



A Friend on My Side: What a Sixth Grader Taught Us About AI Math Coaching at Scale

Michael Yuen , StarCoach AI

In a three-week pilot across three middle schools in a Washington district, an AI math coach designed to affirm every attempt—and reward work with a game—got students who avoided math to seek it out after school and on weekends. Scores rose about 10 points among those who finished.

Digital Object Identifier 10.1109/MC.2026.3705124

It was three o'clock on a quiet Wednesday. I was on a Zoom call with Dana, a middle-school math teacher in a Washington school district, debriefing the first weeks of a pilot of our AI math coach. Behind her, the classroom was empty—until two of her sixth-grade students wandered back in. They had heard that their teacher was talking to the people who made StarCoach, and they insisted on saying hello. Dana told them she was busy. The two girls insisted harder. She relented and waved them over.

They leaned into the camera and told me, plainly, that they loved StarCoach—that it made them happy. I was charmed—and curious. I asked what they liked about it. One of them—I'll call her Briana—leaned in and said:

"It makes me feel like I have a friend on my side."

Dana and I both went soft. Then she shoed them back to recess, and we finished the call. But I couldn't stop thinking about it. I have built games

EDITORS
SORIN FAIBISH IEEE Life Senior Member;
sfaibish@comcast.net
MICHAEL ZYDA University of Southern California;
zyda@mikezyda.com



DISCLAIMER

The author is the founder and CEO of StarCoach AI, the company whose pilot is described in this article. All pilot data reported here were collected on the company's own platform and have not been independently audited; see "How to Read These Numbers" for the principal limitations. No external funding supported this article. "The participating district and schools are not identified, and teacher and student names are pseudonyms, to protect participant privacy."

and apps used by millions of people. But this is an unusual compliment to our product. What would make an 11-year-old say a piece of software felt like a friend on her side—when all we were doing was making her do math problems? So I went and dug up her data.

WHAT HER DATA TOLD ME

Briana had walked into the SBA Booster program with nothing. Before any coaching, every student takes a 10-question diagnostic we built from real Smarter Balanced practice items—the SBA being the annual state math exam that 12 states, including Washington and California give students in grades three through eight.^a Briana answered all 10—and got every one wrong.

Then something turned over. In the week that followed, she finished all six of our coaching workshops—ahead of every other student in her class. One session, on turning the words of a problem into math, ran roughly 120 turns of back-and-forth before it clicked. She kept asking. The coach kept answering. It never once made her feel stupid for not getting it yet.

What that week suggested is the argument of this article; an AI coach can let a single teacher offer every student personalized and patient practice—in class and after school—and do it in a

way engaging enough that students come back on their own. The illustration is Briana. The mechanism is how the coach is designed. The stakes, I'll argue, are access.

THIRTY KIDS, ONE TEACHER, ONE HOUR

Personalized coaching has never scaled inside a normal classroom. One teacher, 30 students, one hour—the arithmetic of attention simply doesn't work. A teacher can lecture to 30 at once, but coaching is one-to-one by nature; it means watching an individual student try, catching the specific place they go wrong, and responding in the moment. There is not enough of one adult to go around. Dana, who teaches grades six through eight, put it bluntly.

"There's one of me, and there's 30 of them. Anything that helps differentiate is always going to be a win."

—Dana – pilot teacher

HIGH COST OF PRIVATE TUTORING

As of 2026, a private math tutor for a middle- or high-schooler typically runs US\$40 to US\$100 an hour, and higher in major metros.^b For a family that is working two jobs, an hour a week at US\$60 is not a line item; it is a nonstarter. So the students who most

need a patient second teacher are exactly the ones who never get one, while wealthier districts a few miles away buy as much tutoring as they like. This is the economic core of why AI coaching matters; it is the first time the marginal cost of one more hour of expert individual attention falls close to zero.

THE FEAR OF THE STUPID QUESTION

There is a quieter barrier, and it may be the more powerful one. Children are afraid of asking a "stupid question." They will not raise a hand when they are the only one lost, and they almost never ask the same question twice of an adult who is grading them. The cost of looking dumb, in front of a teacher and 29 peers, is simply too high.

Briana asked her question more than 100 times in a single sitting. She would not have done that with a human watching. That is the unlock. When the listener cannot judge you—and is endlessly patient—the fear evaporates, and a struggling student will keep going until the idea finally lands. Remove the fear, and you remove the ceiling on how much help a child will accept.

RISK OF GENERAL-PURPOSE CHATBOTS

General-purpose chatbots don't solve either barrier well. A horizontal tool that will discuss any subject in any direction is, by construction, hard to keep safe in a classroom—and several districts have learned that the hard way. The first question I got in a district meeting was not about test scores. It was: How are you different from the general AI tools we're worried about? The answer is in the design.

HOW STARCOACH AI IS DIFFERENT

Four deliberate design choices separate StarCoach from a general chatbot pointed at math.

^aSmarter Balanced Assessment (SBA), administered in mathematics to grades 3–8 by Washington, California, and other member states of the Smarter Balanced consortium.

^bA composite of 2026 U.S. tutoring-rate guides (Tutors.com, Brighterly, TutorCruncher): private math tutoring commonly \$40–\$100/h (middle school = \$25–\$88; high school = \$30–\$100), higher in major metropolitan areas.

Affirm every attempt

StarCoach affirms every student attempt before it corrects anything. The affirmation is varied so that it never sounds canned, and it is never negative. That single rule does most of the emotional work. It removes the shame barrier, so students take risks, make mistakes in the open, and repeat questions without embarrassment.

“We built the coach to affirm every attempt because students will risk failure with AI in ways they won’t with a human evaluating them.”

— Michael Yuen, on the design

You can see it in Briana’s own transcript. Asked to describe a word problem in one sentence, she first answered “swimming”—a joke. The coach laughed with her and redirected, without a trace of correction, until she found the real answer herself.

COACH

Read it through once. Ignore all numbers for now. What is the situation? Tell me in one sentence.

BRIANA

swimming

COACH

Ha—not quite, but I like the creativity! Read the problem

again real quick—what is Frank actually doing?

BRIANA

on mb driving

COACH

Exactly—Frank is driving! So put it together: he wants to know how long it will take him to do what?

BRIANA

to go down 455 miles down with the same speed

COACH

Exactly right! That’s a solid Read 1—its only job is to understand the situation before you touch any numbers.

An unedited fragment of Briana’s ~120-turn translation session. Student spelling preserved.

Specialized by design, safe by design

StarCoach is bound to a real curriculum—Illustrative Mathematics for grades six through eight—and it does one thing: math. That narrowness is a safety feature, not a limitation. Because the coach cannot wander off mathematics, it cannot wander into the territory that forces general-purpose tools to be locked down after something goes

wrong. Horizontal tools are made safe by subtraction, walling off harm after it appears. A vertical coach is safe by construction, from the first turn. For a tool that sits with children all day, that difference is everything.

It listens and speaks 10 languages

Roughly 40% of the students in some of our pilot classrooms are still learning English. A coach that works only in English would quietly fail exactly those students. StarCoach listens and responds in more than 10 languages, so a child can be coached—and can answer—in the language they think in. The math doesn’t get easier; the door just opens. A student who would have sat silent in English can reason out loud in Spanish, and the coach meets them there. For classrooms where many students are still acquiring English, this is less a convenience feature than a precondition for the tool being usable at all.

Do the math—Earn game rewards

Engagement and rewards are kept cleanly separate (Figure 1). After real coaching, students unlock the StarDragon—a small, genuinely fun arcade game where you fly a dragon through a field of stars (Figure 2). It is not math disguised as a game, and it is not a sticker. It is a real reward that real work unlocks. Teachers use it as the carrot, and it pulls hard; students will log in at night to finish a workshop so they can earn a few minutes of flight. Two components, each doing one thing well—coaching that coaches, and a game that’s actually a game.

WHAT TEACHERS SAID ABOUT DIFFERENTIATION

Teachers compared StarCoach to the tools their districts already license. Their views are anecdotal, not a benchmark study—but they were consistent, and they came unprompted. Denise, a 20-year veteran, described the difference in her own classroom.

“With the instant feedback—not only, hey, you got

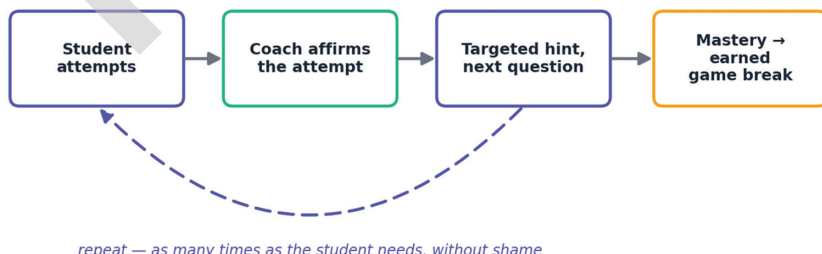


FIGURE 1. The coaching loop: affirmation on every turn, with the game reward kept as a separate, earned component.

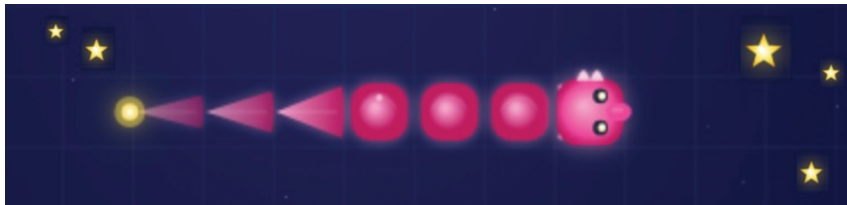


FIGURE 2. The StarDragon reward—a standalone arcade game that students unlock by completing coaching.

this correct, but maybe you didn't, but here's the little mistake you made... We use Khan Academy; it doesn't provide that. We use Prodigy; it does not provide that."

— Denise – pilot teacher

The distinction she is drawing is between content delivery and coaching. Video libraries and practice-game platforms can present material and check answers; what teachers told us they wanted, and saw here, was something that responds to a particular student's particular mistake in the moment—across an entire class at once. That is the capability StarCoach is built around.

STUDENTS STARTED CHOOSING MATH

Across three pilot middle schools, more than 1,000 students enrolled, and about 985 were active. In three weeks, they started 6,530 coaching sessions, completed 1,322 full workshops, and exchanged 64,527 messages with

the coach^c (Figure 3). Classroom engagement ran from 91% to 97% across the most active teachers. And on two weekends, with nothing assigned and no one watching, students logged 1,764 coaching sessions—children choosing math on a Saturday.

AND THEY LEARNED

Among the 79 students who completed all six workshops and took both assessments, scores rose from 46% to 56% correct on average—a 10.1-point gain, with 68% improving^d (Figure 4). On items held at consistent difficulty across both tests, the descriptive gains in coached domains were larger still: number sense up 25.6 points, functions up 23.1 points, and geometry up 21.6 points (Figure 5). For a three-week window layered on top of normal

instruction, those are encouraging early signals—and Briana's arc is the human version of the same curve.

How to read these numbers

This was a single-arm pilot, not a controlled trial; there was no comparison group, and Figures 4 and 5 are descriptive, not tests of statistical significance. The pre/post-sample is also a self-selected subset—79 of roughly 985 active students finished all six workshops and both assessments inside the three-week window, so it skews toward the most engaged students and almost certainly overstates the average effect across all learners. All data come from our own platform and have not been independently audited. I report them as practitioner evidence worth a rigorous external study, not as proof.

SKILLS THAT OUTLAST THE TEST

Test scores aside, the end-of-session reflections were striking. The following three are illustrative rather than systematic evidence of skill transfer, but they were typical: students describing a strategy in their own words rather than parroting an answer.

"I'll use the 3 Reads + Trigger Words Strategy as a habit next time I see a word problem on the SBA."

— Owen – 6th grade – Translation Workshop



FIGURE 3. Cumulative pilot activity over three weeks. Every school day, the curves climbed; weekends held.

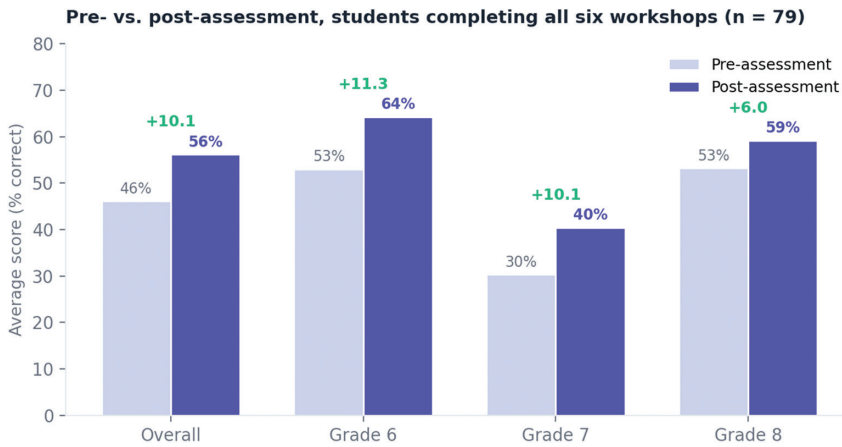


FIGURE 4. Pre- versus post-assessment for students who completed the full program (n = 79).

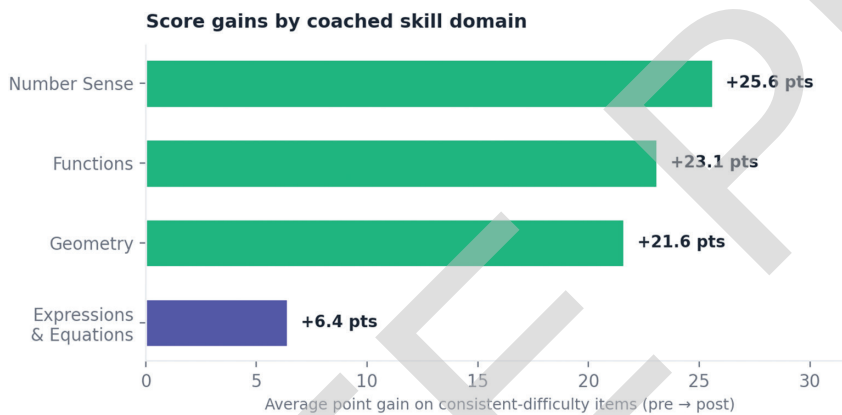


FIGURE 5. Descriptive gains on consistent-difficulty items in coached domains. (Domains whose item difficulty changed between tests are excluded.)

“The one habit I’ll take to the SBA is translating trigger words before touching any numbers. I can build a ‘math sentence’ that acts as a map.”
— Maya – 6th grade – Translation Workshop

“Stop focusing on how the information looked and focus on how the values were organized and related. Once that structure was clear, the answer became much easier to see.”
— Andre – 8th grade – Representation Workshop

coaching block became a routine that students protected.

“I dedicate 15 minutes every day in class—get on StarCoach, do your 15 minutes, and that’s how we warm up each day. Some kids said it’s their new best friend for math.”
— Dana – pilot teacher (97% return rate)

“Fifteen minutes twice a week—regular and continuous. An organic part of the class, not a fire hose right before the test.”
— Paula – pilot teacher

TEACHERS ASKED FOR MORE

Teachers don’t usually beg for more software. These did. The 15-minute

Greg, a 24-year veteran, had only one complaint: that we hadn’t come sooner.

“I wish the district had started this a month earlier.” And Phil Daro, a coauthor of the Common Core State Standards in Mathematics who advises us, watched the pilot and named what he saw.

“The four design principles that drive math learning at scale: mindset, duration, frequency, and recognizable relevance.”
— Phil Daro – co-author, Common Core State Standards in Mathematics

WHERE TUTORING WAS NEVER AN OPTION

This matters most where the US\$60 hour was never realistic. At one pilot school, 61% of students qualify for free or reduced-price lunch,^e and Denise estimates that, on average, about 40% of her students are still learning English.^f A coach that is free at the point of use and fluent in a student’s home language removes the cost barrier and the language barrier at the same time—the two walls that kept private tutoring out of reach.

And the demand is coming from the families themselves. Sarah’s students beg for the reward game—“we promise we’ve done, like, five solid minutes, please give us the game”—and, more tellingly, the requests have started reaching her from parents.

“I’m getting that question from parents as to what summer supports we have... A lot of students are asking if they could have StarCoach during the summer.”
— Denise – pilot teacher

That is the equity argument in one sentence. The communities the system usually leaves behind are the ones asking for more.

^eFree/reduced-price-lunch eligibility at one participating school, ~61% (U.S. Dept. of Education, NCES Common Core of Data, 2024–2025).
^fTeacher interview, 2 June 2026 (teacher’s estimate of English-learner support needs across her sections).

MODEL CHOICE AND SAFETY

A coach that interacts with children all day raises a real engineering question: which model, and under what guardrails? We built StarCoach on a frontier large language model selected for instruction-following reliability and its developer's emphasis on alignment and safety testing (in our current deployment, Anthropic's Claude Sonnet 4.6). But the more important safety property is architectural, not vendor specific. Because the system is scoped to a single subject and bound to a defined curriculum, the space of plausible interactions is narrow; the model is prompted, retrieved, and constrained around mathematics problems and pedagogy. A general-purpose assistant must be made safe by

subtraction—anticipating and walling off harm across an open domain after it appears. A vertical tutor is safe largely by construction because the surface area for misuse is small from the outset. For a classroom deployment, that containment matters more than any single model choice.

A WHOLE SUMMER FOR THE PRICE OF AN HOUR

The pilot that began as a three-week test-prep booster turned into a daily classroom routine and a request we did not anticipate: keep it going over the summer. So that is what we're doing. A full summer of AI coaching—built on the same curriculum, backed by the same advisors—now costs about what a single hour of private tutoring does.

The same US\$60; one whole summer instead of one session. Briana already knows what that feels like. She told me herself: a friend on her side.

ACKNOWLEDGMENTS

The author thanks Phil Daro for his guidance on the pedagogy underlying this work, and Paul Conley, who led the StarCoach AI classroom pilot. Thanks also to the teachers and students of the participating Washington school district, whose participation made this report possible. 

MICHAEL YUEN is the founder and CEO of StarCoach AI, San Francisco, CA 94116, USA. Contact him at michael@starcoach.ai.

When the listener cannot judge you—and is endlessly patient—the fear evaporates, and a struggling student will keep going until the idea finally lands.

StarCoach is bound to a real curriculum—Illustrative Mathematics for grades six through eight—and it does one thing: math.

StarCoach listens and responds in more than 10 languages, so a child can be coached—and can answer—in the language they think in.

After real coaching, students unlock the StarDragon—a small, genuinely fun arcade game where you fly a dragon through a field of stars.

On two weekends, with nothing assigned and no one watching, students logged 1,764 coaching sessions—children choosing math on a Saturday.

A coach that is free at the point of use and fluent in a student's home language removes the cost barrier and the language barrier at the same time.