Chapter 4: The Tutoring Advantage

In 1984, a University of Chicago education professor named Benjamin Bloom published a short paper with an unusual title: “The 2 Sigma Problem.” It has been quoted by educators ever since, and once you understand it, you’ll see why it sits at the center of this book.

Two University of Chicago doctoral students, Anania and Burke, had run a set of experiments. They randomly assigned students to learn the same material in one of three ways. The first was the conventional classroom: about thirty students per teacher, with tests given mainly to hand out grades. The second was what Bloom called mastery learning: the same kind of class, but with regular check-in tests, and extra help for anyone who hadn’t mastered the material before the class moved on. The third was one-to-one tutoring, combined with the same mastery approach.

The results, Bloom wrote, astonished them. The average student who was tutored scored about two standard deviations above the average student in the conventional class. Put in plain English, the average tutored student outperformed 98 percent of the students taught the ordinary way.

Think about what that means. Take a completely average child, a kid in the middle of the class. Give them a tutor, and they don’t move up a little. They move to the top of the room.

The problem Bloom actually posed

Bloom didn’t stop at the good news. He knew schools could not afford a tutor for every child. So he framed a challenge: could researchers and teachers find ways of teaching groups that would get most students close to what tutoring achieves? That was the “problem” in

the title. The two sigma result was the target. The problem was how to reach it without one adult for every child.

For forty years, that problem has mostly gone unsolved. Not because anyone doubted the value of tutoring, but because of simple arithmetic. There are about thirty kids in the room and one teacher. Individual attention is the most powerful thing we know of in education, and it has always been the most expensive.

And, as we saw in Chapter 2, it's the thing wealthy families buy. The two sigma gap is not just an education statistic. It's a meritocracy statistic.

What the modern evidence says

I want to be careful here, because Bloom's two sigma number gets thrown around as if it were a law of nature. It isn't. His studies were small and carefully controlled, and later research has generally found smaller effects. The honest way to use Bloom is as a direction, not a promise.

So what does the newer research say? Quite a lot, and most of it is encouraging.

In 2020, the economists Andre Nickow, Philip Oreopoulos, and Vincent Quan pulled together the experimental research on tutoring from pre-K through twelfth grade. They found that tutoring programs produce "consistent and substantial positive impacts on learning," with an average effect of about 0.37 standard deviations. That's not two sigma, but in education research it's large. Effects were strongest when tutoring was done by teachers or trained paraprofessionals, and in the earlier grades.

In 2024, a team led by Matthew Kraft asked a harder question: what happens when you scale tutoring up? Looking at more than 260 randomized trials, they found that the big averages are driven partly by small programs. For programs serving 400 to 999 students, the average effect was about 0.21 standard deviations. For programs serving 1,000 or more, it was about 0.16. Smaller, but, as the researchers themselves put it, still of “considerable policy importance.”

That finding matters for everything that follows in this book. Tutoring works. But it gets harder to do well at scale, because it depends on finding, training, and supervising enormous numbers of skilled people. The bigger the program, the harder it is to keep quality high.

Proof that it can work in public schools

Some of the most encouraging evidence comes from public schools that serve kids the system usually fails.

In Chicago, a nonprofit called Saga Education has provided daily, in-school math tutoring to ninth and tenth graders, usually two students to one tutor, since 2012. Rather than hiring certified teachers, Saga recruits recent college graduates, career changers, and retirees, and trains them to do one focused job. The University of Chicago Education Lab has studied the program in several large randomized trials with thousands of students, and found it improves students’ math learning. The details vary across the studies, but the lesson is consistent: when adolescents who are behind get daily, focused, personal instruction during the school day, they catch up in ways that many people had assumed was too late.

That is what a system can do when it treats individual attention as a priority instead of a luxury. The obstacle is that people cost money, and good people are hard to find by the tens of thousands.

Where AI enters

This is where artificial intelligence changes the math. Not by replacing the tutor, at least not yet, but by making good tutoring easier to deliver at scale.

The clearest evidence so far comes from a 2024 study at Stanford called Tutor CoPilot. Researchers Rose Wang, Susanna Loeb, Dora Demszky, and their colleagues ran a randomized trial with about 900 tutors and 1,800 K-12 students from Title I schools, the schools that receive federal money because they serve many low-income families. Some tutors were given an AI assistant that suggested, in real time, how an expert teacher might respond: ask a guiding question, give a hint instead of the answer, encourage the student.

Students whose tutors had access to the AI were 4 percentage points more likely to master the math topics they were working on. For the least experienced, lowest-rated tutors, the gain was 9 percentage points. In other words, the AI helped most exactly where help was needed most: making weaker tutors perform more like strong ones.

That is a modest result in a single study, and I don’t want to oversell it. But look at what it points to. The hardest part of scaling tutoring is the shortage of skilled people. AI’s first proven contribution is making ordinary people more skilled. That is a solution to Bloom’s problem that Bloom never had.

Two roads

So there are two ways AI can help close the tutoring gap, and I think we should pursue both.

The first is augmentation. AI makes human tutors and teachers better, as in the Tutor CoPilot study. It suggests the next question, spots the misconception, drafts the practice problems, and frees the adult to focus on the child.

The second is direct tutoring. AI works with the student itself, patiently, at the student's pace, for as long as the student needs. That is the promise people get most excited about, and it's also where the risks are greatest. A chatbot that simply hands kids the answer isn't a tutor. It's a shortcut. In the next chapter, we'll look at what the research on AI tutors really shows, including one study where a generic chatbot helped students during practice and then hurt them on the test.

For now, I want you to hold on to one idea. For forty years, the most powerful tool in education has been locked behind a price tag. The families who could pay for it, paid. Everyone else waited. For the first time, we have a realistic chance to unlock it for every child. The question isn't whether that chance is real. It's whether we'll take it seriously enough to do it right.

Chapter 5: What the New AI Evidence Actually Says

You may have heard that a Harvard study proved AI teaches better than professors. There was a Harvard study. It's real, and it's remarkable. But what it proves is more specific than that, and more useful.

This chapter is about what the new research on AI and learning actually shows. I'm going to be careful, because this is where books like mine usually get into trouble. It's tempting to grab the most exciting headline and run with it. The critics of AI in schools will read this chapter closely, looking for exactly that. I'd rather give them nothing to find.

There are two studies I think every parent, teacher, and policymaker should know. They point in what looks like opposite directions. Read together, they tell the most important story in this book.

The Harvard physics study

In 2025, Greg Kestin and his colleagues at Harvard published the results of a randomized controlled trial in the journal Scientific Reports. The setting was an undergraduate physics course, and the comparison was a tough one. Harvard's physics course was already using active learning, a well-regarded method where students work through problems in small groups instead of just listening to a lecture. This wasn't AI against a bored professor reading slides. It was AI against good teaching.

For two lessons, some students learned the material in class with active learning, and others learned the same material at home with an

AI tutor the researchers had built. Then they switched, so each group experienced both.

The students using the AI tutor learned significantly more. Their median learning gain was more than double that of the students in the active-learning class. And they did it in less time: the median student spent 49 minutes with the AI tutor, compared with a 60-minute class. The students also reported feeling more engaged and more motivated.

That is an extraordinary result. But look closely at what was tested.

This was not a student typing questions into ChatGPT. The researchers built the tutor carefully. It was guided by detailed instructions written by the course's own teachers, grounded in step-by-step solutions they had prepared, and designed to lead students through the reasoning rather than hand them the answer. It was used by Harvard undergraduates, for two lessons, in one subject.

So what the Harvard study proves is this: a well-designed AI tutor, built by experts around good teaching principles, can help students learn more in less time than even a strong classroom lesson. That is a big deal. It is not proof that any chatbot beats any teacher, or that the result would hold for a nine-year-old learning to read.

The crutch study

The second study is the one I think about most.

In 2025, a team led by Hamsa Bastani of the Wharton School published a field experiment in the Proceedings of the National Academy of Sciences. They gave nearly a thousand high school math students access to AI during practice sessions. Some got a version that worked like standard ChatGPT. The researchers called it "GPT Base." Others got a version designed with safeguards to protect learning,

which they called “GPT Tutor.” It gave hints and guided students instead of simply supplying answers.

During practice, both versions helped. In the researchers’ earlier working paper, students using GPT Base scored 48 percent better on practice problems, and students using GPT Tutor scored 127 percent better.

Then the researchers took the AI away and tested the students on their own.

The students who had used the standard chatbot did worse than students who had never had AI at all. In the working paper, their scores were 17 percent lower. They had used the AI, in the researchers’ word, as a “crutch.” It did the thinking for them during practice, and when the crutch was gone, they hadn’t learned to walk.

The students who had used the tutor with safeguards didn’t suffer that harm. The protections built into the design largely prevented it.

What the two studies say together

Put these two studies side by side, and the lesson is almost too clear.

The question is not whether AI is good or bad for learning. The question is how it’s designed and how it’s used. An AI that hands a child the answer is a shortcut, and shortcuts can make children weaker. An AI that asks the next question, gives a hint instead of a solution, and makes the child do the thinking can make them stronger, and faster.

That is the argument at the heart of this book: the problem isn’t the technology, it’s the usage. I was glad to find it’s also what the evidence says.

And notice something else. The harmful version in the Bastani study was the one kids get by default: an open chatbot on a phone or laptop. That's the version many students can already reach at home, on their own, with no guardrails and no teacher involved. Keeping well-designed AI out of schools doesn't keep AI away from kids. It just means the only AI they use is the kind that can hurt them.

What we don't know yet

I promised to be honest, so here is what the research can't yet tell us.

We don't have long-term evidence. Most of these studies last a few weeks or a few lessons. We don't know yet what happens after a full year, or five years, of learning with a good AI tutor.

We don't have much evidence for young children. The Harvard study was about college students. The Bastani study was about high schoolers. Early reading, the skill that everything else is built on, is a different challenge, and it deserves its own careful research.

We don't know how well carefully designed tutors hold up when a whole district uses them, with all the messiness of real schools. As we saw in Chapter 4, even human tutoring gets harder to do well at scale.

And we don't have independent, peer-reviewed evidence for some of the most dramatic claims being made about AI-powered schools. I'll come back to that when we talk about the school models that are already trying this.

None of that is a reason to wait. It is a reason to move carefully, measure everything, and use the kind of AI the evidence supports. That is exactly what I'll propose at the end of this book.

The takeaway for parents

If you remember one thing from this chapter, make it this. When someone tells you AI will make kids lazy and dumb, they're describing the crutch. When someone tells you AI will transform learning, they're describing the tutor. Both are real. They are not the same product.

The debate about AI in schools keeps treating them as one thing. Our job, as parents and citizens, is to insist on the difference, and to demand that our public schools give children the tutor instead of leaving them alone with the crutch.

In the next chapter, I'll take on the strongest version of the case against technology in schools, the one made by Jared Cooney Horvath in The Digital Delusion. He is right about more than you might expect.