TEACHING LIFE SCIENCES
IN THE AGE OF AI

Fast Tools, Slow Thinking

What building with AI taught me about where teaching is going



Chris Deppmann · UVA Biology

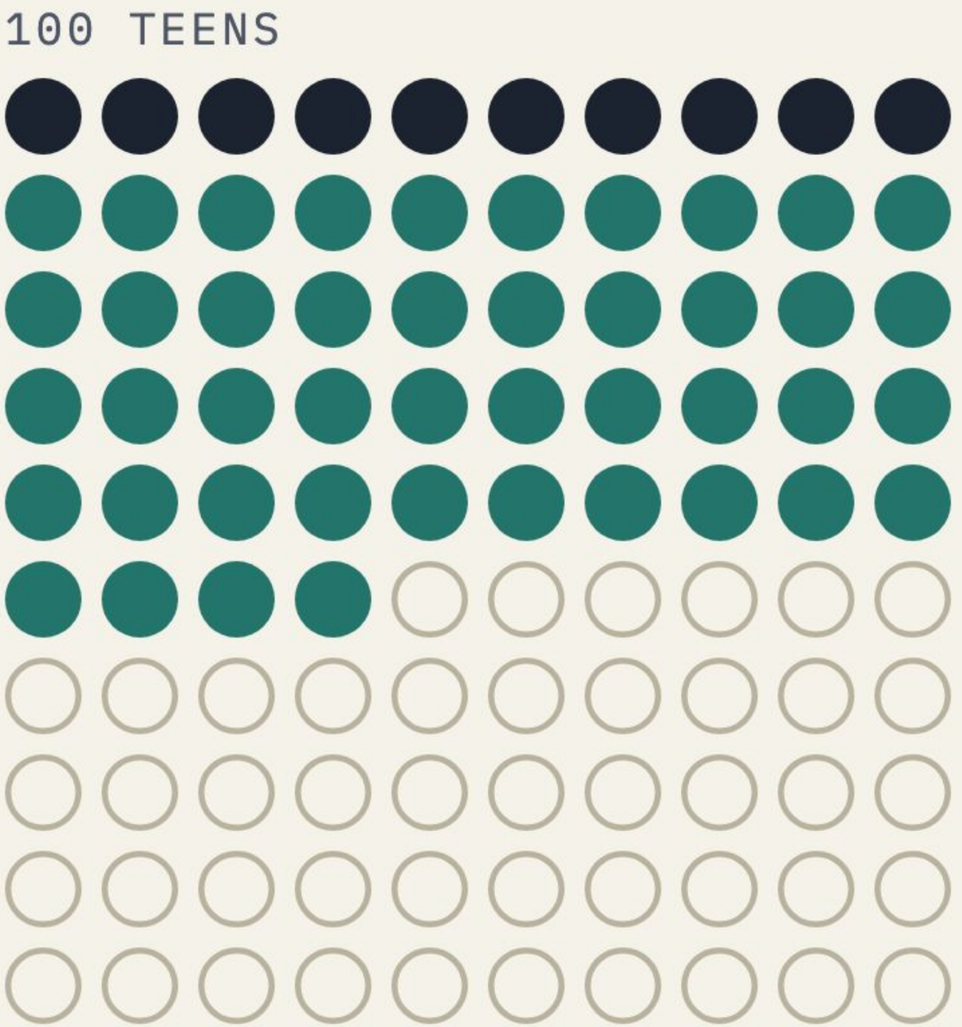
Illustrations are AI-generated.

YOUR STUDENTS

More than half of teens already use AI for schoolwork.

54%

of US teens used an AI chatbot for schoolwork, fall 2025.



Pew Research Center

WHERE I'VE LANDED

Fast tools for us. Slow thinking for them.

FAST TOOLS
· for teachers

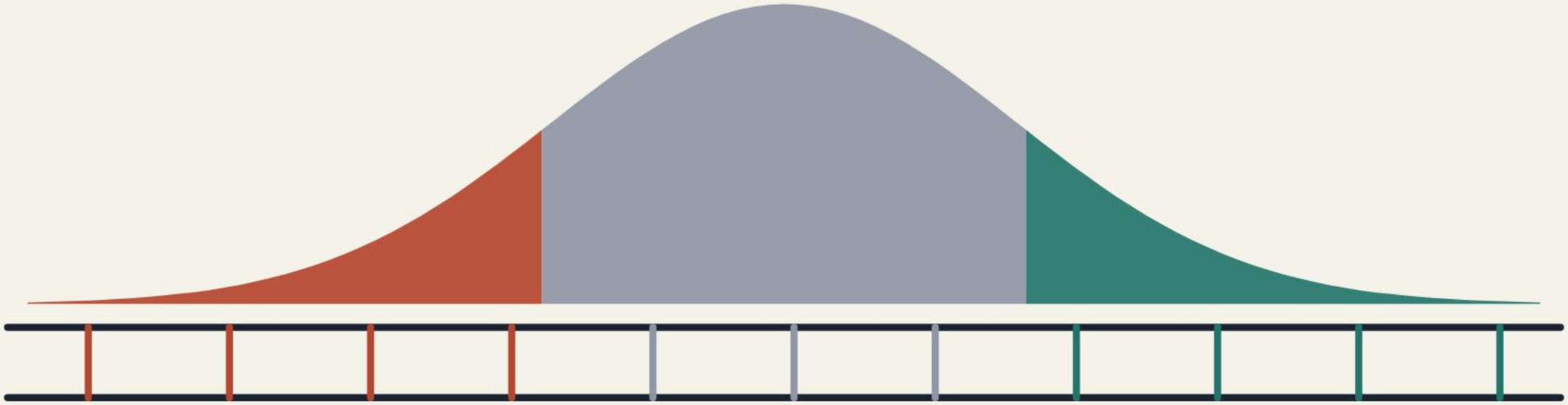
Teachers are essential. Any AI tutor should be under the teacher's control, with every exchange visible to the teacher.

SLOW THINKING
· for students

Leaning on AI hurts students, especially young ones. They need slow thinking to make an idea their own.

THE PROBLEM

Every class spans a ladder. We teach to the middle.



← LOWER RUNGS

HIGHER RUNGS →

LEFT TAIL
Needs more help to reach the standard.

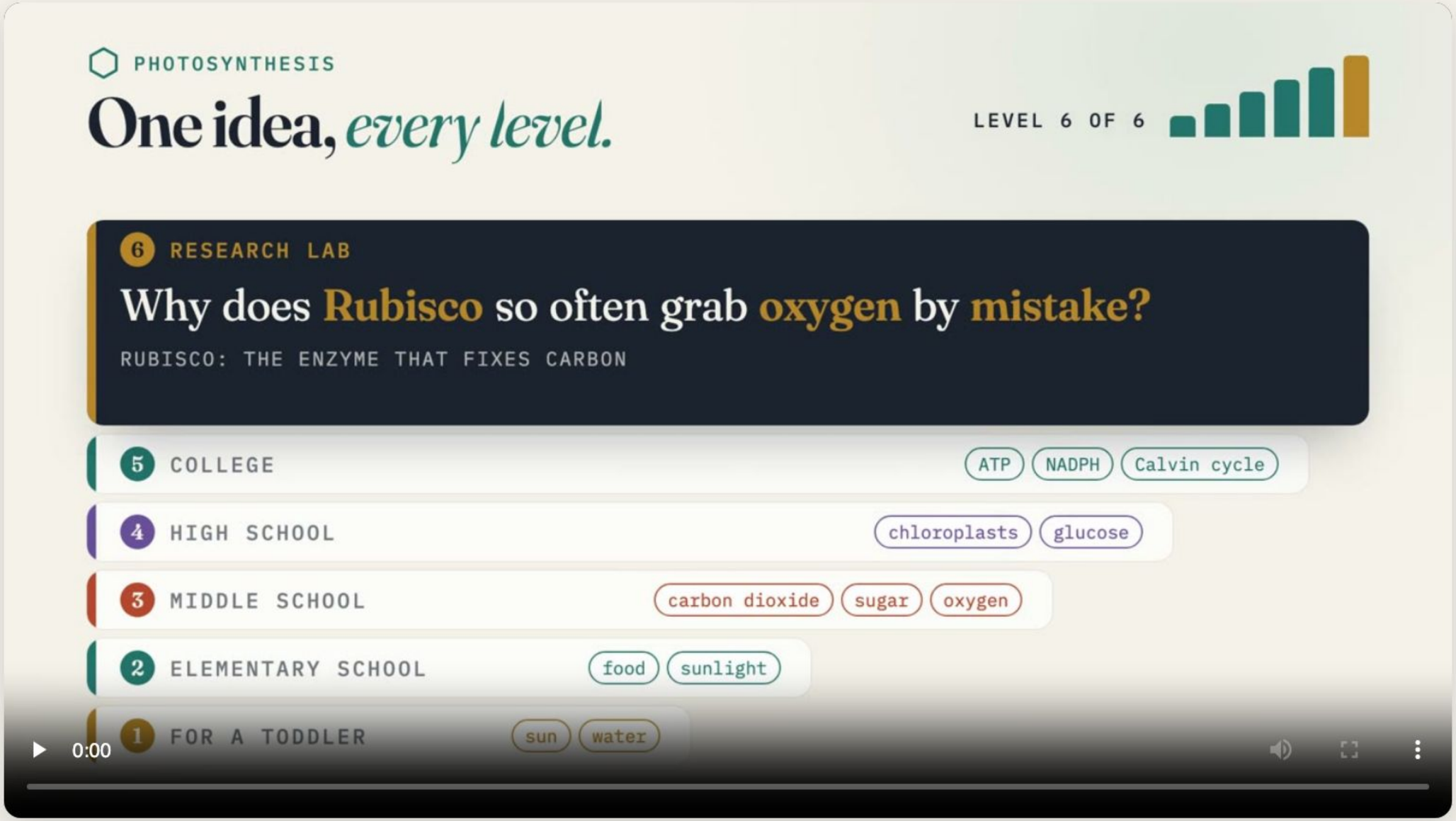
MIDDLE
Where the lesson is aimed.

RIGHT TAIL
Ready to go much further.

Both tails are hard to reach with **one teacher's attention**.

WATCH 55 SECONDS

The same idea lives on every rung.

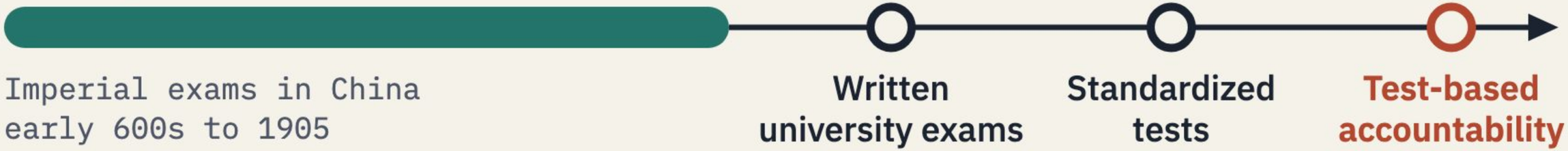


WHY WE TEACH TO THE TEST

When a measure becomes a target, it ceases to be a good measure.

Marilyn Strathern (1997), about exams, after Charles Goodhart (1975)

More than a thousand years of high-stakes exams



CHEAP TO MEASURE

The answer

Quick to grade.

teaching to the test

WHAT WE CARE ABOUT

The understanding

Slow to read.

WHERE WOULD THE ATTENTION COME FROM?

Reaching every rung takes attention nobody has.

49

hours a week

RAND, 2025

$895

of your own money, every year

AdoptAClassroom, 2025

22.6%

of US students chronically absent

2024–25 school year

> **One teacher’s attention is the limit.**

WHAT WE MEAN BY AI

Most of us still picture a chatbot. That picture is already out of date.

2022



A chatbot

You ask. It answers.

one turn at a time

WHAT WE MEAN BY AI

Most of us still picture a chatbot. That picture is already out of date.

2022

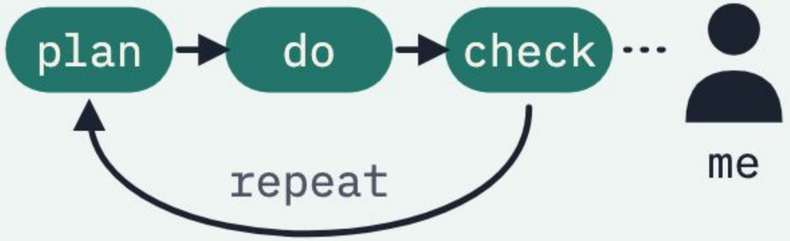


A chatbot

You ask. It answers.

one turn at a time

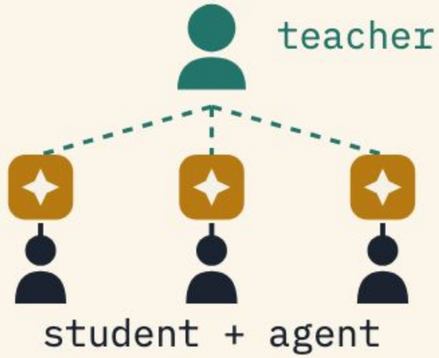
NOW



Agents

> How I built the site: agents drafted, agents fact-checked, I reviewed.

NEXT



An agent for every student

> **The teacher sees every exchange.**

a direction, not a date

PART 1 OF 2

What I built, and what it taught me

What happens when making materials stops being the hard part?



WHAT AGENTS CHANGED

The next 324 lessons took 13 days.

17 weeks



30 lessons · Jun 12 to Oct 10

13 days



bars drawn to scale

324 lessons · Oct 11 to 23

That’s the difference agents made.

Today: 354 lessons · 11 courses · 464 films · **free**

WATCH 60 SECONDS

Here’s the experiment.

biochemistrypedia.com

A switch

sets how much help.

The task stays the same.

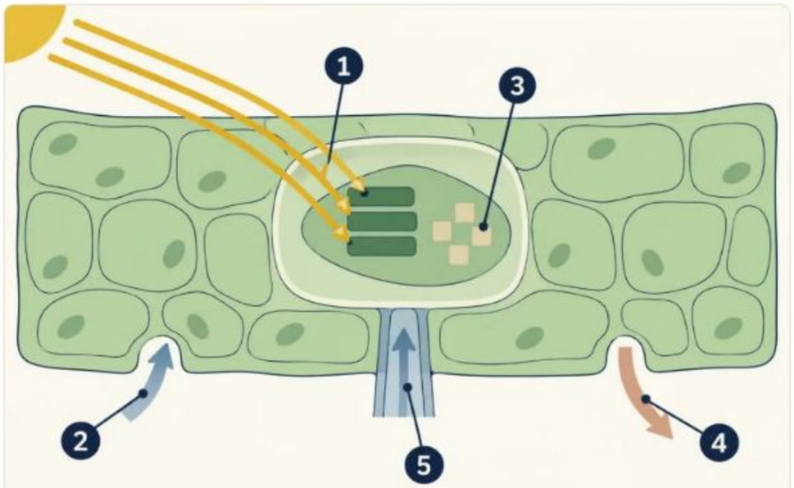
SUPPORTS

- ○ **Just the task** — No extra help
- ○ **Some help** — Word banks, partly labelled diagrams
- ◉ **Most help** — Adds sentence starters, fully labelled diagrams

Supports change how much help you get, not what you learn.

BLANK | PARTLY LABELLED | FULLY LABELLED

CAPTURING SUNLIGHT · BIO.2E



1 light 2 carbon dioxide 3 glucose 4 oxygen 5 water

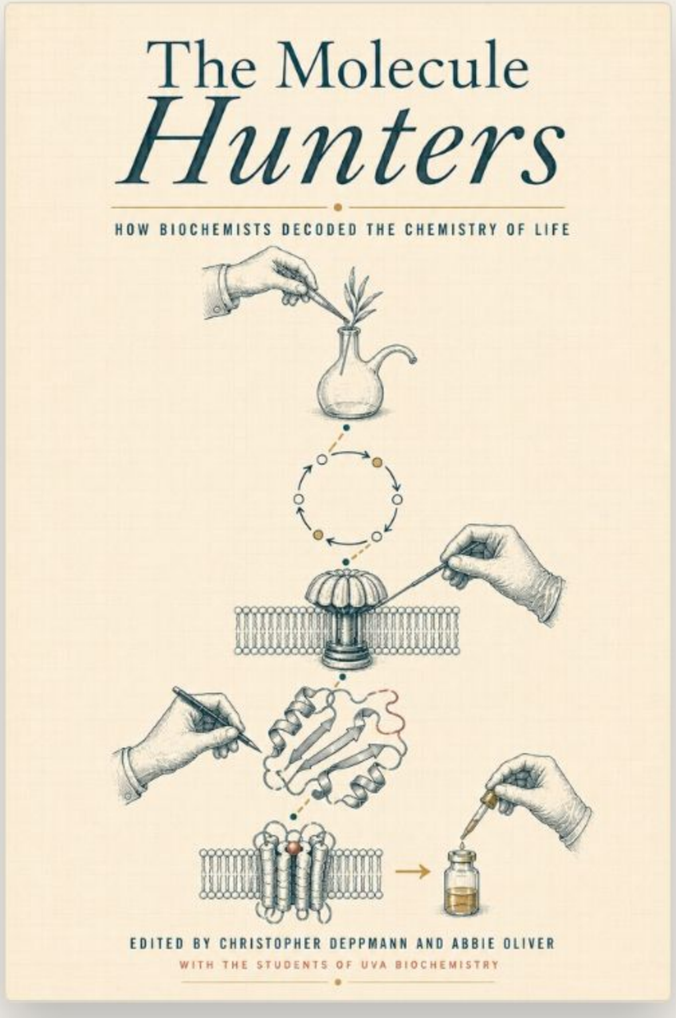
Inside one leaf cell, the chloroplast captures light and uses materials from the air and roots to make sugar and release a gas.

SAME TASK AT EVERY LEVEL

▶ 0:00 / 1:00

ANOTHER WAY I'VE USED AI

500 students wrote a book with me.



133 profiles

190 scientists

500+ UVA student co-authors

1 Every profile began as an AI deep-research dossier.

2 Free as a PDF, an EPUB and an audiobook.

3 Errors are listed in public, on a live fact-check page.

`The Molecule Hunters · deppmannlab.com/book`

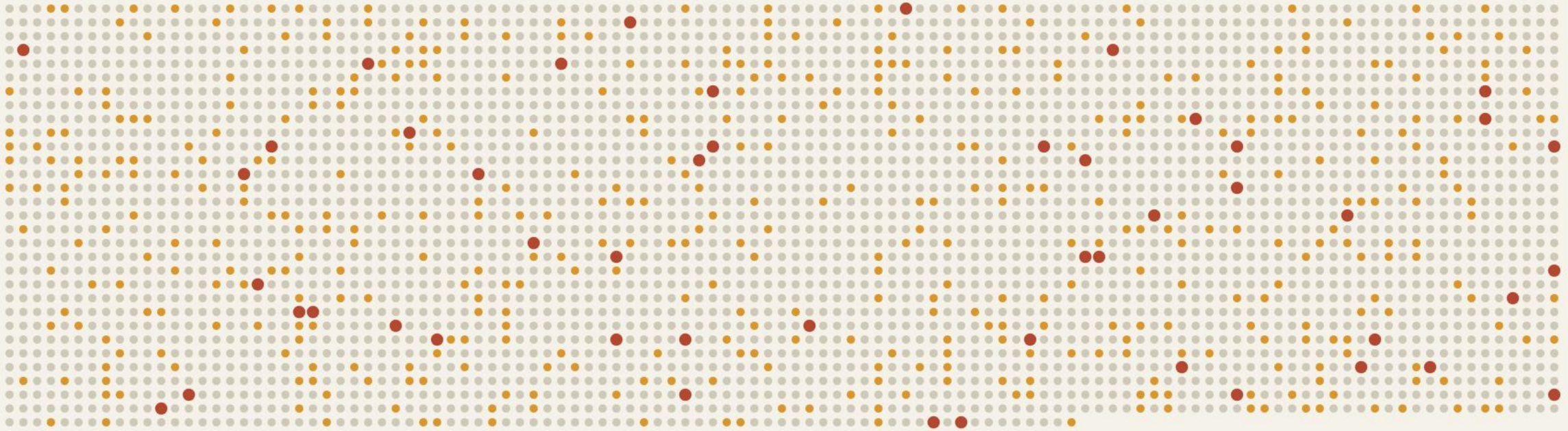
WHAT BUILDING TAUGHT ME

Each thing I built taught me something I didn't expect.

	I BUILT		IT TAUGHT ME
1	354 free lessons	→	Making is easy. Checking is the real work.
2	Pre-work students do before class	→	I could finally see how they think.
3	AI-written exam questions	→	The wording can be ambiguous.
4	A place for their questions	→	Their best questions went unread.

Each answer raised the next question.

Making got cheap. Checking is the job.



49 flat wrong

453 imprecise

of 3,468 claims
in one course

> As we improved the process, wrong claims fell from **4 per lesson to under 1**.

Now I can see how they think.

THURSDAY MORNING



2 finished

FRIDAY, 8:40 AM



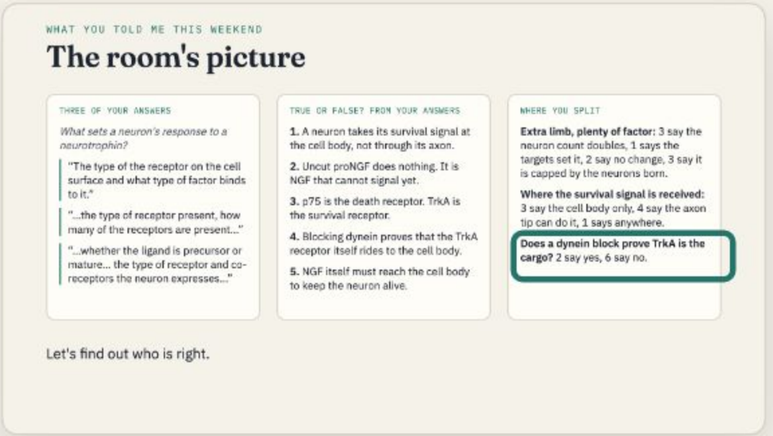
9 finished

+ 3 partly done

■ finished ◧ partly □ not yet

One grad course, 12 students
one 3–4 hour module

A REAL ROOM SLIDE



WHAT YOU TOLD ME THIS WEEKEND

The room's picture

THREE OF YOUR ANSWERS

What sets a neuron's response to a neurotrophin?

> "The type of the receptor on the cell surface and what type of factor binds to it."

> "...the type of receptor present, how many of the receptors are present..."

> "...whether the ligand is precursor or mature... the type of receptor and co-receptors the neuron expresses..."

TRUE OR FALSE? FROM YOUR ANSWERS

1. A neuron takes its survival signal at the cell body, not through its axon.
2. Uncut proNGF does nothing. It is NGF that cannot signal yet.
3. p75 is the death receptor. TrkA is the survival receptor.
4. Blocking dynein proves that the TrkA receptor itself rides to the cell body.
5. NGF itself must reach the cell body to keep the neuron alive.

WHERE YOU SPLIT

Extra limb, plenty of factor: 3 say the neuron count doubles, 1 says the targets set it, 2 say no change, 3 say it is capped by the neurons born.

Where the survival signal is received: 3 say the cell body only, 4 say the axon tip can do it, 1 says anywhere.

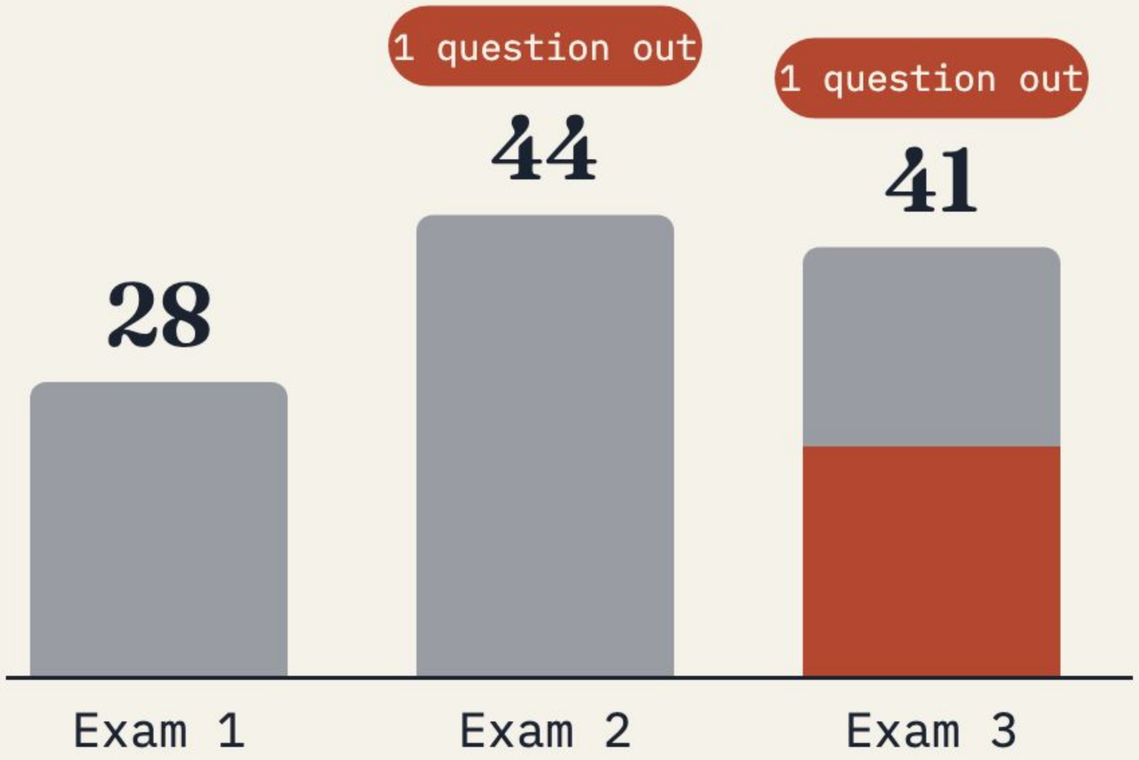
Does a dynein block prove TrkA is the cargo? 2 say yes, 6 say no.

Let's find out who is right.

Class opens with their own words: 2 say yes, 6 say no.

AI-written exams can be ambiguous.

Review requests per exam · ~735 students



Exam 3: 22 of 41 on one question

Out = full credit for everyone.

Students flagged it.

A person reviewed it.

The process changed.

A red-team review was added. “All of the above” was banned.

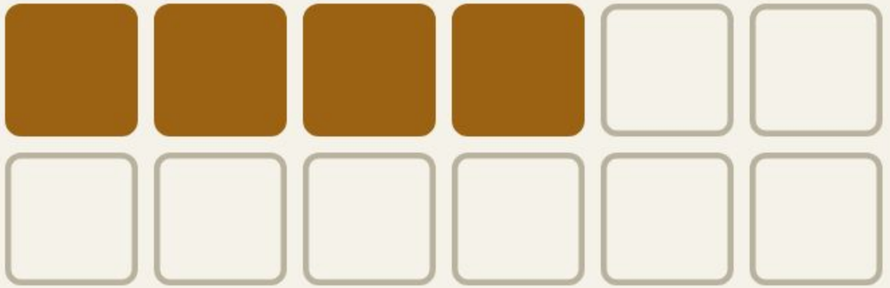
AI keeps getting better at clean wording and the right level. **It isn’t there yet.**

Their best questions went past the lesson.

WHAT THEY ASKED

- *a drug for this*
- *cancer*
- *dopamine and reward*

WHO ASKED



Four of 12 students asked most of them.



This is the right tail asking to go further.

One grad course, 12 students

PART 2 OF 2

Where this is going

Five shifts, each one about reaching the whole ladder.



Meet every student where they are.

NOW

In one grade, students' levels vary widely. Software that taught each student at their own level raised math scores **0.36 SD** in 4.5 months (Delhi, 2019). One study; mixed elsewhere.



BOTH TAILS

NEXT

Every student starts on their own rung: catching up from below, racing ahead from above, checked as they go — and the teacher told who's slipping.

Let curiosity lead the right tail.

NOW

Interest in science falls from **9 to 17**. Curiosity questions drop from about **2.4** (kindergarten) to about **0.5** (grade 5) per two hours.



RIGHT TAIL

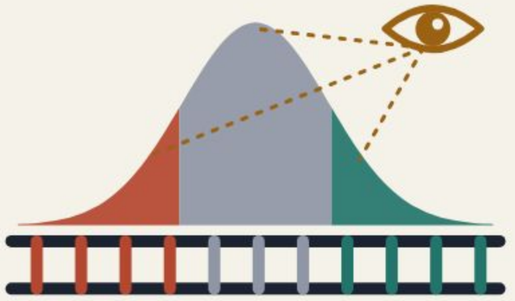
NEXT

Keep teaching the core. A student's own question can become an independent study, built on the fly — and exploring earns credit.

AI carries the volume, so your attention goes where it's needed.

NOW

Teachers who use AI every week save about **5.9 hours** a week, mostly on prep.



YOUR ATTENTION

NEXT

AI reads every answer and builds the review or the stretch. Each morning you see which few students need you, and why.

A tutor the school owns, so you can see the thinking.

NOW

Unguarded GPT-4: practice up **48%**, exam scores down **17%** once it was removed. A guarded tutor didn't do that harm (PNAS 2025).

NEXT

The school gives every student a flexible tutor that asks before it tells. Because the school owns it, the teacher sees how each student uses it and thinks.

Students bringing their own AI is a different problem. I haven't solved it.

How they got there counts more than the answer.

NOW

We grade the answer, and a chatbot now gives any answer. Chegg's revenue fell **39%** in a year.

NEXT

The prediction, the reasoning and the revision become what we look at. The score becomes a by-product.

HOW WE BREAK THE HABIT

When the answer is free, understanding is the only thing left worth grading.

> Goodhart's law bites hardest on **one narrow, high-stakes number**.

1



Measure the thing itself.

The reasoning, not a stand-in for it.

2



Many small looks, not one big test.

Low stakes, so nothing is worth gaming.

3



A person reads it.

Judgment, not just a score.

HONESTLY

State tests and the MCAT aren't going away soon. If understanding is what we look at every day, the test becomes a by-product.

HONEST ABOUT THE RISKS

It can go wrong, sometimes quietly.

1. Free tools help the already-engaged most. Devices and broadband aren't equal.
2. Over-reliance, for students and for us: if we trust the summary, our own judgment gets weaker.
3. A rung can harden into a label. Placement has to stay fluid.
4. AI is wrong sometimes: a summary said a student had done nothing; an exam question can be ambiguous.
5. A record of how a child thinks is sensitive: who owns it, how long it's kept, breaches.
6. Goodhart applies to AI too: optimize engagement or scores, and it bends the same way.
7. One 2026 trial (not peer-reviewed): a course AI tutor lowered grades, most for first-generation students.

THE RULE

AI reads. You decide.

WHERE I LAND

Fast tools for us. **Slow thinking** for them.

What I'd ask of any AI in a classroom

1. **The student tries first.**
   It never gives the answer before they commit.
2. **The teacher sees every exchange.**
3. **It meets each student at their level.**
   And tells the teacher who is falling behind.
4. **It looks at reasoning.**
   And gives second chances.
5. **Everything it makes is checked.**
   Mistakes are fixed in public.

We can do better.

My hope is that this is an enabling technology.

biochemistrypedia.com
/teachers