons to keep evaluation human already exist. The trial that opened this chapter is one. Fan is another. You do not need a civilizational morality play to run this week's loop.

Wrong answers you should be able to hear

1. *"It sounds finished, so someone thought."* Fluency is cheap. Ask for the check.

¹⁶⁵Two lanes (formation versus performance) are a design hypothesis motivated by Bastani, Fan, Stadler, and Liang, constrained by Willingham — not existing district policy and not a fitted life-span schedule.

¹⁶⁶Weixin Liang, Mert Yuksekogunul, Yining Mao, Eric Wu, and James Zou, "GPT Detectors Are Biased against Non-Native English Writers," *Patterns* 4, no. 7 (2023): 100779. Average false-positive rate 61.22 percent on 91 human TOEFL essays; 89/91 flagged by at least one of seven detectors; near-perfect on U.S. eighth-grade essays.

¹⁶⁷Debora Weber-Wulff et al., "Testing of Detection Tools for AI-Generated Text," *International Journal for Educational Integrity* 19, art. 26 (2023). Fourteen tools; neither accurate nor reliable; worse after paraphrase.

¹⁶⁸Nataliya Kosmyna et al., "Your Brain on ChatGPT," arXiv:2506.08872 (2025). Preprint as of 30 August 2026; n = 54 / 18; EEG connectivity during essay writing; not a population finding and not a measure of Ennis's sentence.

2. *“The detector said 90 percent AI, so they cheated.”* Detectors mislabel honest TOEFL-like prose at high rates. Use a live check in the subject.
3. *“We’ll just add an AI-literacy unit.”* Without domain knowledge, “evaluate the chatbot” is a maxim. Teach the loop inside history, science, or math they already know.
4. *“Practice scores went up, so they are learning.”* The opening trial is the dissociation: session performance up, unaided judgment down, the loss unperceived.
5. *“I’ll have the model evaluate the model.”* That is another finished sentence. The student still has to interpret, check, refuse, and say it closed.

Five-minute parent warm-up

Do this before the lesson, no student in the room.

Minute 1. Pick a subject you already teach. Ask a model (or write a fluent wrong paragraph yourself) for a confident ten-sentence explanation of something in that subject. Keep it off the student’s screen for now.

Minute 2. Interpret: what is the main claim? What would have to be true?

Minute 3. Check off the page: one new tab, or one number recoded, or one sentence in the textbook. Mark one error or one uncertainty.

Minute 4. Close the model. Say the conclusion, the check, and what you still do not know, out loud, in thirty seconds.

Minute 5. The sentence you will say tomorrow: “The machine finished the sentence. Evaluation is still ours.” Put the pencil down.

If you can do those five minutes, you are ready. The student checks. You hear whether they checked.

How to teach it this week

Same session shape. The object is a fluent paragraph in a known subject. You preview it. You know at least one check that would catch a problem.

Warm-up (3–5 minutes, unaided). Two facts from this week’s subject. One tiny prompt: “If a paragraph about this topic arrived already polished, what would you

check first?” Paper. Model closed.

Short model (5 minutes). You display a fluent paragraph you generated or copied. You do not ask “does this sound good?” You interpret the claim in one sentence. You leave it — new tab, textbook, or a tree of counts. You mark one check. You close the window and say the conclusion. Then you stop talking.

Student attempt (10–15 minutes). A different paragraph, same subject. They run the four steps in writing. You wait. Struggle before rescue.

One good question. “What would have to be true?” “Where did you check off this page?” “Can you say it with the window closed?”

Mixed practice (5–10 minutes). Yesterday’s live website (Chapter 6) next to today’s generated paragraph next to a clinic percent (Chapter 7). The student names the object and the check: leave, recode, or both.

Exit ticket (5 minutes, unaided). Window closed. Two spoken sentences: the claim, and the check. If they cannot say the check, the sitting was performance, not formation.

Exact wording you can say

“The machine finished the sentence. Evaluation is still ours.”

“What is the claim, in one line?”

“Leave this paragraph the way you left the website. Who, evidence, other sources.”

“If there is a percent, draw a thousand.”

“Refuse a fluent wrong answer, including this one.”

“Close the window. Tell me what you believe, what you would do, and what you still do not know.”

When you are about to take over:

“Your check. I’ll listen.”

Then wait.

First problem for the student

A fluent paragraph in this year’s science or history, previewed by you, with at least one checkable flaw or one missing source — labelled as practice, not as a reported

family.

Say: “This looks finished. Interpret the claim. Check it off the page. Then close everything and tell me.”

Wait. If they paraphrase the paragraph, they treated fluency as evaluation. If they open a new tab or the textbook, they have started the loop. If they freeze, model a *different* paragraph fully, then return to theirs.

Later the same week, the diagnostic: a paragraph that is *right*, fluently. Hear “it sounds good, so I skipped the check.” Hear “I checked anyway; here is the source.” Checking a true paragraph is still the work. The loop is not a hunt for gotchas. It is a refusal to let ease finish the job.

How to fade help

Day one: you run the four steps on a different text. Day two: they interpret and check; you hold the closed-window defense. Day three: they run all four. When they close the window without being told, stop narrating the loop. Bring the model back when the subject changes. A generated proof and a generated civic claim share fluency and do not share the knowledge the check requires.

When to stop talking

After “what is the claim?” After “new tab or textbook.” After modeling a different paragraph. If you are correcting the generated essay while they watch, you have started evaluating for them. Stop. Point at the loop on a card. Wait.

Formation sittings this week: model closed, or hints only. Performance sittings, if you use the model as a draft tool: log that it was used, and the grade (or the kitchen equivalent) is the check plus the closed-window defense, not the polish of the draft. Mixing the lanes without saying which sitting you are on is how a person spends a year with a fluent crutch and calls it expertise. Adults will keep models in the room. Some stakes still have to remain formation: a number that will be used on a body, a source you will share, a diagnosis you must understand. Chapter 7 already showed that physicians invert those numbers. A model that emits 95 percent fluently has not made the inversion rarer. It has made the inversion look like expertise.

Practice that actually builds learning

Blocked. Three generated (or fluent) paragraphs in the *same* subject, four steps each. The student is learning the loop, not collecting “AI tells.”

Mixed. A live website, a pamphlet percent, a generated paragraph, a human-written textbook page. The student names which check applies. Keyword “AI” will not always be on the object. Fluency will.

Retrieval. Two minutes of domain facts before the paragraph. You cannot check a Civil War claim without Civil War knowledge. You cannot recode a screening sentence without prevalence. The model is more likely to be locally fluent in a domain you do not know, which is exactly when you cannot see that it is wrong.

One incorrect example. After they have attempted, a composite student who pasted the prompt, received a fluent essay, submitted it, and could not reconstruct the argument with the window closed. Ask what sitting that was evidence of. Ask what sitting it was not. They name the artefact-without-judgment problem Chapter 5 already taught. The machine made the hole unskippable. It did not invent it.

A take-home essay is no longer, by itself, evidence of the student’s judgment. Norman Frederiksen argued that tests shape the skills we get; convenient artefacts are not the construct.¹⁶⁹ Mike Perkins and colleagues’ AI Assessment Scale is language for aligning allowed tools with evidence — process, artefacts, reflection, demonstration, live elements — not a trial that a scale raises thinking scores.¹⁷⁰ Choosing a level is the start. The work is the evidence. In the kitchen, the live check is the evidence.

Keep practice short enough to finish. Three loops with a closed-window defense beat a dozen generated essays nobody talked through.

No need to run a “spot the AI” contest. Tells change. The durable property is fluency. The durable move is the loop.

A worked incorrect loop, labelled as an illustration. A composite student asks a model for a paragraph on a historical treaty, pastes it into a paper, and adds

¹⁶⁹Norman Frederiksen, “The Real Test Bias,” *American Psychologist* 39, no. 3 (1984): 193–202.

¹⁷⁰Mike Perkins, Jasper Roe, Leon Furze, and Jason MacVaugh, AI Assessment Scale (2024), *Journal of University Teaching and Learning Practice* 21, no. 6. Language for assessment redesign, not an RCT. Mike Perkins, not David Perkins. Dissemination claims are not effect sizes.

one citation the model invented. The prose is clean. The citation does not exist. The live check — “open the treaty page; point to the sentence” — would have caught it. A detector might have called the prose human, or might have called a careful English-learner’s prose machine-made, and would still have missed the fake source. Process without a live check is theatre. Process *with* a live check is this week’s mixed practice: the annotation of where they looked, plus two minutes of pointing.

Sharing as the civic twin. Chapter 6’s accuracy question still belongs at the share button: is this true? A generated paragraph can be shared as easily as a cloaked .org. The loop before sharing is the same loop: who is behind this (a model plus a prompt plus whatever the prompt retrieved), what is the evidence, what do other sources say. If the student cannot answer those with the window closed, the house does not share. Sharing is a doing. Ennis includes do.

What this week is not. It is not a ban. It is not an always-on copilot as a personality. It is not a promise that evaluating chatbot output will, by itself, transfer to a clinic or a jury a year later. Near practice on fluent pages in a known subject is what you can defend. Far transfer remains the default failure Chapter 3 already taught. Teach near, on purpose, and let similarity do the small work it actually does.

For the student

You will live in a world where a polished paragraph is cheap. That does not make you cheap. It makes evaluation your job.

A machine can finish your sentence. It can finish a stranger’s sentence and hand it to you as a source. It can sound kind, certain, and complete. Sounding complete is not the same as being checked.

Here is the loop.

Interpret: what is this claiming?

Check: leave the page, recode the number, or open a source you already trust in this subject.

Refuse: if it is fluently wrong, you do not have to adopt it. Saying “I don’t know yet” is a legal answer.

Defend: close the window. Say what you believe, what you would do, and what you still do not know.

Tiny worked example. A fluent paragraph says a rare-disease test is “95 percent accurate, so a positive result means you almost certainly have it.” Sounds like expertise. Interpret: it is swapping “how often the test is right about well people” for “how likely you are to be sick given a positive.” Check: draw a thousand, as in Chapter 7. Refuse the 95 percent as the answer to the second question. Close the window. Say about 2 percent, and that you would ask for the count before you consented.

Two tries.

1. Your parent or teacher will give you a fluent paragraph in a subject you already study. Run the four steps. Write them.
2. Close everything. Say the claim, the check, and one thing you still do not know.

Explain it back. Why is a better essay not automatically better thinking? Why is a detector score not the live check?

Challenge. This week, take one paragraph a machine (or a person) finished for you — mail, a summary, a homework draft. Run the loop. If you cannot defend it closed, it is still a draft. You are allowed to use a tool on a performance sitting. You are not allowed to confuse the polish with your judgment.

You are allowed to struggle. The students in the opening trial did not perceive what they had lost. Noticing the seam is the skill, not finishing faster.

If it isn't clicking

1. They paraphrase the fluent page and call it a check. Next move: take the paragraph away before they write. They get the topic and one question (“what would count as evidence?”). They must leave, then return. Paraphrase becomes impossible without the costume on screen.

2. They hunt “AI tells” (odd phrases, too-perfect structure) and miss the claim. Next move: ban tell-spotting for a week. Only the four steps. A true human paragraph and a generated one, unlabelled. The loop has to work on both.

3. They freeze when the window closes. Next move: shrink the defense. One sentence of claim, one sentence of check. Every sitting, two weeks. Lengthen when those two sentences appear without panic. If the freeze is the subject, go back to knowledge. A loop cannot supply the chemistry.

When to slow down: they cannot yet leave a website (Chapter 6) or recode a percent (Chapter 7). Those moves are the check. Teach them in the subject first. Then point them at a generated page.

When to go ahead: they run the four steps on two new fluent paragraphs in a known subject, catch at least one fluent error or one missing source, and defend with the window closed.

When to get a human tutor: the bottleneck is the subject, or writing itself, not the machine. A subject tutor who will sit the live check is the right valve. An “AI literacy” course with no content is the maxim again. A detector dashboard is not a tutor.

If the house is tired of policing tools, name the sitting instead. “This hour is formation: window closed.” “This hour is performance: you may draft with a tool; you will still defend closed.” Mixing without naming is the confusing part. Naming is cheap. The closed-window minute is the expensive, useful part.

Tools, including AI

This chapter is about the tool. The box is still short.

A box of rules

- Child attempts first on formation sittings.
- Parent holds the key: the claim, a decent check, the live-defense questions.
- Use a language model to explain the loop to *you*, to generate extra fluent paragraphs in this week’s subject with hidden flaws you have verified, or to write a hint script that withholds the answer.
- A hint-only use — questions, not finished solutions — is the tutor-shaped use that did not harm unaided exams in the opening trial’s guarded arm.¹⁷¹

¹⁷¹Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakçı, and Rei Mariman, “Generative AI without Guardrails Can Harm Learning,” *Proceedings of the National Academy of Sciences* 122, no. 26 (2025): e2422633122. Unguarded GPT-style practice about +48 percent on session work; about −17 percent versus never-AI on a closed-book isomorphic exam; students did

Unguarded “write this for me” is the other arm.

- During the attempt: no “what’s the answer,” no photo-to-key, no unsupervised chatbot on a formation sitting.
- Detectors are not the live check.

You may use a model to make extra practice with answers hidden. You still verify the subject content. Models invent citations. A practice paragraph with a fake source is a bad item unless catching the fake source is the point, and you know it is fake.

Two-lane language, not a schedule: formation hours build the skill; performance hours judge the human as evaluator with the tool in the room. There is no fitted age at which formation ends. A consent number does not end. An interview story does not end. Email tone might be performance earlier. You choose the sitting by the error-cost, not by the calendar.

If a co-op or a relative wants a detector report in the file, you now have a calm answer. We can talk about the draft. We will not treat a classifier as a character test. Liang’s TOEFL result is enough reason.¹⁷² The live defense in the subject is the object Chapter 5 already named. This chapter only adds that the artefact is now cheap, so the defense is no longer optional.

A short hint script you can steal, then fade:

“What is the claim, in one line?”

“Where will you check that is *not* this paragraph?”

“What would change your mind?”

“Window down. Tell me.”

Ten seconds of silence after each line is still teaching.

not perceive the loss. Hint-tutor arm: practice up, unaided exam approximately at control. Cited here as the life condition, not retold as classroom-AI design.

¹⁷²Weixin Liang, Mert Yuksekgonul, Yining Mao, Eric Wu, and James Zou, “GPT Detectors Are Biased against Non-Native English Writers,” *Patterns* 4, no. 7 (2023): 100779. Average false-positive rate 61.22 percent on 91 human TOEFL essays; 89/91 flagged by at least one of seven detectors; near-perfect on U.S. eighth-grade essays.

What “done enough” looks like

- You can say, in one sentence, that a machine finishes the sentence and evaluation remains the human work.
- You can hear the five wrong answers and know that each skips the loop.
- The student runs interpret / check / refuse / closed-window defense on two fluent paragraphs in a subject they know.
- A detector score is not in the folder as evidence of cheating or of virtue.
- You can name, for tomorrow’s hour, whether it is formation or performance.
- The folder has dated loops, not a pile of polished drafts nobody talked through.

This chapter was one chapter. It will not grow into a product tour, an age-band standard operating procedure, or a sequel. The deeds that remain are ordinary: pause, leave the page, recode the number, notice when you do not know, refuse a fluent wrong answer, including from a model, including when refusal is costly.

Move on to records when the loop can be run without a lecture. Chapter 9 is the paper trail a stranger can read. You will not title a magic course “Critical Thinking 101” and expect a registrar to understand the work of these weeks. You will embed the work in the subjects where it already lives.

You can do this. Tomorrow’s sentence is already written: “The machine finished the sentence. Evaluation is still ours.”

Chapter 9

Records a stranger can read

The teaching job was the rest of this book: Ennis's sentence in plain language, knowledge first, maxims that fail without practice, a score that is not the skill, leaving the page, recoding a number, evaluating a sentence a machine finished. This chapter is the paper trail that work has to leave the house on.

There is no national homeschool diploma, and there is no national “critical thinking” credit a registrar already knows how to file. A parent, an umbrella school, a church-school statute, or a private-school affidavit can put a graduation date on a transcript and a signature under it. Colleges will still want titles they can map onto English, history, government, biology, mathematics. Some will want course descriptions. Some will want an outside signal — an SAT or ACT score, an AP score, a dual-enrollment transcript. None of those objects is produced by stamping *Critical Thinking 101* on a line.

State compliance is not college evidence. Some states can be a legal year without a portfolio. Other states already want a plan, a portfolio, or a test. Those files are compliance. They are not, by themselves, a high-school transcript a stranger can read. Read your own state's statute. An advocacy card is a finding aid, not the code. This chapter is not a fifty-state legal manual.

Two readers. Two calendars. A family can satisfy a silent state and still owe a stranger a transcript. A family can satisfy a thick state file and still owe the stranger titles, credits, and grades. An umbrella school can put a name and a signature on the page. It still needs United States History written as United States History.

Titles a stranger can read

Practice, from college homeschool pages and from athletic-eligibility example transcripts in the mathematics book of this series, is a translation document, not a scrapbook.¹⁷³ The fields a stranger can use without a phone call are few:

- Student name and address.
- Ninth-grade start date, month/day/year if athletics are in view.
- Course titles a registrar can map: English 9–12, United States History, World History, American Government / Civics, Biology, Chemistry, Physics, Algebra I, Geometry, Algebra II, and so on.
- Grades, credit units, a grading scale, academic year, graduation date.
- Administrator signature.

Dual-enrollment courses listed on the homeschool transcript and backed by the official college transcript. Credits in clean increments: a year is 1.0; a semester is 0.50.

Title the year as the subject, not as “Critical Thinking 101.” Embed the work in the subjects.

A registrar who sees “Critical Thinking 101, A, 1.0” has been given a slogan. They do not know whether the year was informal logic, a test-prep sitting, a media unit, or a philosophy circle. They cannot map it onto the core. A line that says “American Government, A, 1.0” with a course description that names Civic Online Reasoning’s three questions, live pages, and a folder of dated lateral-reading tasks has done the translation twice. A line that says “Biology, A, 1.0” with a description that names control-of-variables, natural-frequency recoding of a screening claim, and oral defense of lab conclusions has done it in science. The evaluation habits live *inside* the subject. The title is the subject.

Publisher and program names are not titles. A SHEG lesson sequence is not “eleventh grade.” A Philosophy for Children enquiry hour is not English 10 unless the year’s English work actually was English. Rationale maps are not a social-studies credit by themselves. Khan Academy’s free map is not United States His-

¹⁷³Translation fields follow the same homeschool-transcript practice used in this series’ mathematics manual: NCAA Eligibility Center example transcript and Home School Toolkit (2025–26); College Board homeschool assessment guide (Jacob Mishook, 29 June 2026), SAT code 970000; Common App counselor flow. Confirm live pages in the year of application. Pattern, not a national rule.

tory until you write what was actually read, written, and checked.

Honesty on the title is the same rule as honesty on a case in Chapter 5. If the year's government work never left a page, American Government is still the right title only if the civics content was taught; the description should not claim lateral reading you did not do. If the year's work was a stand-alone logic workbook and no history, World History is the wrong title. A beautiful portfolio titled "Thinking 12" has satisfied one reader and confused the next.

A *hypothetical* illustration, labelled as such: a parent who used a standard U.S. history text, six Civic Online Reasoning lessons inside that year, and a government semester writes two lines — "United States History (narrative text; document work; lateral reading on modern claims), A, 1.0, 2024–25"; "American Government (constitution; six live-page evaluations; oral defense), A–, 1.0, 2025–26." A dual-enrollment introductory statistics sitting that spring is listed on the home-school transcript *and* sent as an official college transcript, which is where the natural-frequency work of Chapter 7 can live without inventing a "Health Thinking" credit. The titles and the years are hypothetical. The rule is not.

Placement and acceleration are record-keeping choices as well as teaching choices. A fourteen-year-old may be ready for document-based history and not yet ready to recode a clinic percent. A sixteen-year-old may tell a strong situated work story from a chemistry year and still stay on the page when a cloaked .org looks finished. Skill, not birthday. A "grade 8 critical thinking workbook" is a publisher's scope, not a legal grade, and not a transcript line.

Informal learning does not arrive as "Civics, 1.0, B+." A plan, a portfolio, or an athletic worksheet still wants a title and a grading basis after the fact. The translation is possible. It is work. That is a fact about documents, not a verdict on the practice.

Common App is the other translation surface most families will actually meet. Practice, from practitioner guides: the student marks homeschooled; invites the parent, or an umbrella administrator, as counselor; the counselor account uploads the School Report, the transcript, a school profile, and course descriptions. Parent-as-counselor is standard, not a workaround. Confirm against Common App help text in the year you apply. A school profile in this room is a one-page description of the homeschool: who taught, what the grading scale is, whether the school is the kitchen or an umbrella, how dual enrollment and AP were used. It is not a marketing brochure, and it is not a paragraph that says the student is a critical

thinker. Let the course descriptions carry the habits.

A language model may draft a transcript *template* with those fields, or a course-description paragraph that lists the text, the major topics, the evaluation habits actually practiced, the grading basis, and the credit. You check every line against the actual work. You sign. Asking a model to invent hours, invent a diploma, invent a Critical Thinking 101 that was not taught, or back-fill lateral reading the course did not do is a false statement on a document a college will treat as signed. An agent that drafts a template has not signed.

What the course description actually says

Write once. Reuse. A stranger should be able to see the subject, the texts or primary sources, the named moves, and how you graded.

History. Texts and document collections by title. Moves a stranger can picture: sourcing, contextualization, corroboration; unseen documents; a live check. If Civic Online Reasoning lived here, say so in a clause — “live-web evaluation of modern claims that reuse this period’s themes” — not as a second course.

Government / civics. Constitution, institutions, and, if you taught it, the three questions of Chapter 6 on live pages, with dated tasks in a folder. Wineburg 2022 lived in a required government course for a reason.¹⁷⁴ Put the habit where the knowledge is.

Biology, chemistry, health. Experiments, texts, and, if you taught it, control of variables, and recoding of a screening or risk claim as counts. Title the year Epidemiology only if that was the course. Keep the recoding inside biology, health, or statistics rather than in a separate thinking credit.

English. Novels, essays, rhetoric. Argument maps, if you used them, belong in the description as a method for claim–reason–objection on *these* texts, not as a software credit.

Mathematics. Algebra I, Geometry, Algebra II, Statistics — the mathematics, as the mathematics manual of this series already taught. A natural-frequency tree can

¹⁷⁴Sam Wineburg, Joel Breakstone, Sarah McGrew, Mark D. Smith, and Teresa Ortega, *Journal of Educational Psychology* 114, no. 5 (2022): 893–909. Six COR lessons in a required high-school government course. The trial’s home is a reason to transcript the habit inside government, not as a stand-alone thinking credit.

be named in a statistics or biology description. It is not a substitute for Algebra II.

Work and electives. If a student completed a real project with a situated story — anticipated a problem, in a context, with information — describe the project under the subject it belonged to (robotics, journalism, running a household budget as consumer math). NACE’s interview item is a story in a domain.¹⁷⁵ It is not a transcript title.

Grade the subject the way you grade the subject: work completed, accuracy, the live checks you actually gave. A high thinking-test percentile, if someone sat one, can be an outside signal in the file. It is not the course grade. Chapter 5 already refused to let the sitting be the meaning.

Outside signals, kept in their lanes

Each receiving college sets policy. Credit is not automatic.

SAT / ACT / CLT. Admissions tests. The homeschool SAT high-school code is commonly 970000; confirm in the year you sit.¹⁷⁶ They are not a critical-thinking diploma and not a substitute for history titles.

AP. Homeschoolers may self-study. Parents cannot order exams. Find a school through the AP Course Ledger; the coordinator enrolls the student. A course labelled “AP” on a homeschool transcript is a different fact from a May score. Title United States History or AP United States History only if the work matches. Government, Biology, Statistics, English Language — these are places the evaluation habits of this book already live. An AP score is not a Watson–Glaser.

CLEP. Subject exams a parent can order without a school. Each college sets score and hour policy. CLEP is not an NCAA core course if athletics are in view. CLEP Introductory Sociology is not “Critical Thinking 101” under a new name.

¹⁷⁵National Association of Colleges and Employers, Career Readiness Competencies: Critical Thinking (December 2025). Sample interview item: describe a time when you anticipated a problem. A story in a domain, not a course title.

¹⁷⁶Translation fields follow the same homeschool-transcript practice used in this series’ mathematics manual: NCAA Eligibility Center example transcript and Home School Toolkit (2025–26); College Board homeschool assessment guide (Jacob Mishook, 29 June 2026), SAT code 970000; Common App counselor flow. Confirm live pages in the year of application. Pattern, not a national rule.

Dual enrollment. The cleanest outside transcript the kitchen can buy when you are the ceiling. Introductory statistics, American government, English composition, a lab science — official college transcript, listed at home as well. Admissions offices already know how to file it. Keep the syllabus. If the college course included document work, labs, or a seminar, the description can say so. The college’s title is the title.

Inventories. Watson–Glaser, CCTST, CLA+ are sittings. If a destination or an employer asks for one, sit it and put the score report in the file as what it is. Keep it in its lane: a sitting, not the homeschool’s proof of a life skill, and not a subscore that steers the transcript. Nursing’s long use of CCTST did not produce larger generic-test gains than other majors; that is a reason to keep the sitting in its lane.¹⁷⁷

NCAA, only if athletic scholarships at member schools are in view. Core courses are college-preparatory high-school work in English, mathematics (Algebra I or higher), science, social science, world language, comparative religion, or philosophy. “Critical Thinking 101” does not name a core. “American Government” does. “United States History” does. One worksheet per core course: teacher of record, texts, content, grading basis, signature. Grade and credit on the worksheet must match the transcript. CLEP and credit-by-exam are not NCAA core courses. Evaluation happens after a school requests it. Read the current Home School Toolkit rather than a forum restatement.

What to keep in the file

1. A transcript with titles a stranger can map onto the core. No magic thinking course as a substitute for those titles. Publisher or method in a parenthetical or on a description page. Credits in clean increments. Your signature.
2. Course descriptions that name the text, the major topics, the evaluation habits *actually practiced*, the grading basis, and who taught. Government that left the page should say so. Biology that recoded a percent should say so. History that used unseen documents should say so.
3. A thin folder of dated work: lateral-reading notes, natural-frequency trees, closed-window defenses, situated stories. That folder is your evidence of

¹⁷⁷Christopher R. Huber and Nathan R. Kuncel, “Does College Teach Critical Thinking?,” *Review of Educational Research* 86, no. 2 (2016). Nursing’s mandated generic-test use did not add a reliable increment over other majors. An inventory in the file is a sitting.

judgment. It is not a diploma.

4. Dual enrollment: list the course at home and send the college transcript. Keep the syllabus.
5. AP: the score, if a seat was found. The course titled honestly whether or not a seat was found.
6. SAT / ACT / CLT as the destination requires. PSAT as an away student with lead time if National Merit is in view.
7. Optional inventory score report, labelled as a sitting, if a destination asked.
8. NCAA packet, only if a coach and a member school are in view.
9. State compliance in its own folder: notice, portfolio, evaluator letter, required test. That folder is not the transcript.

The paper trail is what the kitchen can actually put in a stranger's hand. Write the titles as subjects. Keep the descriptions honest about the moves. Sit an outside signal the destination actually reads. When you are the ceiling, buy a transcript a college already knows how to file. You sign.

If a relative, a co-op director, or an umbrella administrator wants a thinking course "so colleges can see it," offer them the description paragraph instead of a new title. Two or three honest clauses inside Government, History, and Biology will do more for a stranger than a slogan line. If they still want a seminar named as an elective, title it by its content: Informal Logic, Introduction to Philosophy, Media and Civic Life, Epidemiology Topics. Then write what was read and what was checked. Electives are allowed. Magic titles are the problem.

About twelve percent of homeschooled students in 2016 were taught with mostly or strictly informal learning.¹⁷⁸ Lived judgment does not arrive pre-titled. After the fact, still pick the subject the year was actually about. A year of kitchen chemistry with cases and recoding is Chemistry, with a description. A year of news-hour conversation with no documents, no live-page departures, and no folder is not American Government just because the television was on.

¹⁷⁸National Center for Education Statistics, *Homeschooling in the United States: Results from the 2012 and 2016 Parent and Family Involvement Survey* (NCES 2020-001). Informal learning about 12 percent. Ever-taught high-school mathematics figures live in this series' mathematics records chapter; this chapter only needs the informal-learning share as a documentation fact.

A life of this work will not look like a badge. It will look like history, government, biology, statistics, and English, taught on purpose, with a folder a second adult could open. Chapter 10 names programs and helps by fit, without a sales pitch, so you can choose the next book without renaming it Critical Thinking 101.

Chapter 10

Programs, placement, and extra help

This book is not a curriculum. Civic Online Reasoning, Reading Like a Historian, Philosophy for Children, argument-mapping software, Gigerenzer-style risk literacy, and a handful of free handbooks are programs and tools you will actually meet. Name them. Compare grain, parent load, and where they sit in a subject. No homeschool randomized trial comparing them head-to-head was opened for this project, because none was found. Choose by fit: whether you can hear a wrong stay, a swapped percent, or a fluent un-checked paragraph in that program's work; whether the student will generate the attempt; special needs; budget; college target. A catalogue star is not a finding.

Pick a tool whose pictures you can learn well enough to notice a wrong turn. Place the student by skill in the *subject*, not by birthday, and not by a “critical thinking grade-level test.” When you are the ceiling for a year's government, biology, or statistics, a tutor, a co-op seat, or a community-college course is a normal high-school plan. National Center for Education Statistics figures already counted those valves among homeschoolers: any tutor about 23 percent; any local group or co-op about 31 percent; any online course about 23 percent (higher in the 9–12 band).¹⁷⁹ The kitchen is one adult, sometimes supplemented.

Exact list prices move. Confirm on the publisher's page before you buy. Several of the strongest civic objects below are free.

¹⁷⁹National Center for Education Statistics, *Homeschooling in the United States* (NCES 2020-001). Any tutor 23 percent; any co-op 31 percent; any online course 23 percent (about 30 percent of the 9–12 band). Mother was the main provider for 78 percent; formal curriculum 77 percent, in the same survey.

Common programs, by fit

Civic Online Reasoning (Stanford History Education Group / Digital Inquiry Group). Three questions on live pages: who’s behind the information, what’s the evidence, what do other sources say. Free classroom tasks.¹⁸⁰ Fit: government and history, middle and high school, live internet, constructed response. Parent load is real on previewing URLs and sitting beside the first departures; lighter once the student leaves without a nudge. The 2022 district study — six lessons of about fifty minutes in a required government course — moved credibility-task scores; university communications reported a rough doubling and still about half the points left on the table.¹⁸¹ That is a useful trial, not a promise your kitchen will match it. Steal the object. Put it inside civics or history, as Chapter 9 already required of the transcript. COR is not a stand-alone “media literacy diploma,” and it is not CRAAP under a new acronym.

Reading Like a Historian (SHEG / DIG). Document-based history: sourcing, contextualization, corroboration. Avishag Reisman’s 2012 urban high-school quasi-experiment found gains on historical thinking, transfer of those strategies to contemporary issues, factual knowledge, and reading comprehension.¹⁸² Quasi-experiment, not a randomized trial, not a generic-inventory result. Fit: United States and world history, ordinary high-school readers included. Parent load is the documents and the questions, not a franchise script. Title the year United States History. Name the method in the description.

Philosophy for Children (Lipman; UK SAPERE-style enquiry). A community of enquiry around a story, image, or prompt. Two large independent English trials tell different stories. Gorard, Siddiqui, and See (2015): 48 primary schools; small attainment gains, larger for pupils eligible for free school meals. The 2021 Education Endowment Foundation effectiveness trial: 198 schools, highest security rating, **0 months** extra progress on reading and maths, including for FSM pupils.¹⁸³ Teacher enthusiasm stayed high. Fit: an hour of dialogue you value as

¹⁸⁰Civic Online Reasoning, <https://cor.stanford.edu/>, fetched 30 August 2026. Free lessons; the site is not itself an outcome trial.

¹⁸¹Sam Wineburg et al., *Journal of Educational Psychology* 114, no. 5 (2022): 893–909; Stanford GSE news, 19 April 2022, for the doubling / remaining-half summary. Useful trial, not a kitchen guarantee.

¹⁸²Avishag Reisman, “Reading Like a Historian,” *Cognition and Instruction* 30, no. 1 (2012): 86–112. Quasi-experiment; 236 eleventh-graders; five San Francisco high schools.

¹⁸³Stephen Gorard, Nadia Siddiqui, and Beng Huat See, Education Endowment Foundation P4C

dialogue, usually inside English or a short seminar, not as an engine of generic thinking scores and not as a replacement for knowledge-rich subjects. Parent load is facilitation: wait, ask, keep the reason on the table. If you want P4C for the hour itself, the 2021 trial also showed no harm to attainment. If you want it to raise a Watson–Glaser, the trials were not even measuring that.

Argument mapping (Rationale and cousins; van Gelder’s line). Box-and-arrow claims, reasons, objections. Software-supported intensive practice is the modern claim. Tim van Gelder has been explicit that no large randomized trial of argument mapping versus matched traditional instruction exists; the empirical base is mostly pre–post, often on CCTST, in university semesters.¹⁸⁴ Fit: the argument-analysis *slice* of a logic, rhetoric, or philosophy elective, or as a method inside English and history. Parent load is learning the maps well enough to hear a restatement posing as a reason. Mapping is visible process — useful as Chapter 5’s process evidence, with a live check. It is not probability, diagnosis, or civic departure, and it should not be sold to you as the whole of Ennis’s sentence.

Control-of-variables / inquiry science. Chen and Klahr showed that the logic of unconfounded tests can be taught explicitly, in a simplified domain, to elementary students. A later meta-analysis of control-of-variables interventions reported a mean *g* of 0.61, with assessment format a major source of variance.¹⁸⁵ Fit: the science year you are already teaching. Parent load is the experiment, not a thinking workbook. Measure the move (can they design an unconfounded test?), not a generic inventory. Title the year Biology or Physical Science.

Risk literacy (Gigerenzer and colleagues). Natural frequencies: counts in a thousand rather than isolated conditionals. Papers and teaching restatements, including the Casscells item and the mammography absolute-risk example, are the source.¹⁸⁶

efficacy trial (2015): 48 schools; small KS2 attainment gains, larger for FSM. EEF / NFER effectiveness trial (2021): 198 schools; 5/5 padlock; 0 months on FSM reading and on cohort reading and maths. SAPERE training package. Two large independent RCTs; they do not agree on attainment.

¹⁸⁴Tim van Gelder, chapter in Davies and Barnett, *The Palgrave Handbook of Critical Thinking in Higher Education* (2015): no large RCT of argument mapping versus matched traditional instruction; pre–post pool, often CCTST. Developer-adjacent; not an independent meta-analysis in *Review of Educational Research*.

¹⁸⁵Zhe Chen and David Klahr, *Child Development* (1999), control-of-variables training. Marcus Schwichow, Corinne Croker, Christopher Zimmerman, Tim Höffler, and Hendrik Hartig, meta-analysis (2016): 72 intervention studies; mean *g* = 0.61 (95% CI 0.53–0.69) after excluding outliers. Scientific reasoning, not a general thinking course.

¹⁸⁶Gerd Gigerenzer et al., *Psychological Science in the Public Interest* 8, no. 2 (2008); Ward Cass-

Fit: biology, health, statistics, and family pamphlets. Parent load is drawing the thousand well enough to hear 95 percent as the wrong question. There is no home-school boxed set this research treated as a trial. You do not need one. Chapter 7 is the week. Elwyn's choice / option / decision talks are a clinic sequence you can role-play once the count exists.¹⁸⁷ They are a practice model, not a population-health result.

The Debunking Handbook 2020. Twenty-two-author consensus, free. Corrections are still worth making; worldview backfire is rare in later work; continued influence is the surviving problem; lead with the fact, name the myth, name the move, return to the fact, and offer a true cause.¹⁸⁸ Fit: family sharing rules and civic writing, not a course. Parent load is one evening's read. It will not replace leaving the page.

Inoculation / prebunking (short videos; Bad News game). Technique recognition — emotional language, false dichotomies, scapegoating, ad hominem — with lab effects in the medium range and a small YouTube field effect, decaying without boosters.¹⁸⁹ Fit: a Friday supplement, with a later revisit, never a substitute for COR in government. Parent load is low on a ninety-second video and real on scheduling the booster. A one-shot assembly is not this literature.

Khan Academy and other free maps. Useful as practice in the *subject* (civics, statistics, biology). Not a complete homeschool program of record unless you build the record: titles, grades, credits, a description of what was actually checked. A chatbot layer over practice, where it exists, is a hint tool only if you stay in the room and withhold the finished answer. Chapter 8 already named that split.

Others you will see named. Paul–Elder wheels, elements, and standards are a curriculum brand. This book stole, in earlier chapters, the 1981 warning that training on “neutral” cases can make students more sophistic rather than less so, and then left the franchise.¹⁹⁰ Watson–Glaser and CCTST prep books train the sitting.

cells, Arno Schoenberger, and Thomas Graboys, *New England Journal of Medicine* 299 (1978); Ulrich Hoffrage and Gigerenzer, *Academic Medicine* 73 (1998).

¹⁸⁷Glyn Elwyn et al., *Journal of General Internal Medicine* 27 (2012): 1361–67. Practice model.

¹⁸⁸Stephan Lewandowsky, John Cook, Ullrich K. H. Ecker, et al., *The Debunking Handbook 2020*, <https://doi.org/10.17910/b7.1182>.

¹⁸⁹Jon Roozenbeek, Sander van der Linden, et al., *Science Advances* (2022); Rakoen Maertens et al., *Journal of Experimental Psychology: Applied* (2021).

¹⁹⁰Richard Paul, 1981 strong-sense / weak-sense warning, as used in this book's earlier chapters: steal the warning; leave the franchise. This resources chapter does not rank or sell Paul–Elder materials.

Chapter 5 already told you what a sitting is. Checklist media-literacy packets that score .org, typos, and About pages are the contrast case Chapter 6 retired as the week's object. NAMLE's Core Principles are a professional association's framework, not a trial.¹⁹¹ Twenty-five state media-literacy laws are laws; they have not been evaluated as classrooms.¹⁹² First-year-college "thinking across the curriculum" stamps did not add a reliable increment over ordinary college in Huber and Kuncel's synthesis.¹⁹³ Accountable Talk and other discussion designs can increase student talk; more talk is not automatically more judgment.¹⁹⁴

Public-school virtual and charter-at-home are often public-school enrollments with a teacher of record and state tests. They are not the parent-as-teacher kitchen. Co-ops, umbrella schools, cottage schools, and microschools are legal and social wrappers. Umbrella schools can issue transcripts. They do not magically install lateral reading.

Placement tests exist because the labels lie

A "grade 8 critical thinking workbook" is a publisher's scope, not a legal grade. A child of twelve may source well in a known history unit and invert a clinic percent the same afternoon. That is not a contradiction. It is two domains.

Place by the subject's own gate.

- Before live-web government work: can they read the search results, and do they know what a trade association, an ad, and a primary source are? If not, teach those as content, then leave the page.
- Before natural-frequency trees: can they take a fraction of a thousand and say "percent of" versus "percent off"? If not, the math year is the placement, not a thinking workbook.
- Before closed-window defense of a generated paragraph: can they already leave a website and recode a percent in that subject? The loop in Chapter 8

¹⁹¹National Association for Media Literacy Education, Core Principles (2023 revision). Framework, not a trial.

¹⁹²Media Literacy Now policy map (updated January 2026): 25 states; MLN has not evaluated classroom implementation.

¹⁹³Christopher R. Huber and Nathan R. Kuncel, *Review of Educational Research* 86, no. 2 (2016).

¹⁹⁴P. Karen Murphy, Ian A. G. Wilkinson, Anna O. Soter, Maeghan N. Hennessey, and John F. Alexander, *Journal of Educational Psychology* 101 (2009): 740–64. Several discussion approaches increased student talk; few increased critical-thinking/reasoning outcomes as measured.

sits on those moves.

- Before a philosophy enquiry hour: can they stay on one question long enough to give a reason? If not, shorter waits in ordinary English discussion are the on-ramp.

Willingham's age-lock refusal, in spirit: what is thinkable depends mainly on prior opportunity to learn in the domain, not on being "too young for critical thinking." Even young children can think critically in a domain they know; trained adults fail outside theirs.¹⁹⁵ Accelerate a student who has the prerequisites in *this* subject. Slow a student who does not. Skipping civic knowledge because a child is "good at arguing" is not acceleration.

New York plans and some dual-enrollment offices still want a grade. NCAA wants a ninth-grade start date. You can organize by skill *and* still export a grade-labelled transcript. Chapter 9 is that export. This chapter's job is to let the subject speak before the birthday does, and before a thinking-test percentile does.

When a tutor or a college course helps

You do not have to be a historian, a biologist, or a statistician. You do need today's idea well enough to hear a stay, an inversion, or a fluent un-checked paragraph. When that hearing is not yet available for this year's course, extra help is ordinary.

A weekly tutor is help. Unless the tutor's institution issues a transcript, you remain the grader and the administrator of record. Name the tutor on a course description if you want the help visible. Hire in the *subject* — government, biology, statistics, English — not a generic thinking coach with a prep book.

A co-op class with a named instructor, a named text, and a grade the instructor assigned is easier to describe than a year you are not sure how to narrate. You still sign as administrator unless an umbrella school does.

Dual enrollment is a college course that produces a college transcript. American government, English composition, a lab science, introductory statistics: admissions offices already know how to file those. Cost ranges from free, in some state programs, to ordinary community-college tuition. State dual-credit laws vary. Keep the syllabus. List the course at home. Send the college transcript.

¹⁹⁵Daniel T. Willingham, "Critical Thinking: Why Is It So Hard to Teach?," *American Educator* 31, no. 2 (Summer 2007). Domain knowledge, not a birthday, is the gate.

A *hypothetical* illustration, labelled as such: a parent who last took civics years ago enrolls the student in a community-college American government course, remains the homeschool administrator of record, and still runs two live-page departures at home so the COR habit is practiced, not only lectured. The rustiness is hypothetical. The 23 percent / 31 percent tables are not. That plan is a normal high-school year.

Handing the same student an unconstrained chatbot in place of the tutor is not the same valve. Chapter 8 named the split: formation versus performance; hints versus finished sentences; live defense with the window closed.

Special education at home is not an IEP with a public-school duty in the way public school is. If you are the ceiling for a learner whose bottleneck is reading, buy reading expertise rather than a media-literacy app. If the bottleneck is mathematics for the thousand-person tree, buy math help. The evaluation habit will wait on the decoding.

A short chooser, not a ranking

- If you want the civic habit with trial evidence, Civic Online Reasoning is built that way. Put it in government or history. Preview the pages. Leave before you read.
- If you want document-based history, Reading Like a Historian is built that way. Title the year as history.
- If you want an enquiry hour and can live with 2021's zero extra months on attainment, Philosophy for Children is built that way. Value the hour as the hour.
- If you want visible argument structure, mapping software is built that way. Check the maps live. Measure that slice.
- If you want clinic and pamphlet numbers, draw Gigerenzer's thousand inside biology or statistics. You already have Chapter 7.
- If you want a sharing rule for the house, the Debunking Handbook plus Chapter 6's three questions are enough for a week of evenings.
- If you need free subject practice, Khan Academy and SHEG's free lessons are built that way. You still write the record.

None of those bullets is "best." Fit is whether you can hear a wrong answer in that program's representations, and whether the student will generate the work. When

the student has unfinished subject knowledge, go back, whatever the calendar says. When you are the ceiling, buy a tutor, a co-op seat, or a community-college transcript. When the destination is a stranger with a transcript in hand, title the year as the subject — Government, History, Biology, Statistics, English — and let the description carry the habits. A life of deciding what to believe or do will not arrive as a franchise. It will arrive as Tuesday, in a domain, with a check.