

The 5 Percent Problem

Online mathematics programs may benefit most the kids who need it least

IN 1924, SIDNEY PRESSEY, a professor from Ohio State University, invented a teaching machine. The mechanical device, about the size of a portable typewriter, allowed students to press one of five keys to answer questions curated by expert instructors. A later version dispensed candy for correct answers.

Education optimists were fascinated, and Pressey promised the technology would accelerate student learning. But the machine was a commercial flop.

Exactly a century later, similar programs spangle U.S. classrooms: i-Ready, DreamBox, Khan Academy, IXL, and many others. They are driven by clever algorithms rather than finger power. Though none feature candy dispensers as rewards, some have animations or videos explaining what a student got wrong.

By LAURENCE HOLT





PHOTO-ILLUSTRATION / SHUTTERSTOCK & ISTOCK

The pandemic mania for teaching kids on computers prompted a great surge in the adoption of such programs.

Do they work? In August 2022, three researchers at Khan Academy, a popular math practice website, published the results of a massive, 99-district study of students. It showed an effect size of 0.26 standard deviations (SD)—equivalent to several months of additional schooling—for students who used the program as recommended.

A 2016 Harvard study of DreamBox, a competing



Sidney Pressey's 1924 invention was among the first attempts to employ technology to improve student learning.

mathematics platform, though without the benefit of Sal Khan's satin voiceover, found an effect size of 0.20 SD for students who used the program as recommended. A 2019 study of i-Ready, a similar program, reported an effect size in math of 0.22 SD—again for students who used the program as recommended. And in 2023 IXL, yet another online mathematics program, reported an effect size of 0.14 SD for students who used the program as designed.

Those gains, and many others like them reported each year, are impressive. Since use of these tools is widespread, one could be forgiven for asking why American students are not making impressive gains in math achievement. John Gabrieli, an MIT neuroscientist, declares himself “impressed how education technology has had no effect on . . . outcomes.” He was talking about reading but could equally have called out

mathematics, the other big area in which education technology is widely used but growth in achievement has not followed.

A clue is in those wiggly words “students who used the program as recommended.” Just how many students *do* use these programs as recommended—at least 30 minutes per week in the case of Khan Academy? The answer is usually buried in a footnote, if it's reported at all. In the case of the Khan study, it is 4.7 percent of students. The percentage of students using the other products as prescribed is similarly low.

Imagine a doctor prescribing a sophisticated new drug to 100 patients and finding 95 of them didn't take it as prescribed. That is the situation with many online math interventions in K–12 education today. They are a solution for the 5 percent. The other 95 percent see minimal gains, if any.

Worse, some studies report that the 5 percent who do see results skew towards higher income, higher performing students. A 2022 study of Zearn, another math learning platform, in Washington, DC, public schools found that students who used the program most were more likely to be white or Asian and from high-income areas of the city and less likely to be considered at risk. (Other studies, including the one of Khan Academy, show no particular pattern across student groups.) Learning gains for any group of students are to be welcomed, but it may be that the 5 percent of learners who achieve strong results with these programs could achieve the same strong results with any practice program, including paper-and-pencil practice. At the very least, district leaders who adopt online learning programs with the aim of reducing equity gaps in math should be aware that they may be widening them.

It's not at all clear that the program vendors are at fault, any more than you would blame a pharmaceutical company for the failure to see results among patients who didn't take their drug. Indeed, the vendors point to data that students who use their program more show higher performance. But that is a correlation. As Hilary Yamtich, a fourth grade math teacher at a school in Oakland, California, who conducted a study of her own, points out, “students who are more motivated to learn are more likely to choose to use Khan.”

There may be other reasons beyond motivation that some students use these programs to such different extents. One possibility is that some teachers are more committed to implementing the programs than others—if they selected the program themselves, for instance. And “programs that have been carefully integrated into the curriculum, rather than seen as supplemental to it, likely see more consistent usage patterns,” according

Imagine a doctor prescribing a new drug to 100 patients and finding 95 of them didn't take it as prescribed. That is the situation with many online math interventions in K-12 education today. They are a solution for the 5 percent.

to Sarah Johnson of Teaching Lab, a nonprofit focused on teacher coaching. The Harvard study of DreamBox found that the variation in student usage was driven more by “teacher- and school-level practices” than by “student preferences.”

A second theory focuses on student behavior: perhaps

some students use the program at home while others do not. The pandemic forced many school districts to address disparities in access to technology, but not all students have parents who badger them to do their homework or have time to assist with it.

Other students may simply be more motivated to do well in math, as Yamtich says, or more assiduous in following their teacher's instructions. Another Zearn study found that high-usage students were more likely to believe they can improve in math, an attitude researchers refer to as a “growth mindset.” The researchers concluded that using their program led to a better mindset, but the causal arrow could equally point in the opposite direction: students with a growth mindset invest more time in trying to improve.

Third, the programs may have been unintentionally designed to fit high achievers better, says Stacy Marple, a researcher at WestEd who has studied several online programs. Marple tells the story of a 7th grade classroom where she observed an online program that asked a student, “Using the principles of equivalency and inverse



AP PHOTO / MICHAEL MELIA

Heather Dalton, a teacher at Charles Barnum Elementary School in Groton, Conn., helps her fifth grade students use Dreambox, a web-based platform for teaching math. Studies find that consistent use of such programs depends on the intention and commitment of teachers.

At the very least, district leaders who adopt online learning programs with the aim of reducing equity gaps in math should be aware that they may be widening them.

operations, isolate the variable” in an equation. The student clicked each possible answer in turn. “Do you know what [the question] is asking you?” Marple asked the student. “Umm, not really,” she replied. (Your correspondent is similarly unsure.) The program offered no way for students to look up the meaning of a word like “isolate.”

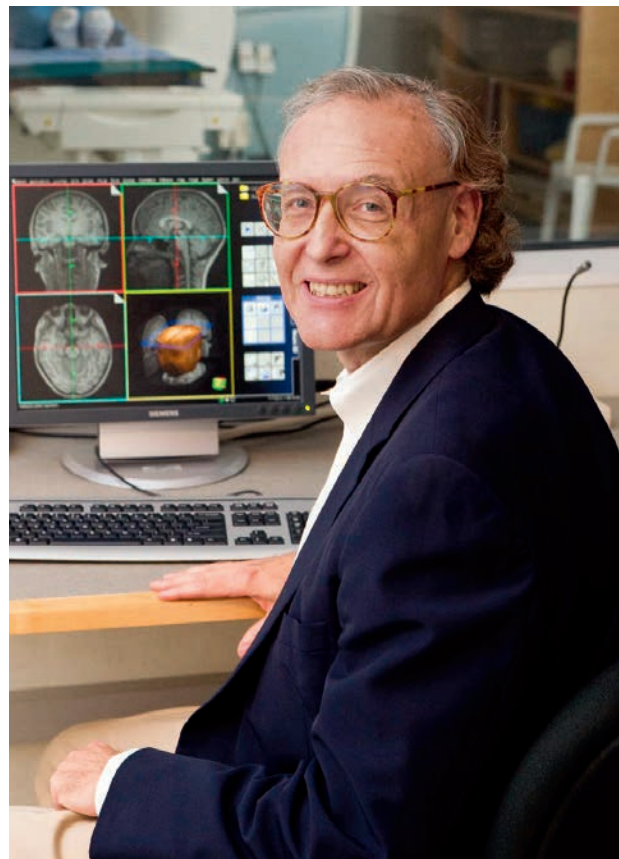
In another classroom, students were provided with video explanations of concepts they struggled with. But few students ever watched a video, since doing so was considered a “hint” by the program and resulted in points being deducted from their score. That, in turn, might result in them having to repeat a problem set from the beginning. What is meant to help students instead makes some feel like a bird in a box, which they are most desperate to avoid.

Whatever the reason for low usage—and it is likely a combination of all of these—schools should assume the impact of online learning programs will be limited unless they take steps to ensure the students who need them most get the recommended dosage. This is especially important since, as Ken Koedinger, an expert on education technology at Carnegie Mellon University, points out, there is solid evidence that the amount of practice a student does directly impacts their learning. Recognizing this, an initiative in Texas, for example, offered grants only to those districts that submitted a plan to achieve fidelity to a learning program and a way to track it. At the very least, districts should make a habit of looking at usage data from these platforms.

Since schools pay for these programs, it would be



Oakland math teacher Hilary Yamtich found students who used Khan Academy were already motivated to learn.



MIT neuroscientist John Gabrieli noted how, as in math, reading technologies have had no effect on outcomes nationally.

fair for taxpayers to ask if their dollars are being wasted. That could lead to schools seeking agreements with vendors so they pay only for time used. “Even better,” says Raymond Pierce, president of the Southern Education Foundation, a non-profit founded to advance education opportunities, “would be to pay solely for growth in student achievement.” That seems sure to focus the minds of vendors’ executives who can expect, on current performance, to lose their shirts.

Districts may have been lulled into a false sense of security by the research reports published by vendors. Federal rules for schools’ use of Title I funds in low-performing schools require them to purchase only interventions that have evidence of effectiveness for a sample of at least 300 students. But it says nothing about what percentage of the student body that 300 should represent. The urgent question is not just whether the tools are effective but for whom. One hundred years after Pressey, we still don’t know.

Laurence Holt is a Senior Advisor at XQ Institute and author of The Science of Tutoring.