Support Builds For Making the SAT Untimed for Everyone

A POSSIBLE SOLUTION TO THE “GAMING THE SYSTEM” PROBLEM

THE FIRST TIME SHE TOOK THE SAT, in December 2013, Savannah Treviño Casias got what she calls “a really low score”—1040 out of 1600. It put her squarely in the 50th percentile—not so good for someone aspiring to be a psychologist.

At the time a junior at the Girls Leadership Academy of Arizona, a small public charter school in Phoenix, Treviño Casias had been diagnosed in 6th grade with the math learning disability known as dyscalculia. The diagnosis qualified her for extra time on classroom tests and quizzes, along with other accommodations. But she’d sat for that first SAT without requesting extra time, taking it in four hours along with hundreds of other students at a nearby high school.

Treviño Casias vowed to do better. She arranged for a family friend to tutor her over the following six months—and she asked for extra time on the next test.

She is not alone. Requests for more time and other accommodations on the SAT have soared in recent years, to an estimated 160,000 in 2015–16 from 80,000 in the 2010–11 school year. At the same time, the jaw-dropping “Varsity Blues” scandal in 2019 shed new light on parents determined to get their kids an advantage on such tests, at any cost.

As schools grapple with these realities, a few educators, researchers, and psychologists have begun to wonder whether it’s time to make a fundamental change to tests like the SAT so that they’re harder to game. More broadly, they ask: If success in college is about 21st-century skills such as critical thinking, close reading, and collaboration, should gate-keeping tests such as the SAT be timed at all? Advocates argue that making the test untimed for everyone would make it harder for rich or well-connected parents to game the system, and also might do a better job of measuring students’ true capabilities.

“How do we measure whether people have the capacity

by GREG TOPPO
to do deep thinking and be thoughtful?” asked Ohio State University law professor Ruth Colker. “My hypothesis is: it’s by giving them enough time to do deep thinking and be thoughtful.” The ability to answer test questions quickly, she added, is itself “a skill—and it’s a skill that can be learned. But I think we tremendously overweight that skill.”

Others, such as Gregory Cizek, a professor of educational measurement and evaluation at the University of North Carolina School of Education, say college admissions tests like the SAT must be timed because they’re essentially trying to predict a student’s chances of success in freshman year.

“The college environment is one where you don’t get unlimited time to do stuff,” he said. “You have to perform under a sort of time pressure. Your term paper is due by the end of next week. You’re going to have a quiz in class. You’re going to have to stand and give a group report or discussion. And you’re not going to have unlimited time to prepare it.”

But Colker, who is among the foremost advocates for an untimed or extended-time SAT, has written that standardized test developers shouldn’t be allowed to implement “speeded exams” unless they can show that the strict time limits are truly required for validity.

Shifting to unspeeded exams, she wrote, “would mark the implementation of a new universal design principle that would make standardized testing more equitable for a range of people.” “Universal design,” in this context, would entail creating testing conditions that optimally serve everyone, including members of racial minorities, girls, people of low socioeconomic status, older applicants, and individuals with disabilities—both students with Individualized Education Plans, or IEPs, which provide for specialized instruction and services for children with disabilities; and those with Section 504 plans, which allow for accommodations for children with disabilities. Under either designation, students can qualify for extra time on tests.

Think of an unspeeded SAT, then, as offering all test-takers the cognitive version of curb cuts, automatic doors, or closed captioning—designs that are especially beneficial to users with disabilities but that end up serving many others as well.

To be sure, many students come to school with genuine learning challenges. Nationwide, about 13 percent of K–12 students had IEPs in the 2015–16 school year. About 2.3 percent of students had a 504 plan, up from 0.9 percent in 2003–04. A 2015 analysis by the Advocacy Institute, a nonprofit that supports people with disabilities, found that 504 students are “overwhelmingly white and disproportionately male.” Other analyses have found that 504 students are more likely to be enrolled in non-Title I schools, which serve a smaller percentage of low-income students.

Recent amendments to federal law allow students to ask for testing accommodations under 504 plans because of difficulty reading, concentrating, and doing manual tasks; that provision also applies to students who have major physical disabilities that wouldn’t necessarily earn them an IEP. More time is just one possible accommodation—students can request more (and longer) breaks, calculators, computers to type essays, and large-print, Braille, or audio tests, among other accommodations.

Advocates of unspeeded tests often point to a 2002 study by the testing company Pearson, which found that scores on its Stanford 10 Achievement Test were equally valid under speeded and unspeeded conditions. Researchers learned that giving students with disabilities more time helped them show what they knew. It also didn’t unfairly inflate the scores of the non-disabled students who got the extra time. And researchers found that in most classrooms, students actually didn’t need that much more time—about 20 minutes, on average—to finish the exam to their own satisfaction.

The president and CEO of the National Center for Learning Disabilities, Lindsay Jones, said it’s no coincidence that many state end-of-year assessments are now untimed for all students. “Our states have figured out how to do that,” she noted. “The premise is they want to know what you know. They’re not testing how quickly you can get it done.”

In another often-cited study, from 2004, an Indiana University law professor, William Henderson, found that when he looked at law students’ performance on three key indicators—in-class exams, take-home exams, and papers—a student’s score on the tightly timed Law School Admissions Test turned out to be a “relatively robust” predictor of just one of the three: in-class exams. It was a relatively weak predictor of take-home exams and papers.

In other words, a timed test showed how well students would do . . . on other timed tests.

By contrast, undergraduate grade point average was a “relatively stable predictor” of all three indicators.

Henderson concluded that heavy reliance on the timed LSAT—which like other standardized tests suffers from a performance gap between white and minority students—could make
it more difficult for minority students to be admitted to good law schools, even if they apply with a high grade point average. Relying on such time-pressured tests, Henderson concluded, “may skew measures of merit in ways that have little theoretical connection to the actual practice of law.”

Others, such as a professor of electrical and computer engineering at Boston University, Ari Trachtenberg, say the whole field of time-related accommodations is squishy at best. He wrote recently that extra time on college-level tests skew the results in “a manner that is not rigorously and objectively analyzed or understood.” Trachtenberg dug through the research and couldn’t find any objective basis for calculating, for instance, how much extra time a student with a certain disability actually needs.

According to Trachtenberg, the system disadvantages both students with disabilities who don’t know how to advocate for themselves and non-disabled students, neither of whom enjoy extra time. “It is inappropriate to give an objective test with a clearly delineated grading policy,” he wrote, “if some students get uncalibrated bonuses.”

At the same time, he explained, extra-time accommodations may actually “re-victimize” engineering students with disabilities later in life, setting them up for failure on “high-pressure tech interviews and subsequent jobs that do not, and cannot, honor time extensions for deadline-driven work.”

Is there a way to break out deadline-driven work from all the rest and test it accordingly? An Atlanta neuropsychologist, Marla Shapiro, suggested that it’s possible. She recalled working, earlier in her career, with a medical group (she won’t say which one) that developed a two-part certification test. The first part, a written exam that “had nothing to do with the practice of that field,” offered time accommodations. But for the clinical portion of the test, the group set a time limit by surveying practitioners and calculating the longest time duration that “a reasonable practitioner” should need to review a case. The medical group allowed test-takers no exceptions to the time limit.

“They held firm to it,” Shapiro recalled, “because the test was meant to determine whether a person could practice competently, and if test-takers could not do it within a generously large period of time, the board did not think it was fair to find them competent to practice in a field where you bill by the hour.”

At the moment, the College Board’s position on speededness is basically that it doesn’t make much of a difference. “Extra testing time leads to very small score gains,” said spokesman Zachary Goldberg. He suggested that a better way to give students a chance to do well on the test would be to give them access to good test prep: Goldberg noted that 20 hours using free Khan Academy SAT materials yields, on average, a 115-point score gain.

All the same, he said, new versions of the test give students 43 percent more time per question than any similar exam. The move is part of a shift that also gives students taking some Advanced Placement tests nearly twice as much time on multiple-choice questions than in past years.

Gaming the System

Accommodations, deserved or undeserved, have been under the microscope in 2019. They played a prominent role in this year’s Varsity Blues college admissions scandal, with prosecutors alleging that wealthy parents persuaded willing psychologists to say their child needed extra time in special testing centers—in a few cases, ringers proctored the exam and cheated on a student’s behalf. The New York Times reported from court filings that in one case, William Singer, the scandal’s mastermind, told a parent that for $4,000 or
$5,000, a psychologist he worked with would attest that his child needed more time.

In May, the Wall Street Journal twisted the knife, finding that the number of public high-school students with diagnoses that allow more time on tests such as the SAT has “surged” in schools located in wealthier areas. Overall, the newspaper reported, the number of requests for special accommodations received by the College Board jumped 200 percent from 2010–11 to 2017–18. Eligibility for such accommodations is growing fastest, federal data suggest, in the nation’s highest-income school districts. The Journal found that in schools where no more than 10 percent of students are low-income, 4.2 percent of students, on average, have 504 designations, entitling them to test-taking allowances such as extra time. By contrast, just 1.6 percent of students in low-income areas have 504 designations, despite the fact that more students in these schools may actually qualify under federal law for services under designations such as 504s or IEPs.

In July, the New York Times did its own analysis of federal 504 data and found that in wealthy areas the share of students with the designation is twice the national average. In a few communities, such as Weston, Connecticut, the 504 rate is 18 percent, eight times that of nearby Danbury.

That alone should raise alarm bells, said Nicole Ofiesh, a cognitive behavioral scientist who directs Stanford University’s Schwab Learning Center. Ofiesh has done extensive pro-bono testing among low-income students and says these students often have the highest incidence of learning disabilities, such as ADHD and dyslexia—as well as related mental-health issues. “They are often under-identified dramatically because the resources in the schools that they attend are not adequate to identify these students and get them on 504 plans and IEPs.”

Like Ohio State’s Colker, Ofiesh pointed out that most observers don’t even stop to ask, Why are high-stakes tests like the SAT timed at all? “Very few test agencies will give you an adequate answer to that,” she said. “It’s almost always an administrative reason: ‘It’s because that’s the only way we can do it in an efficient way.’”

For his part, College Board CEO David Coleman wrote in a subsequent letter to the Journal that protecting the SAT’s security and credibility is “an ongoing battle.” Families seeking test accommodations without genuine need is “abhorrent,” he said, and he promised that the College Board would take a hard look at non-special needs schools that have inexplicably high numbers of requests for accommodations.

“We have heard from schools that know they have a problem with this,” Coleman wrote. “They want our help, and we are providing it.”

A longtime admissions expert and former dean of admissions at Pomona College, Bruce Poch, said the College Board helped create the current dilemma in 2003, when it stopped telling colleges about students’ extra-time requests, removing the so-called “asterisk” on scores earned by students who got extra time.

“In some sense, the College Board handed the keys to this problem to the world,” Poch said. “The removal of that asterisk sparked a massive increase in the number of kids applying under 504.”

At the time, the move was viewed as a bid for fairness, a way to remove any stigma attached to getting extra time for those students who genuinely needed it. The College Board’s Goldberg said it was “the right thing to do, so we stand by it for obvious reasons.”

Poch, now dean of admission and executive director of college counseling at the Chadwick School, an independent
Think of an unspeeded SAT as offering all test-takers the cognitive version of curb cuts, automatic doors, or closed captioning—designs that are especially beneficial to users with disabilities but that end up serving many others as well.

A school near Los Angeles, said wealthy parents will always find a way to make the system work for themselves.

“This is an arms race, so whatever you do to make it fair, people are going to find something else to game it another way,” he said.

Shortly after the change, Harvard researcher Samuel J. Abrams wrote in *Education Next* that the College Board had granted “new opportunities to the strategic, while leaving behind the less savvy and less financially well-endowed” (see “Unflagged SATs,” *features*, Summer 2005).

Poch recalled that when he got into the admissions business in the late 1970s, dyslexia was the wealthy family’s diagnosis du jour for otherwise typical students who earned grades lower than A in class. Then attention deficit disorder became fashionable—then attention deficit hyperactivity disorder. “Now it’s ‘general anxiety disorder,’” he said.

He observed that the Varsity Blues scandal has actually inspired helpful conversations on campus. To date, authorities haven’t identified any Chadwick families as conspirators, Poch said. But the blatant cheating among well-to-do parents has prompted teachers, students, and parents there to think about how well they uphold the school’s core values, which include “honesty” and “fairness.”

“It led to a good, frank conversation and more rigorous follow-up with [504] evaluations,” he said.

The scandal has also been a bracing reminder about the possibilities for dishonesty and unfairness among certain independent schools in upscale communities, Poch said. “You can, with just a few little cocktail parties, find out the people who will write those notes that say the kid needs extended time.”

Stanford’s Ofiesh sees the obvious disparity in accommodations granted to wealthier, more connected students—especially for conditions like anxiety and depression. But short of unlawful behavior like bribes, she said, most of us are viewing the dilemma through the wrong lens.

“While we often want to say that it’s the wealthiest who are gaming the system, it’s usually the wealthiest who are able to do something about the fact that their kids have these mental-health conditions—and then can get them the accommodations,” she said. “They’re not cheating the system or gaming the system—they have the resources to maneuver the system that those who don’t have money don’t.”

Marla Shapiro, the Atlanta neuropsychologist, would agree. For students who come seeking help, she said, “I give a big old battery and lots of measures that overlap” to zero in on students’ behavioral and cognitive deficits. And she requests extensive school records to get a full picture of a student’s performance before she will consider recommending accommodations like extra time.

“If I have someone coming in for testing and I can’t get records, I don’t test them,” she said.

But she and others reported that most school districts don’t have the capacity to do the proper screening. And our health-care system is not very helpful—it basically blocks most families’ access to good, thorough—and expensive—psychological testing, the kind that would get many students the accommodations they need while rooting out those who don’t need them. “I pay my plumber more than Blue Cross would pay me for testing,” Shapiro said, remarking that an insurance company recently offered her $54 an hour for her services.

Others, like Ohio State’s Colker, question why timing most tests even makes logical sense. In the case of content-based tests, for instance, neither timing nor giving vulnerable students more time seems constructive. If a student doesn’t know how to apply *Plessy v. Ferguson* or the Pythagorean theorem, she observed, “I could give you all day. If you don’t know how to do it, more time won’t help you.”

In cases like these, she said, speededness is good at producing one key product: “a beautiful bell-shaped curve” that helps colleges detect fractional differences between top applicants.

Colker suggests imagining two whip-smart students sitting for the same multiple-choice reading comprehension test—two students with equal reading skills and vocabulary, but one of whom works at a faster rate. Let’s say the test is a section of a newspaper article. Both students read and understand the article completely. But one student, working more quickly, finishes the test with minutes to spare. The second, working more slowly, soon realizes that she’s running out of time and begins “engaging in rapid guessing behavior,” filling in as many blank circles as she can.

In the real world, the two readers would be virtually indistinguishable when it comes to understanding the day’s news. But on this test the slower reader gets a lower score. What was accomplished?

Plenty, said Colker: “If you’re looking for a test that can especially distinguish the top one or two percentile of the curve, speededness is very helpful.” Get rid of it, she said, and you’ll have more perfect scores, which to a psychometrician is an unfortunate, avoidable inconvenience.

“If you thought that your job was to produce a perfect
bell-shaped curve and not have too many perfect scores on a standardized testing instrument, then absolutely you would add a speeded element, because then fewer people will get a perfect score,” Colker said. “But I think that’s a pretty crappy reason to have the test speeded.”

UNC’s Cizek, who is also a member of the National Assessment Governing Board, observed that the goal of most test-makers isn’t to produce a bell curve. “If we asked a thousand people to run the 100-yard dash, some people would be really speedy, some people would be really slow, and probably you and I would finish in the middle of the pack,” he said. “It’s just an artifact of a lot of human aspects that turns out to be kind of bell-shaped, but they don’t design it to get that.”

Hard to Say No

Perry Zirkel, a professor emeritus of education and law at Lehigh University, noted that the rate of 504 diagnoses has more than doubled since 2003, when the College Board got rid of the “asterisk.”

Most of the growth came after 2008, when Congress may have unwittingly made it easier for parents to game the system (see Figure 1). Seeking to give returning veterans better access to rehabilitation services, it allowed students to test their performance without medication, making it more likely that they’d qualify for help. So, for instance, a student who is prescribed Ritalin may test in the 50th percentile in reading while taking the drug, but in the 15th percentile without it. Under the 2008 rules, he’d qualify for an accommodation if he sat for a screening without Ritalin—even though he may take the medication every morning before school and do just fine.

At an annual “504 institute” Zirkel holds at Lehigh, school officials each year mostly want to know more about current legal standards. “The implicit agenda is, ‘How do we say no? How we do withstand this pressure?’” he said.

“It isn’t just a simple matter of sitting down with a team of people and looking at these legal standards. When you say no, it causes all kinds of pressure and backlash and it’s just very, very hard,” Zirkel said.

In general, cash-strapped school districts “don’t want to say no” to parents who insist on a testing accommodation, Zirkel added. In practical terms, giving a student more time on tests is cheap. Fighting a well-connected family could be expensive. “The inclination is to sort of give in.”

Zirkel recalled that when he taught undergraduates, he found that many students who legally qualified for extra time on tests throughout their education hadn’t learned to advocate for themselves as adults. “The day before the final they’d say to me, ‘I’m entitled to extra time.’ And I’d say, ‘Look at the university policy—you

### More Students Are Accessing Testing Accommodations under 504 Plans (Figure 1)

The rate of 504 diagnoses, which allow for accommodations for children with disabilities, has soared in recent years.

<table>
<thead>
<tr>
<th>Academic year</th>
<th>Rate of 504 diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009-10</td>
<td>0</td>
</tr>
<tr>
<td>2011-12</td>
<td>1</td>
</tr>
<tr>
<td>2013-14</td>
<td>1.5</td>
</tr>
<tr>
<td>2015-16</td>
<td>2.5</td>
</tr>
</tbody>
</table>

SOURCE: Civil Rights Data Collection
Eventually, Zirkel found, keeping track of all the extra-time requests became an administrative nightmare. “I gave everybody extra time,” he said. “As much as you want.” He discovered that those who had better basic skills did better on the tests. “I was taking away this false advantage that these kids had,” he said.

It was, in a way, revelatory. If you really care about a student, he said, often “the Band-Aid of a 504” is not going to be beneficial. Instead, he suggests, schools should do a better job helping all students attain better reading fluency, reading comprehension, time management, and self-advocacy.

For his part, Cizek said having untimed state assessments in kindergarten through 12th grade is a very bad idea. “It really disadvantages kids” with accommodations, “and people just aren’t willing to stomach the political fallout from it,” he said.

Because the state tests are often used to measure teacher effectiveness, teachers have little incentive to push students to finish in a timely fashion, Cizek said. “Their incentive is to say to the kid, ‘Hey look: If you need more time, take more time. Have you checked your answers? Maybe you want to sit down and check them again.’”

Students whose IEPs or 504 plans lay out the need for more time on tests generally get it, Cizek said. But when educators make these tests untimed, they give extra time to students who don’t necessarily need it, giving them an unfair advantage. What’s perhaps most insidious about this arrangement, he noted, is that students who end up taking more time are the very ones who miss out on instruction when the other students finish on time and go back to the classroom.

“A kid who was behind anyway is falling further behind because they’re now losing out on the instruction, when somebody should have just said to him, ‘Look: This is a one-hour test. You’ve been at it for two hours. You should probably wrap it up now.’ But it’s hard to say that to a low-performing kid. I get why it’s not an attractive policy position, but by and large it’s actually harmful to kids,” Cizek said.

The speededness discussion may soon be moot, as test-makers are increasingly tinkering with computer-adaptive assessments that turn the timed-test paradigm on its head. Instead of requiring all students to sit for a defined number of questions over a set time, these tests pose questions in a progressively customized sequence, quickly generating a probability that the user knows the required material.

Cizek said that perhaps half of the states rely on adaptive tests for K–12 assessment. What’s more significant, about 80 percent of professional licensure certification is now adaptive, including tests for medical and legal licenses and for other professions such as real estate.

The College Board’s Goldberg said they’re looking into the possibility of using adaptive assessment technology for the SAT, but that it’s “really in an exploratory phase and not so far along.”

As for Treviño Casias, the Phoenix charter-school student, the second time around she got twice as long on each section of the SAT. She was allowed to sit by herself, over two days, in a classroom at her old, familiar high school. One of her teachers proctored the test. “It just helped me feel more comfortable and I guess just have less testing anxiety,” she said.

Her second score: 1550, more than 500 points higher than the first, putting her among the top 1 percent or so of test-takers. “I was really surprised,” she said. “I just expected a little improvement. I didn’t expect 500 points.”

Now 23, Treviño Casias graduated in May from Arizona State University. She earned the bachelor’s degree in psychology she’d been seeking—and began studying for a master’s degree in counseling in August.

Even with twice as much time on her second SAT, she said, she couldn’t do her best work. “I still could have used even more time to figure out a math problem—and even more time to actually have my mind process the information.”

She still doesn’t quite see the point of speededness, especially for someone training to be a therapist. “I wouldn’t want to give any spur-of-the-moment advice,” she said. “I would be working with a person—their whole being would be kind of in my care. So I wouldn’t want to rush things.”

Greg Toppo, former national education writer for USA TODAY, is author of The Game Believes in You: How Digital Play Can Make Our Kids Smarter.